

The Evolutionary Rate Variation Among Genes of MVA and MEP Pathways in Plant Terpenoid Biosynthesis

Ji-peng Mao

South China Agricultural University

Lin-wang Huang

South China Agricultural University

Jing Hao

South China Agricultural University

Tian-yi Liu

South China Agricultural University

Shao-wei Huang (✉ shwhuang@scau.edu.cn)

South China Agricultural University <https://orcid.org/0000-0001-8702-0626>

Research article

Keywords: MVA pathway, MEP pathway, Evolutionary rate, Selective constraints, Positive selection

Posted Date: December 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-134044/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

The evolutionary rate variation among genes of MVA and MEP pathways in plant terpenoid biosynthesis

Ji-peng Mao^{1,2,3}, Lin-wang Huang¹, Jing Hao⁴, Tian-yi Liu¹, Shao-wei Huang^{1*}

1. Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China.
2. Taishan Hongling Seed Orchard, Taishan, China.
3. State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, South China Agricultural University, Guangzhou, China.
4. College of Life Sciences, South China Agricultural University, Guangzhou, China.

*Corresponding author: E-mail: shwhuang@scau.edu.cn.

Abstract

Background: Terpenoids are one of the most important compounds in plants, play an significant biological defense and developmental roles in numerous plant species, and widely used for industrial chemicals. Many previous studies have completed the identification of terpenoid biosynthetic pathway and related genes. However, few studies have focused on the molecular evolution analysis of terpenoid pathway genes in plants. In this study, we researched the evolutionary rate variation pattern of 16 terpenoid pathway genes in 12 species with a broad taxonomic span.

Results: We retrieved 14 genes in MVA and MEP pathways and 2 extra genes from 12 species, respectively. The evolutionary parameters d_N values and d_N/d_S ratios are varied significantly among genes, and the d_N/d_S ratios of most genes are varied substantially among lineages. The MVA and MEP pathways genes have different evolutionary rate variation pattern, although no significant difference in d_N/d_S ratios between two pathways genes. For MVA pathway, the downstream genes exhibits the greater d_N/d_S ratio than upstream genes. For MEP pathway, the three midstream genes evolves more rapidly than other genes, and most of MEP pathway genes were detected the signature of positive selection under random sites models. Moreover, the d_N/d_S ratios of MVA and MEP pathways genes are negatively correlated with pathway position and PPI, and coding sequence length, respectively.

Conclusions: Taken together, the results indicated that the evolutionary rate variation of MVA pathway genes is mainly attributed to differential selective constraint rather than the positive selection. However, the differential selective constraint relaxation and positive selection collectively shaped the evolutionary rate heterogeneity of MEP pathway genes.

Keywords: MVA pathway, MEP pathway, Evolutionary rate, Selective constraints, Positive selection.

Background

Terpenoids are the most abundant natural products, and present in all living organisms, and characterized by complex skeletons, kaleidoscopic structures and extensive biological activity [1,2]. According to incomplete statistics, more than 55,000 terpenoids have been identified in extant organisms. There are known to have many momentous biological and physiological functions, and are widely used in pharmaceutical, cosmetic, food, beverages, pigments and condiment industries [3-6]. Such as, some terpenoids are crucial components of cytoderm, membrane structures and electron transfer systems in plants [7]. The lineage-specific terpenoids plays an important role in plant defense against herbivores and pathogens [8,9]. Partial terpenoids of secondary metabolism have been essential to many facets of human life (e.g. artemisinin for malaria, taxol for ovarian/breast cancer) [10,11].

All terpenoids are derived from the common five-carbon structural units precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), though there have a variety of types with distinct structures and functions [12,13]. IPP/DMAPP can be generated via two independent and cooperative pathways: the mevalonate (MVA) and 2-C-methyl-D-erythritol-4-phosphate (MEP) pathways occurred in cytoplasm and plastids, respectively [1] (Figure.1). However, most organisms only use one of the MVA and MEP pathway for the biosynthesis of IPP/DMAPP [14,15]. Specifically, higher plants use both the MVA and MEP pathways for IPP/DMAPP biosynthesis, such an advantage maybe for overcoming the sessile lifestyle constraints [16].

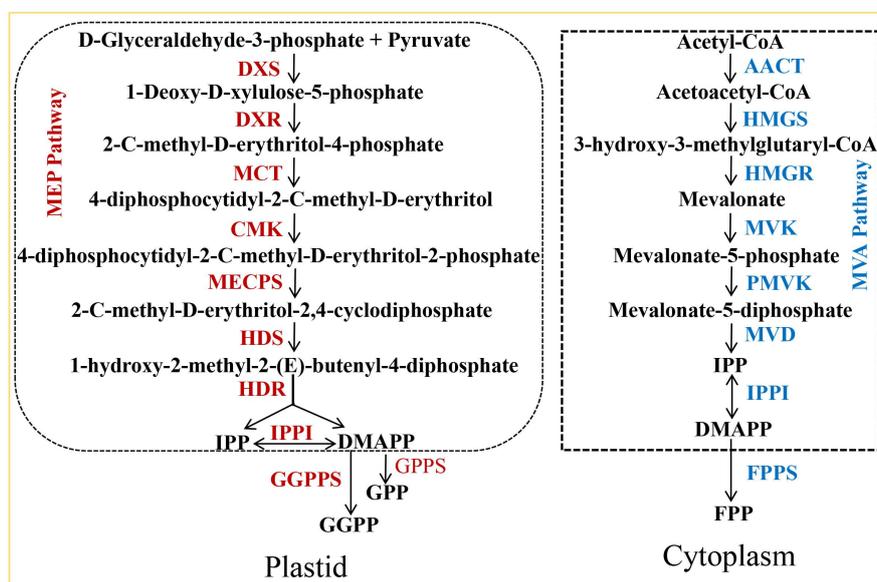


Figure 1. MVA and MEP pathways in plant terpenoid biosynthesis

The first step of MVA pathway is to take acetyl-CoA as the initial substrate and catalyzed by acetoacetyl-CoA thiolase (AACT). AACT condenses two molecular of acetyl-CoA to form acetoacetyl-CoA. And then, acetoacetyl-CoA is converted to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) under the catalysis of HMG-CoA

synthase (HMGS). HMG-CoA is reduced to MVA by HMG-CoA reductase (HMGR), which is a key enzyme that regulates the MVA pathway flux [17,18]. MVA is then converted to IPP via two successive phosphorylation reactions and one step decarboxylation reaction catalyzed by MVA kinase (MVK), phosphomevalonate kinase (PMVK) and dipshosphomevalonate decarboxylase (MVD), respectively. Finally, IPP is isomerized to DMAPP under IPP isomerase (IPPI) [1].

The initial steps of MEP pathway is to take pyruvate and glyceraldehyde-3-phosphate as the initial substrates and catalyzed by 1-deoxy-D-xylulose-5-phosphate (DXP) synthase (DXS), which is the first rate-limiting enzyme [19]. DXP is then converted to MEP via the intramolecular rearrangement and reduction catalyzed by DXP reductoisomerase (DXR), which is the second key enzyme [20,21]. MEP is converted to 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME), which is catalyzed by 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (MCT). CDP-ME is subsequently converted to 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP) via three successive reactions catalyzed by CDP-ME kinase (CMK), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MECPS) and HMBPP synthase (HDS), respectively. Finally, HMBPP is reduced to a mixture of IPP and DMAPP by HMBPP reductase (HDR), which is the third key enzyme in the MEP pathway [22]. IPP and DMAPP generated by MVA pathway are mainly used for the biosynthesis of farnesyl diphosphate (FPP). IPP and DMAPP formed by MEP pathway are used for geranyl diphosphate (GPP) and geranylgeranyl diphosphate (GGPP) biosynthesis. FPP, GPP and GGPP are synthesized by the corresponding synthase (FPPS, GPPS and GGPPS), respectively [23].

Since the discovery of MVA and MEP pathways in higher plants, most researches have focused on the identification and cloning of MVA and MEP pathway genes, transcriptional and post-transcriptional regulatory mechanisms, synergy between MVA and MEP pathways [23,24]. However, few studies have been devoted to the evolutionary rate variation among genes of MVA and MEP pathways, and the potential factors that contribute to rate variation. Previous studies have shown that the genes present in different metabolic pathway position are subjected to unequal evolutionary forces. For example, the upstream genes evolve more slowly than the downstream genes in the anthocyanin biosynthetic pathway [25-27]. The same evolutionary rate variation pattern also present in the biosynthesis pathway genes of isoflavonoid [28], isoprene and terpenoid phytohormones abscissic acid, gibberellic acid, and brassinosteroids [29,30]. However, study on evolutionary rate variation of phenylpropanoid biosynthetic pathway genes in *Arabidopsis thaliana* [31], gibberellin biosynthetic pathway genes in rice tribe [32], and endosperm starch biosynthetic pathway genes in *Oryza sativa* [33] reveals no correlation between the evolutionary rates and pathway positions. Additionally, pathway properties has been elucidated to affect evolutionary rates. For example, regulatory genes have been reported evolve more rapidly than structural genes [34,35]. Other characteristics of gene such as expression level, multifunctionality, 5'-UTR length, 3'-UTR length, tissue specificity, protein length, gene copy number, domain number, GC3 content, G+C content and codon bias also correlate with the evolutionary rate variation [36-40].

In this study, we investigated the evolutionary rate variation of MVA and MEP pathway genes and two additional genes from 12 representative species. In order to address the following questions. First, whether statistically significant variation in d_N/d_S ratios exists among genes and lineages? Second, whether significant difference in evolutionary rate between MVA and MEP pathway genes? Third, are there exist different evolutionary rate variation pattern in MVA and MEP pathway genes? If there is, whether this variability is explained by differential selective constraint and/or positive selection? Fourth, are there exist evidence of positive selection in any analyzed pathway genes? Finally, whether d_N/d_S ratios varies with any network parameters and/or gene structure?

Results

Sequences retrieval and analyses

A total of 192 protein coding sequences of terpenoid pathway genes were obtained from the 12 species (Additional file 1). The mean coding sequence length of the 16 genes ranged from 714 bp (*MECPS*) to 2217 bp (*HDS*) (Table 1). The ENC and G+C content were similar for the same gene across species and among genes. The average ENC and G+C content of the 16 terpenoid pathway genes were ranged from 47.81 to 55.32, and 0.43 to 0.56, respectively (Table 1). However, the average GC3 content was varied greatly among genes with the average ranging from 0.29 to 0.62 (Table 1). Gene copy number were differed among species for per gene. *Elaeis guineensis*, *Vitis vinifera*, *Morus notabilis* and *Arabidopsis thaliana* had mostly one or two copy each gene. Whereas most *Glycine max* terpenoid pathway genes had more than two copies (Additional file 2). According to the KEGG pathway database, we calculated the PPI value of 16 terpenoid pathway enzymes. The results showed that in addition to, AACT, HMGS, HMGR, MVK and DXS were involved in a number of pathways, the other 11 pathway enzymes only participate in terpenoid pathway (Table 1 and Additional file 3).

Table 1. Summary statistic used in the multivariate analysis of 16 pathway genes

Gene	d_N/d_S	Position	GC3	G+C	ENC	CDS length (bp)	PPI
<i>AACT</i>	0.068 (0.065)	1	0.29	0.46	49.94	405	12
<i>HMGS</i>	0.080 (0.079)	2	0.41	0.44	53.91	469	5
<i>HMGR</i>	0.089 (0.093)	3	0.59	0.54	52.21	580	3
<i>MVK</i>	0.142 (0.139)	4	0.43	0.47	53.89	387	2
<i>PMVK</i>	0.145 (0.147)	5	0.33	0.45	52.31	512	1
<i>MVD</i>	0.135 (0.124)	6	0.42	0.47	54.31	415	1
<i>DXS</i>	0.063 (0.071)	1	0.47	0.50	48.56	720	2
<i>DXR</i>	0.076 (0.082)	2	0.37	0.46	52.18	472	1
<i>MCT</i>	0.193 (0.194)	3	0.39	0.45	52.20	309	1
<i>CMK</i>	0.163 (0.161)	4	0.37	0.44	53.58	392	1

<i>MECPS</i>	0.132 (0.136)	5	0.51	0.52	51.82	238	1
<i>HDS</i>	0.062 (0.060)	6	0.38	0.45	52.34	739	1
<i>HDR</i>	0.082 (0.091)	7	0.47	0.46	55.18	469	1
<i>IPPI</i>	0.142 (0.143)	7/8	0.51	0.49	53.51	282	1
<i>FPPS</i>	0.091 (0.088)	NA	0.44	0.43	55.32	353	1
<i>GGPPS</i>	0.097 (0.115)	NA	0.62	0.56	47.81	373	1

Note: d_N/d_S ratios within parentheses calculated based on 21 branches pairs, d_N/d_S ratios out of parentheses calculated by one-ratio model of PAML. NA represent the path position is not defined.

Substitution models and phylogenetic analysis

A total of 24 nucleotide substitution models were performed for each gene. We determined the best substitution model for the all of 16 terpenoid pathway genes is the GTR+I+G model, with the highest log likelihood (lnL) and lowest AIC value (Additional file 4). The results of substitution saturation showed that there were no any mutational saturation for the 16 terpenoid pathway genes, indicating that the sequences datasets were appropriate for subsequent phylogeny and substitution rate analyses. We concatenated the sequences of the 16 genes with the same order for each species, and then constructed the maximum likelihood species tree with the GTR+I+G model, delineated in figure 2.

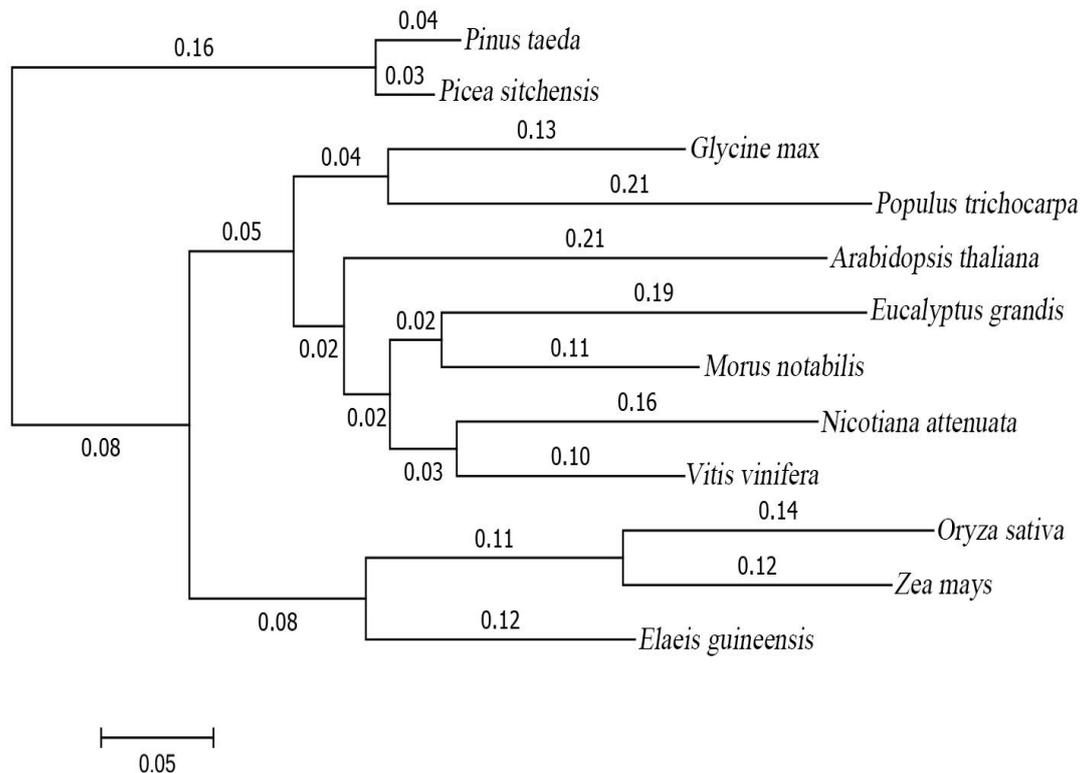


Figure 2. The maximum likelihood tree obtained using combined CDS of 16 terpenoid pathway genes and the GTR+I+G model. Numbers on branches are estimated branch lengths.

Comparing d_N , d_S and d_N/d_S among genes

To compare d_N and d_S among genes, we estimated the d_N and d_S for each branch of the pruned tree of the 12 species. There are total of 22 evolutionary independent branches (Figure 2). The distribution of d_N and d_S value of the 21 branches pairs estimated by Codeml of the PAML v4.9 for MVA and MEP pathways genes and the average of d_N and d_S for each gene are shown in Figure 3. Unlike d_S value, d_N varied significantly among genes ($P < 0.01$), with the highest d_N (0.133 for *GGPPS*) being 5.5 times the lowest (0.024 for *HDS*). Both d_N and d_S , there was no significant difference between MVA and MEP pathways genes. For MVA pathway, the most upstream gene *AACT*

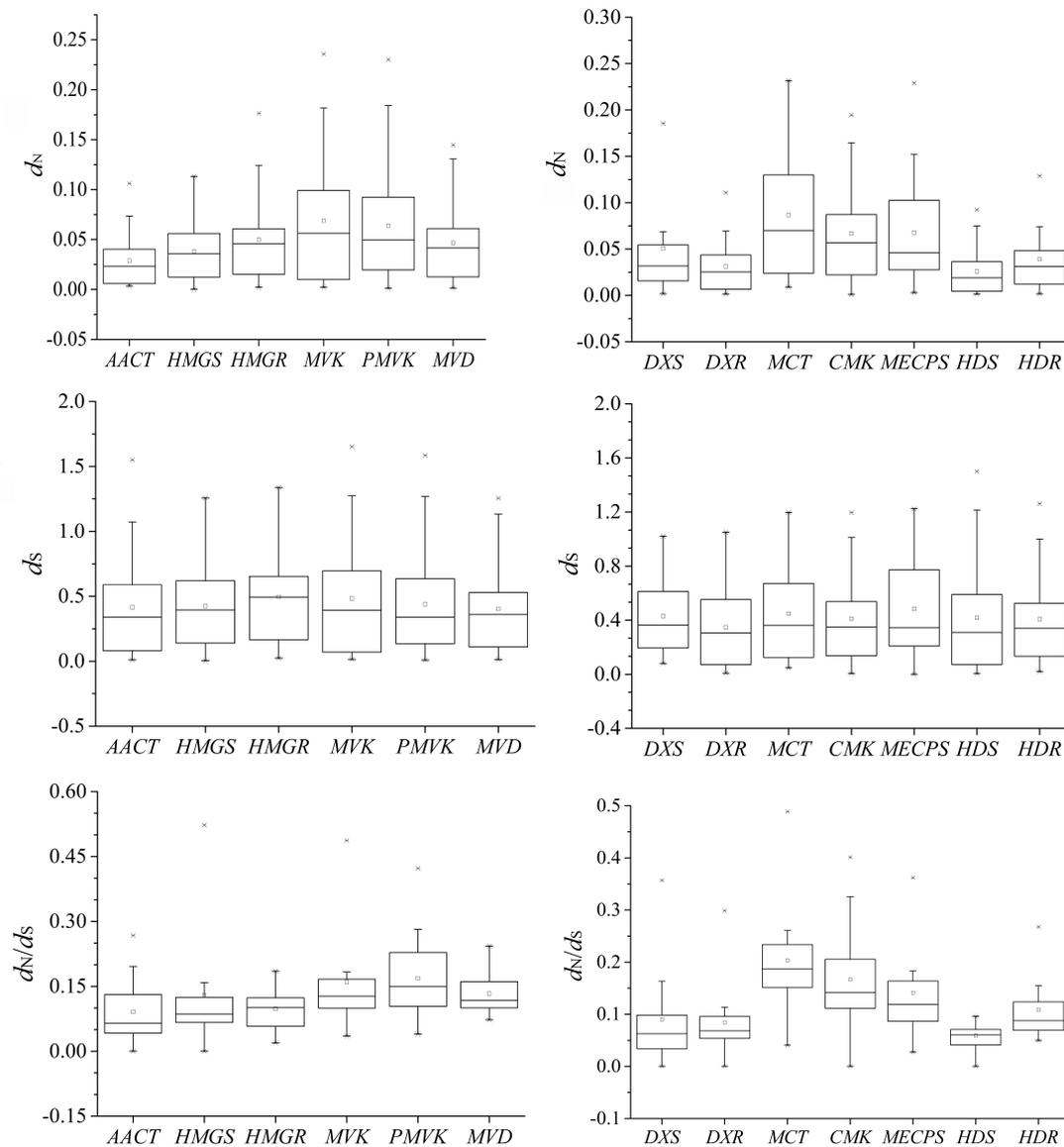


Figure 3. The evolutionary parameters of 21 branches pairs of MVA and MEP pathways genes. Average values of d_N , d_S and d_N/d_S ratio are represented by horizontal bars within the boxes at 95% confidence intervals.

exhibits the lowest d_N value (*IPPI* gene not included in the analysis of MVA and MEP pathway, the same below), which is varied significantly from that of all other genes ($P < 0.05$, with Bonferroni correction). However, compared with the several early steps genes (*HMGR*, *MVK* and *PMVK*), the downstream genes *MVD*, even including *FPPS* gene ($d_N = 0.037$), did not show higher d_N values. For MEP pathway, both d_N and d_S , there were no differ significantly between the early and late steps genes. However, three midstream gene *MCT*, *CMK*, *MECPS* of MEP pathway and the downstream gene *GGPPS* ($d_N = 0.133$) of MEP pathway, exhibits the higher d_N values than other MEP pathway genes.

The d_N/d_S ratios estimated using the one-ratio model of CodeML are similar to the values of ratios calculated using the average nonsynonymous and synonymous substitution rates of 16 terpenoid pathway genes (Table 1). The significant difference comparisons showed that the d_N/d_S ratios varied significantly among the 16 genes ($P < 0.001$) with the highest value (0.193 for *MCT*) being 3.1 times the lowest (0.062 for *HDS*). However, there was no significant difference in d_N/d_S ratios between MVA and MEP pathways genes. For MVA pathway, according to compared pairwise statistically results, the downstream genes (*MVK*, *PMVK* and *MVD*) exhibit a significantly greater d_N/d_S ratios than the upstream genes (*AACT*, *HMGS* and *HMGR*) at level of $P < 0.05$. To MEP pathway, the downstream genes (*HDS* and *HDR*) did not exhibit a higher d_N/d_S ratios than the upstream genes (*DXS* and *DXR*), but the midstream genes (*MCT*, *CMK* and *MECPS*) exhibit a significantly higher d_N/d_S ratios than other MEP pathway genes. The results of two-ratios model with monocots, dicots, gymnosperms and each species as foreground branch showed that such evolutionary rate variation patterns of MVA and MEP pathways genes also exist in each individual species or lineages (Additional file 5). Additionally, we also used the free-ratio model to test whether the d_N/d_S ratios of the terpenoid pathway genes differed among lineages by comparing the likelihood values with one-ratio model. The LRT results showed that except for *MVK*, *MVD* and *HDS* genes, twice the log likelihood values difference ($2\Delta l$) of the other 13 genes were all greater than 31.41 ($P < 0.05$, under Chi-square test with $df = 20$), indicating that the d_N/d_S ratios of these 13 genes are different among lineages.

Detecting positive selection

In this study, we applied the two-ratios model and random-sites models to detect whether there were positive selection occurred on species-specific or lineage-specific branches and codon sites of terpenoid pathway genes, respectively. Two-ratios model results suggested that for all 16 genes only *MECPS* gene detected positive selection signature ($\omega_1 = 9.908$) in branch *pinus taeda*. The LRT statistic for two-ratios model to one-ratio model is $2\Delta l = 2 \times 5.11 = 10.22$, with $P = 0.0013$ and $df = 1$, indicating that the d_N/d_S ratio of *MECPS* gene for branch *pinus taeda* is significantly different for other branches. To test whether ω_1 is significantly greater than 1, we reestimate the log likelihood value of the two-ratios model with $\omega_1 = 1$ fixed [59], calculating the $\ln L$ value is -6904.03. The LRT statistic for this comparison is $2\Delta l = 2 \times 5.01 = 10.02$, with $P = 0.0015$ and $df = 1$, suggesting that ω_1 is significantly higher than 1 at 1% significance level.

Six random sites models were used to assume various classes of d_N/d_S ratios among the codon sites to the 16 terpenoid pathway genes separately. First, we compared models M3 and M0 by LRT to test whether d_N/d_S ratios is significantly different among sites. For all of 16 genes, the LRTs comparing M3 and M0 has the statistic $2\Delta l = 345.37 \sim 1674.48$ much greater than the Chi-square critical values 13.28 at 1% level with $df = 4$. The results showed extreme variation in d_N/d_S ratios among codon sites. Second, we compared models M2a and M1a, models M8 and M7 to detect the positive selection sites. Comparison between M2a and M1a results suggested that, for all 16 genes, the LRTs statistic are $2\Delta l = 0.00 \sim 1.22$, significantly lower than a χ^2 significance value 5.99 at 5% level with $df = 2$. Indicating that models M2a were not significantly fit the data than the null models M1a. By contrast, LRTs ($2\Delta l = 6.32 \sim 120.58$, $0.001 < P < 0.05$, with $df = 2$) for the comparison of M8 and M7 were statistically significant for 14 of the 16 genes (except for *MVD* and *IPPI*), and identified several sites under positive selection for 7 of terpenoid pathway genes (Table 2). Interestingly, 5 of the 7 genes appear in the MEP pathway, while only 2 genes appear in the MVA pathway and its downstream.

Table 2. Log-likelihood values and parameter estimates under M7 and M8 models for the 7 positively selected genes

Gene	Model	lnL	Estimates of Parameters	Positively Selected Sites
<i>DXS</i>	M7	-15372.40	$p = 0.185, q = 2.297$	Not allowed
	M8	-15365.67	$P_0 = 0.984, p = 0.201, q = 3.248$ ($P_1 = 0.016$), $\omega = 1.012$	10M ($P = 0.96$) 20R ($P = 0.95$)
<i>DXR</i>	M7	-9954.16	$p = 0.665, q = 5.729$	Not allowed
	M8	-9893.87	$P_0 = 0.938, p = 1.386, q = 23.873$ ($P_1 = 0.062$), $\omega = 1.135$	22S ($P = 0.96$) 24K ($P = 0.96$)
<i>MCT</i>	M7	-7071.57	$p = 0.159, q = 0.403$	Not allowed
	M8	-7048.18	$P_0 = 0.763, p = 0.340, q = 4.270$ ($P_1 = 0.237$), $\omega = 1.272$	6K, 22I, 25C, 26L, 29R, 35G, 36N, 39V, 44K, 49K ($0.96 < P < 0.99$)
<i>CMK</i>	M7	-6244.34	$p = 0.223, q = 0.859$	Not allowed
	M8	-6236.92	$P_0 = 0.935, p = 0.286, q = 1.693$ ($P_1 = 0.065$), $\omega = 1.465$	18S, 21S, 28S ($0.95 < P < 0.97$)
<i>HDS</i>	M7	-13344.40	$p = 0.245, q = 2.834$	Not allowed
	M8	-13340.91	$P_0 = 0.987, p = 0.275, q = 3.802$ ($P_1 = 0.013$), $\omega = 1.236$	39A ($P = 0.98$)
<i>AACT</i>	M7	-7704.92	$p = 0.274, q = 2.793$	Not allowed
	M8	-7699.14	$P_0 = 0.986, p = 0.319, q = 4.018$ ($P_1 = 0.014$), $\omega = 1.082$	174N ($P = 0.99$)
<i>FPP</i>	M7	-7210.03	$p = 0.304, q = 2.355$	Not allowed
	M8	-7206.88	$P_0 = 0.997, p = 0.313, q = 2.543$ ($P_1 = 0.003$), $\omega = 1.675$	20A ($P = 0.96$)

Note: P is the value of posterior probabilities calculated by BEB.

Multivariate analyses

To determine whether the evolutionary rate variation of terpenoid pathway genes were affected by network structure and intramolecular characteristics, Spearman's rank correlation were used to estimate the correlations between these factors and the evolutionary parameters (d_N and d_N/d_S) estimated by one-ratio model of 16 terpenoid pathway genes. As shown in Figure 4, there were no significant correlations between d_N/d_S ratios and other factors, except for with CDS length ($\rho = -0.592$, $P = 0.016$, with FDR correction, the same below). Similarly, the d_N values was only negatively correlated with CDS length ($\rho = -0.559$, $P = 0.024$). when conducted correlation analyses for MVA and MEP pathways genes respectively, the correlations between CDS length and d_N/d_S ($\rho = -0.893$, $P = 0.012$), CDS length and d_N ($\rho = -0.857$, $P = 0.023$) of MEP pathway genes were remained significant, while those correlations about MVA pathway genes were disappeared (d_N : $\rho = 0.028$, $P = 0.726$; d_N/d_S : $\rho = 0.086$, $P = 0.919$). However, a significant correlation between pathway position and d_N/d_S ($\rho = 0.829$, $P = 0.048$), PPI and d_N/d_S ($\rho = 0.898$, $P = 0.015$) of MVA pathway genes were observed.

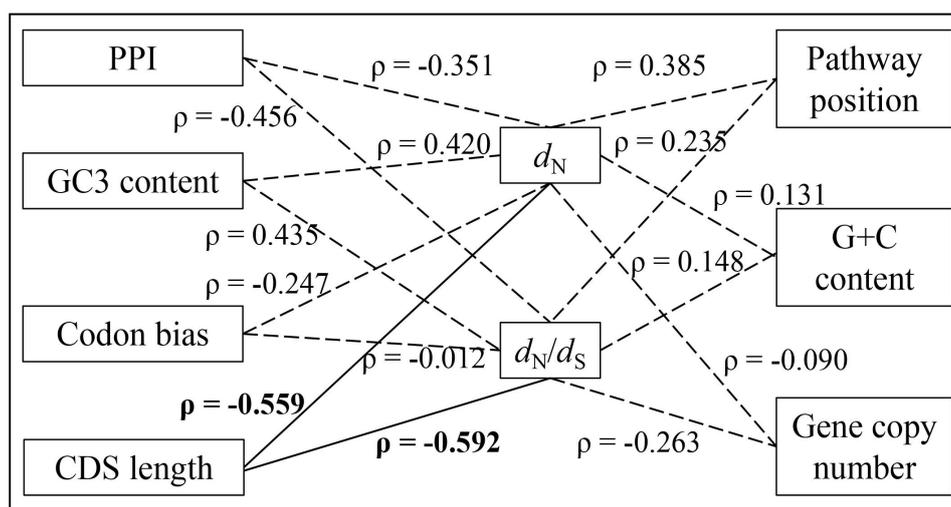


Figure 4. Spearman's rank correlation coefficient of d_N , d_N/d_S and pathway position, G+C, GC3, copy number, codon bias, PPI and CDS length. Solid and dotted lines represent significant and non-significant, respectively.

Since pathway position and PPI are interrelated and are each correlated with d_N/d_S ratios of MVA pathway genes, the partial correlation analysis were performed to estimate the relative contributions of these two factors. Partial correlation analysis showed that when controlling for pathway position, the correlation between PPI and d_N/d_S remained significant ($\rho = 0.672$, $P = 0.036$). However, when controlling for PPI, the significant correlations between pathway position and d_N/d_S were disappeared ($\rho = 0.527$, $P = 0.143$). These results suggest that PPI is an important factor affecting the evolutionary rate variation of MVA pathway genes.

Discussion

In previous research, Ramsay et al [30] have elucidated the evolutionary rate patterns of genes involved in the biosynthesis pathway of four terpenoids: lutein, abscissic acid, brassinosteroid and gibberellic acid. For this study, the fourteen MVA and MEP pathway genes and its downstream genes *FPFS* and *GGPPS* of 12 representative species with a large taxonomic span were used for the analysis of evolutionary rate variation. The results revealed that with a broad taxonomic span including monocots, dicots and gymnosperm, the d_N values and d_N/d_S ratios were varied significantly among the 16 terpenoid pathway genes. However, with the similar d_S values among 16 genes, suggesting that the differential mutation rates cannot explain the evolutionary rate variation of 16 terpenoid pathway genes [25]. In fact, the variation among genes in substitution rates is ubiquitous. Several other previous studies conducted on various biological pathways such as carotenoid pathway [60], anthocyanin pathway [25-27], isoflavonoid pathway [61], starch pathway [33], phenylpropanoid pathway [31], isoprene pathway [29] and gibberellin pathway [32] have demonstrated it. Unlike other precursors, the precursor IPP of terpenoids can be generated by MVA and MEP pathway, respectively. Interestingly, our analysis suggests that MVA and MEP pathway genes have different evolutionary rate variation patterns, although there were no significant difference in d_N/d_S ratios between two pathways genes. For MVA pathway, the downstream genes evolved more rapidly than the upstream genes. This pattern are similar to the anthocyanin biosynthetic pathway genes [25,26] and isoflavonoid biosynthetic pathway genes [61]. One possible explanation for this pattern is that the upstream enzymes are subject to greater selective constraint than downstream enzymes, because of upstream enzymes exert more control over metabolic fluxes, and influence more end products than downstream enzymes [26,62]. First, in our study the evolutionary parameters d_N and d_N/d_S ratios are much lower for the upstream genes than for the downstream genes, which means that purifying selection is stronger on upstream genes. Second, KEGG pathway analysis showed that the most two upstream enzymes (*AACT* and *HMGS*) of MVA pathway involved in 12 and 5 KEGG pathways, respectively. In contrast, the most two downstream enzymes (*PMVK* and *MD*) are only participated in the terpenoid backbone biosynthesis pathway (ko:00900). In addition, subsequent correlation analysis also indicated that the d_N/d_S ratios of MVA pathway genes were significant correlation with pathway position and PPI. An alternative explanation for this pattern is that downstream enzymes undergone more frequent position selection [27]. However, in this study neither the two-ratios branch model nor the two pairs random sites models detected any signature of positive selection for any species-specific branch or lineage-specific branch and codon sites of MVA pathway genes (except for *AACT* gene has one site under positive selection). Consequently, we demonstrate that the higher d_N/d_S ratios of the downstream genes in the MVA pathway is mainly put down to the relaxation of selective constraint rather than subject to more frequent positive selection. It should be noted, however, that the power to detect positive selection with the above methods is limited, particularly when data from across the taxonomic range of species are

pooled together to analysis [63,64]. So we cannot completely rule out the effect of positive selection on the accelerated evolution of downstream genes in the MVA pathway. Some previous studies have provided the evidence that positive selection is responsible for accelerated rates of evolution [65-67].

In this study, the MEP pathway genes did not exhibit the same evolutionary rate variation pattern as MVA pathway genes, anthocyanin biosynthetic pathway genes and carotenoid biosynthetic pathway genes [25,60]. In the MEP pathway, the downstream genes *HDS* and *HDR* did not show a higher d_N/d_S ratios than upstream genes *DXS* and *DXR*. In contrast, three midstream genes *MCT*, *CMK* and *MECPS* exhibit greater d_N/d_S ratios than other MEP pathway genes. Subsequent correlation analysis revealed that there were no significant correlation between d_N/d_S ratios and pathway position, d_N/d_S ratios and PPI. This evolutionary rate variation of MEP pathway genes seems better explained by the pathway fluxes hypothesis, which illuminated that in biosynthetic and metabolic pathways, natural selection more likely target enzymes that control flux, and therefore trend to experience greater evolutionary constraints [68-70]. In MEP pathway, the early step enzymes *DXS* and *DXR*, and the last step enzyme *HDR* with lower d_N/d_S ratios are considered as key enzymes [1], consistent with the fluxes theory. We also examined this evolutionary rate variation could be explained by positive selection. Surprisingly, most of MEP pathway genes were detected with positive selection signature (Table 2), especially for the *MCT* gene with the highest d_N/d_S ratios, there were 10 positive selection sites with *P*-value of BEB greater than 0.95, and for the *MECPS* gene, which detected d_N/d_S ratio > 1 in branch *pinus taeda* under two-ratios model. In addition, 3 positive selection sites were also detected in the midstream gene *CMK*. In conclusion, the three midstream genes with higher d_N/d_S ratios were all detected stronger signature of positive selection. Consequently, we speculated that the evolutionary rate variation of MEP pathway genes is mainly attributed to the differential constraint and positive selection.

Several previous studies have demonstrated that the evolutionary rate of genes within networks and pathways are also affected by multiple network parameters and gene structure [36-40]. For this study, the Spearman's rank correlation test was conducted between the pathway position, PPI, GC3 content, G+C content, codon bias, gene copy number, CDS length and evolutionary rates of 16 terpenoid pathway genes. The results showed that the d_N/d_S ratios of all 16 genes were negatively correlated with CDS length. Several prior studies have also provided evidence that protein length play an key role in shaping the evolutionary rate of TLR signaling pathway genes of Suidae [71], HOG-MAPK pathway genes of *Saccharomyces cerevisiae* [72] and orthologs between *Populus tremula* and *Populus trichocarpa* [73]. While no significantly correlated with other factors. Similar to Ramsay et al [30] reported that no apparent variation in d_N/d_S ratios with gene copy number. Contrary to the significant correlations observed in paralogs between *Arachis duranensis* and *Arachis ipaënsis* (d_N/d_S ratios negatively correlated with GC3 content) [74], orthologs between *Brassica rapa* and *Brassica oleracea* (d_N/d_S ratios significantly correlated with G+C content and codon bias) [75] and pathway genes of converting glucose to the ABA,

GA and brassinosteroids (d_N/d_S ratios correlate with pathway position) [30]. When performed Spearman's rank correlation test to MVA and MEP pathway genes, respectively. The MEP pathway genes remained significantly correlated with CDS length, while the correlation with MVA pathway genes disappeared, suggesting that protein length have a greater affects on MEP pathway genes than MVA pathway genes. In addition, two new significant correlations were observed between d_N/d_S ratios and pathway position, d_N/d_S ratios and PPI. Therewith, we boldly suggested that the MVA pathway genes are subjected to stronger selective constraint than MEP pathway genes. However, limited to the identified of MVA and MEP pathways genes of plants, only a small dataset was used in this study. With the publication of more plants genome data, and the improvement of assembly level, a large dataset should be used for future analysis to provide stronger evidence for the evolutionary rate variation patterns of MVA and MEP pathways genes.

Conclusion

In this study, we have detected the d_N and d_N/d_S ratios were varied significantly among the 16 terpenoid pathway genes. However, there were no significant difference in d_N and d_N/d_S ratios between MVA and MEP pathways genes. The MVA and MEP pathways genes have different evolutionary rate variation patterns and attributed to differential selective constraint, selective constraint relaxation and positive selection, respectively. In addition, *MECPS* gene was detected d_N/d_S ratios >1 in branch *Pinus taeda* under two-ratios model, and most MEP pathway genes were detected with positive selection signature under M8 sites model. Furthermore, Spearman's rank correlation test results showed that CDS length were negatively correlated with the d_N/d_S ratios of all 16 genes and MEP pathway genes but MVA pathway genes, pathway position and PPI were significantly correlated with d_N/d_S ratios of MVA pathway genes. Our results will contribute to improving the understanding of natural selection in pathways.

Materials and methods

Species and genes analyzed

Twelve representative plants were selected from the species that could be retrieved all of 14 MVA and MEP pathways genes, and the other 2 extra genes *FPPS* and *GGPPS* (Figure.1). We retrieved coding sequence (CDS) of the 16 terpenoid pathway genes of *Picea sitchensis*, *Elaeis guineensis*, *Zea mays*, *Oryza sativa*, *Vitis vinifera*, *Nicotiana attenuate*, *Morus notabilis*, *Eucalyptus grandis*, *Arabidopsis thaliana*, *Populus trichocarpa* and *Glycine max* from NCBI or corresponding species genome database. The CDS of 16 terpenoid pathway genes of *Pinus taeda* were retrieved from the full-length transcriptome sequencing database (unpublished data). The retrieval methods refer to the research of Li et al [41] and Han et al. [42]. Non-redundant NCBI blast and genomic blast search were performed using the CDS of 16 terpenoid pathway genes of *Arabidopsis thaliana* as query sequences, and then retrieve orthologs sequences based on bidirectional blast-hit (BDBH) algorithm.

In *Arabidopsis thaliana*, several terpenoid pathway genes occur in multigene families, such as *AACT*, *HMGR*, *MVD*, *DXS*, *IPPI*, *FPPS* and *GGPPS* genes. However, *HMGS*, *MVK*, *PMVK* and other MEP pathway genes are known only as a single copy in *Arabidopsis thaliana*. For this, we selected *AtAACT2* as the query sequence, which is six times more efficient than *AtAACT1* and has an essential role in isoprenoid biosynthesis [1,43], and *AtHMGR1* as query sequence, which is generally more highly expressed than *AtHMGR2* [44-46], and *AtMVD1* as query sequence, which is associated with the peroxisomes in vivo [47,48], and *AtDXS1* as query sequence which encodes a functional DXS enzymes [49], and *AtIPPI1* as query sequence which has overlapping functions in isoprenoid biosynthesis [50], and *AtFPPS1* as query sequence which encodes two protein in the cytosol [51], and *AtGGPP1* as query sequence, which is ubiquitously expressed in all organs [52].

Statistical analyses

To get a better understanding of factors that contribute to variation in evolutionary rate among terpenoid pathway genes, we counted the CDS length, pathway position, GC3 content, G+C content, gene copy number, pathway pleiotropy index (PPI) which measured using the pathway number of protein participate in and codon usage bias which evaluated by effective number of codons (ENC) of terpenoid pathway genes among each species. PPI were obtained from KEGG pathway database. ENC and G+C content were calculated with DnaSP 6.0 [53]. Pathway position were defined by the method that enzymes of MVA and MEP were sequentially numbered from DXS to IPP (1-8) and AACT to IPP (1-7), respectively. Gene copy number were obtained using the method described in Chu et al [28]. First, we conducted Spearman's rank correlations for the bivariate correlation analysis between these factors and evolutionary parameters (d_N and d_N/d_S) estimated by one-ratio model. And then, the partial correlation analysis were performed to estimate the relative contributions of the correlated factors. All of these statistical analyses were performed by applying statistical package R <http://www.r-project.org>.

Multiple sequence alignment and phylogenetic analyses

Multiple sequence align codons were performed using ClustalW program of MEGA 7.0 with default parameters [54], and then delete stop codons and refined manually. Finally saved the aligned sequences files as PHYLIP and FASTA format for subsequent analyses. The sequences substitution saturation were detected using DAMBE 5.0 software [55]. The relative suitabilities of 24 nucleotide substitution models were evaluated for the 16 terpenoid pathway genes by the JmodelTest software, respectively. We selected the best substitution model based on Akaike Information Criterion (AIC), the lower the AIC value, the better the substitution model. The maximum likelihood phylogenetic tree of the 12 species were constructed based on combined CDS of 16 terpenoid pathway genes and the best substitution model by MEGA 7.0 with the default options.

Comparing d_N , d_S and d_N/d_S among genes

The pairwise nonsynonymous and synonymous substitution rates (d_N and d_S) of 12 species were estimated for the CDS of 16 terpenoid pathway genes using the CodelML program in PAML version 4.9 [56], with the parameters described in Yang and Nielsen [57]. We then used the method depicted by Yang et al [32] to detect whether the d_N or d_S differed among genes. We used the one-ratio model of CodelML program of PAML 4.9 to estimate a single value of d_N/d_S for the entire tree along each gene. we used the free-ratio model to assumes an independent d_N/d_S ratio for each branch of the tree along each gene. However, in consideration of the rich parameters need to be estimated in free-ratio model maybe increase the bias in the estimates, with a concomitant decrease in the variance. We also used two-ratios model to assumes the d_N/d_S for each branch in the phylogeny along each gene. The significance of differences in d_N/d_S ratio among genes was analyzed using the method described in the research of Lu and Ransher [26]. We conducted likelihood ratio test (LRT) to detect whether the free-ratio and two-ratios model fits the data significantly better than the one-ratio model.

Detection of positive selection

The nonsynonymous to synonymous substitution rate ratio ($\omega = d_N/d_S$) provides a practical measure of natural selection at the gene level, with $\omega > 1$, $\omega = 1$ and $\omega < 1$ indicating positive selection, neutral evolution and purifying selection, respectively. In this study, except for using the two-ratios model to examine whether there existed a few specific-branches with the d_N/d_S ratio significantly greater than 1. We also applied six random sites models to detect whether specific sites are subject to positive selection, which estimate variable d_N/d_S ratio among sites, but no variation among lineages. Model M0 assumes a single ω value for all sites. M1a (neutral) assumes the conserved sites with $\omega_0 < 1$ and the neutral sites with $\omega_1 = 1$. Compared to M1a, model Ma2a allow a free ω_2 value estimated from the data. M3 (discrete) assumes three site classes. M7 (beta) assumes a beta distribution of ω over sites. M8 (beta & ω) allow an extra site class with a free ω ratio estimated from the data. We then examined for a site class with $\omega > 1$ by comparing likelihood values under M2a vs. M1a and M8 vs. M7. The Bayes Empirical Bayes (BEB) analysis was used to calculate the posterior probabilities of positive selection sites [58]. We also detected for variable evolutionary rates among sites by comparing likelihood value under M3 and M0.

Abbreviations

IPP: isopentenyl diphosphate; DMAPP: dimethylallyl diphosphate; MVA: mevalonate
MEP: 2-C-methyl-D-erythritol-4-phosphate; AACT: acetoacetyl-CoA thiolase;
HMG-CoA: 3-hydroxy-3-methylglutaryl-CoA; HMGS:3-hydroxy-3-methylglutaryl
-CoA synthase; HMGR: 3-hydroxy-3-methylglutaryl-CoA reductase; MVK:
mevalonate kinase; PMVK: phosphomevalonate kinase; IPPI: isopentenyl
diphosphate isomerase; DXP: 1-deoxy-D-xylulose-5-phosphate; DXS: 1-deoxy-D
-xylulose-5-phosphate synthase; DXR: 1-deoxy-D-xylulose-5-phosphate
reductoisomerase; CDP-ME: 4-diphosphocytidyl-2-C-methyl-D-erythritol; MCT:

2-C-methyl-D-erythritol 4-phosphate cytidyltransferase; HMBPP: 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate; CMK: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MECPS: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; HDS: 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase; HDR: 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase; FPP: farnesyl diphosphate; GPP: geranyl diphosphate; GGPP: geranylgeranyl diphosphate; AIC: Akaike Information Criterion; LRT: likelihood ratio test; BEB: Bayes Empirical Bayes; ENC: effective number of codons; CDS: coding sequence.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information. **Additional file 1: Table S1.** The coding sequence information of 16 terpenoid pathway genes. **Additional file 2: Table S2.** Gene copy number of terpenoid pathway genes among species for per gene. **Additional file 3: Table S3.** Summary statistic participated in multiple pathways of the five terpenoid pathway genes. **Additional file 4: Table S4.** Comparison of 24 nucleotide substitution modles for each gene. **Additional file 5: Table S5.** The d_N/d_S ratios of each foreground under two-ratios model.

Competing interests

The authors declare that they have no competing interests.

Finding

This work was funded by the “948” program of State Forestry Administration P.R. China (2014-4-72) and National Key R&D Program of China (2017YFD0600502-3).

Authors' contributions

MJP and HSW conceived and designed the research and drafted the manuscript together. MJP, HLW and HJ performed the research and analyzed the data. LTY participated in the data analysis. All authors have read and approved the final manuscripts.

Acknowledgements

We thank xin-sheng Hu of the department of College of Forestry and Landscape Architecture University for his guidance on PAML 4.9 software.

References

1. Vranová E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*, 2013, 64: 23.1–23.36.
2. Muhlemann JK, Klempien A, Dudareva N. Floral volatiles: from biosynthesis to function. *Plant Cell Environment*, 2014, 37: 1936-1949.
3. Brokl M, Fauconnier ML, Benini C, et al. Improvement of ylang-ylang essential oil characterization by GC×GC-TOFMS. *Molecules*, 2013, 18: 1783-1797.
4. Dudareva N, Klempien A, Muhlemann JK, et al. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist*, 2013, 198: 16-32.
5. Yadav PA, Suresh G, Gao MSA, et al. Limonoids from the leaves of *Soymida febrifuga* and their insect antifeedant activities. *Bioorganic and Medicinal Chemistry Letters*, 2014, 24(3): 888-892.
6. Abbas F, Ke YG, Yu R, et al. Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. *Planta*, 2017, 246(5): 803-816.
7. Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds?. *New Phytologist*, 2018, 220: 692-702.
8. Heiling S, Schuman MC, Schoettner M, et al. Jasmonate and ppsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell*, 2010, 22: 273-292.
9. Hijaz F, Nehela Y, Killiny N. Possible role of plant volatiles in tolerance against huang long bing in citrus. *Plant Signal Behavior*, 2016, 11(3): 1-12.
10. Berteaux CM, Freije JR, Woude HVD. Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*. *Planta Medica*, 2005, 71: 40-47.
11. Schwab W, Fuchs C, Huang FC. Transformation of terpenes into fine chemicals. *European Journal of Lipid Science*, 2013, 115: 3-8.
12. Cheng AX, Luo YG, Mao YB, et al. Plant terpenoids: biosynthesis and ecological functions. *Journal of Integrative Plant Biology*, 2007, 49(2): 179-186.
13. Bohlmann J, Zülke KG. Terpenoid biosynthesis and specialized vascular cells of conifer defense. *Journal of Integrative Plant Biology*, 2010, 52(1): 86-97.
14. Wilding EI, Brown JR, Bryant AP. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate in Gram-positive Cocci. *Journal of Bacteriology*, 2000, 182(15): 4319-4327.
15. Rodríguez C. Early steps in isoprenoid biosynthesis: Multilevel regulation of the supply of common precursors in plant cells. *Phytochemistry Review*, 2006, 5: 1-15.
16. Lichtenthaler HK. The deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology & Plant Molecular*, 1999, 50: 47-65.
17. Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature*, 1990, 343(6257): 425-430.
18. Akhtar N, Gupta P, Sangwan NS. Cloning and functional characterization of 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from *Withania somnifera*: an important medicinal plant. *Protoplasma*, 2013, 250(2): 613-622.

19. Walter MH, Hans J, Strack D. Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid accumulating mycorrhizal roots. *Plant Journal*, 2002, 31: 243–254.
20. Morrone D, Lowry L, Determan MK, et al. Increasing diterpene yield with a modular metabolic engineering system in *E. coli* comparison of MVA and MEP isoprenoid precursor pathway engineering. *Applied Microbiology and Biotechnology*, 2010, 85(6): 1893-1906.
21. Chang K, Qiu F, Chen M, et al. Engineering the MEP pathway enhanced ajmalicine biosynthesis. *Biotechnology and Applied Biochemistry*, 2014, 61(3): 249-255.
22. Ma D, Li G, Zhu Y, et al. Overexpression and suppression of artemisia annual 4-hydroxy-3-methylbutenyl diphosphate reductase 1 gene (*Aoids1*) differentially regulate artemisinin and terpenoid biosynthesis. *Frontiers in Plant Science*, 2017, 8: 24-36.
23. Dubey VS, Bhalla R, Luthra R. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences*, 2003, 28(5): 637-646.
24. Eisenreich W, Rohdich F, Bacher A. Deoxyxylulose phosphate pathway to terpenoids. *Trends Plant Science*. 2001, 6: 78–84.
25. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution*, 1999, 16(2): 266-274.
26. Lu YQ, Rausher MD. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution*, 2003, 20(11): 1844-1853.
27. Rausher MD, Lu Y, Meyer K. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *Journal of Molecular Evolution*, 2008, 67(2):137-144
28. Chu S, Wang J, Cheng H, et al. Evolutionary study of the isoflavonoid pathway based on multiple copies analysis in soybean. *BMC Genetics*, 2014, 15(1): 76-76.
29. Sharkey TD, Yeh S, Wiberley AE, et al. Evolution of the isoprene biosynthetic pathway in Kudzu. *Plant Physiology*, 2005, 137: 700-712.
30. Ramsay H, Rieseberg LH, Ritland K. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Molecular Biology and Evolution*, 2009, 26(5): 1045-1053.
31. Ramos-Onsins SE, Puerma E, Salguero D, et al. Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Molecular Ecology*, 2008, 17: 1211-1223.
32. Yang YH, Zhang FM, Ge S. Evolutionary rate patterns of the Gibberellin pathway genes. *BMC Evolutionary Biology*, 2009, 9(1): 206-218.
33. Yu GQ, Olaen KM, Schaal BA. Molecular evolution of the endosperm starch synthesis pathway genes in rice (*Oryza sativa* L.) and its wild ancestor, *O. rufipogon* L.. *Molecular Biology and Evolution*, 2011, 28(1): 659-671.
34. Tucker PK, Lundrigan BL. Rapid evolution of the sex-determining loci in Old World mice and rats. *Nature*, 1993, 364: 715-717.
35. Purugganan MD, Wessler SR. Molecular evolution of the plant R regulatory gene family. *Genetics*, 1994,138: 849–854.
36. Kitami T, Nadeau JH. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genetics*, 2002,

32: 191-194.

37. Jordan IK, Marino-Ramirez L, Wolf YI, et al. Conservation and coevolution in the scale-free human gene coexpression network. *Molecular Biology and Evolution*, 2004, 21: 2058–2070.
38. Makino T, Suzuki Y, Gojobori T. Differential evolutionary rates of duplicated genes in protein interaction network. *Gene*, 2006, 385: 57-63.
39. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. *Genome biology*, 2006, 7(5): R39.
40. Song H, Sun J, Yang G. Comparative analysis of selection mode reveals different evolutionary rate and expression pattern in *Arachis duranensis* and *Arachis ipaënsis* duplicated genes. *Plant Molecular Biology*, 2018, 98: 349-361.
41. Li C, Li QG, Jim MD, et al. Divergent evolutionary pattern of starch biosynthetic pathway genes in grasses and dicots. *Molecular Biology and Evolution*, 2012, 29(10): 3227-3236.
42. Han M, Qin S, Song XJ, et al. Evolutionary rate patterns of genes involved in the Drosophila Toll and Imd signaling pathway. *BMC Evolutionary Biology*, 2013, 13: 245-255.
43. Jin H, Song Z, Nikolau BJ. Reverse genetic characterization of two paralogous acetoacetyl CoA thiolase genes in *Arabidopsis* reveals their importance in plant growth and development. *Plant Journal*. 2012, 70: 1015–1032.
44. Enjuto M, Balcells L, Campos N, et al. *Arabidopsis thaliana* contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. *Proceedings of the National Academy of Sciences USA*, 1994, 91: 927–931.
45. Enjuto M, Lumbreras V, Marin C, et al. Expression of the *Arabidopsis HMG2* gene, encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase, is restricted to meristematic and floral tissues. *Plant Cell*, 1995, 7: 517–527.
46. Lumbreras V, Campos N, Boronat A. The use of an alternative promoter in the *Arabidopsis thaliana HMG1* gene generates an mRNA that encodes a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase isoform with an extended N-terminal region. *Plant Journal*, 1995, 8: 541–549.
47. Simkin A, Guirimand G, Papon N, et al. Peroxisomal localisation of the final steps of the mevalonic acid pathway in plants. *Planta*, 2011, 234: 903–914.
48. Guirimand G, Guihur A, Phillips M, et al. A single gene encodes isopentenyl diphosphate isomerase isoforms targeted to plastids, mitochondria and peroxisomes in *Catharanthus roseus*. *Plant Molecular Biology*, 2012, 79: 443–459.
49. Estevez JM, Cantero A, Romero C, et al. Analysis of the expression of *CLAI*, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol-4-phosphate pathway in *Arabidopsis*. *Plant Physiology*, 2000, 124: 95–104.
50. Phillips MA, D’Auria JC, Gershenzon J, et al. The *Arabidopsis thaliana* type I isopentenyl diphosphate isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. *Plant Cell*, 2008, 20: 677–697.
51. Leivar P, Gonzalez VM, Castel S, et al. Subcellular localization of *Arabidopsis* 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Plant Physiology*, 2005, 137: 57–69.
52. Okada K, Saito T, Nakagawa T, et al. Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in *Arabidopsis*. *Plant Physiology*. 2000, 122: 1045–1056.
53. Julio R, Albert FM, Carlos BJ, et al. DnaSP 6: DNA sequence polymorphism analysis of

- large data set. *Molecular Biology and Evolution*, 2017, (12):12.
54. Sudhir K, Glen S, Koichiro T. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 2014, 33: 1870–1874.
 55. Xia XH. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution, *Molecular Biology Evolution*, 2013, 30(7): 1720–1728.
 56. Yang Z. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 2007, 24: 1586-1591.
 57. Yang ZH, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 2000, 17(1): 32–43.
 58. Anisimova M, Bielawski JP, Yang ZH. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution*, 2002, 19(6): 950-958.
 59. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology Evolution*, 1998, 15:568–573.
 60. Livingstone K, Anderson S. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *Journal of Heredity*, 2009, 100: 754–761.
 61. Chu SS, Wang J, Cheng H, et al. Evolutionary study of the isoflavonoid pathway based on multiple copies analysis in soybean. *BMC Genetics*, 2014, 15: 76-88.
 62. Crabtree B, Newsholme EA. The derivation and interpretation of control coefficients. *Biochemical Journal*, 1987, 247: 113–120.
 63. Kosakovsky PSL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, 2005, 22: 1208–1222.
 64. McClellan DA, Palfreyman EJ, Smith MJ, et al. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Molecular Biology and Evolution*, 2005, 22: 437–455.
 65. Flowers JM, Sezgin E, Kumagai S, et al. Adaptive Evolution of metabolic pathways in *Drosophila*, *Molecular Biology and Evolution*, 2007, 24(6): 1347–1354.
 66. Lu YQ. Molecular evolution at the self-incompatibility locus of *Physalis longifolia* (Solanaceae). *Journal Molecular Evolution*, 2002, 54: 784-793.
 67. Swanson WJZ, Yang MF, Aquadro CF. Positive Darwinian selection derives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences USA*, 2001, 98: 2509–2514.
 68. Pal C, Papp B, Lercher MJ. An integrated view of protein evolution. *Nature Reviews Genetics* 2006, 7(5): 337-348.
 69. Cork JM, Purugganan MD. The evolution of molecular genetic pathways and networks. *Bioessays*, 2004, 26(5): 479-484.
 70. Eanes WF. Analysis of selection on enzyme polymorphisms. *Annual Review Ecology and Systematics*, 1999, 30(1): 301-326.
 71. Kwame A, Darfour O, Hendrik JM et al. Evolutionary patterns of Toll-like receptor signaling pathway genes in the Suidae. *BMC Evolutionary Biology*, 2016, 16: 33-44.
 72. Wu XC, Chi XQ, Wang PM, et al. The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. *Biology Direct*, 2010, 5: 46-56.
 73. Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Molecular Biology and Evolution*, 2007, 24: 836-844

74. Song H, Sun J, Yang GF. Comparative analysis of selection mode reveals different evolutionary rate and expression pattern in *Arachis duranensis* and *Arachis ipaënsis* duplicated genes. *Plant Molecular Biology*, 2018, 98: 349–361.
75. Guo Y, Liu J, Zhang J, et al. Selective modes determine evolutionary rates, gene compactness and expression patterns in Brassica. *The Plant Journal*, 2017, 91: 34–44.

Figures

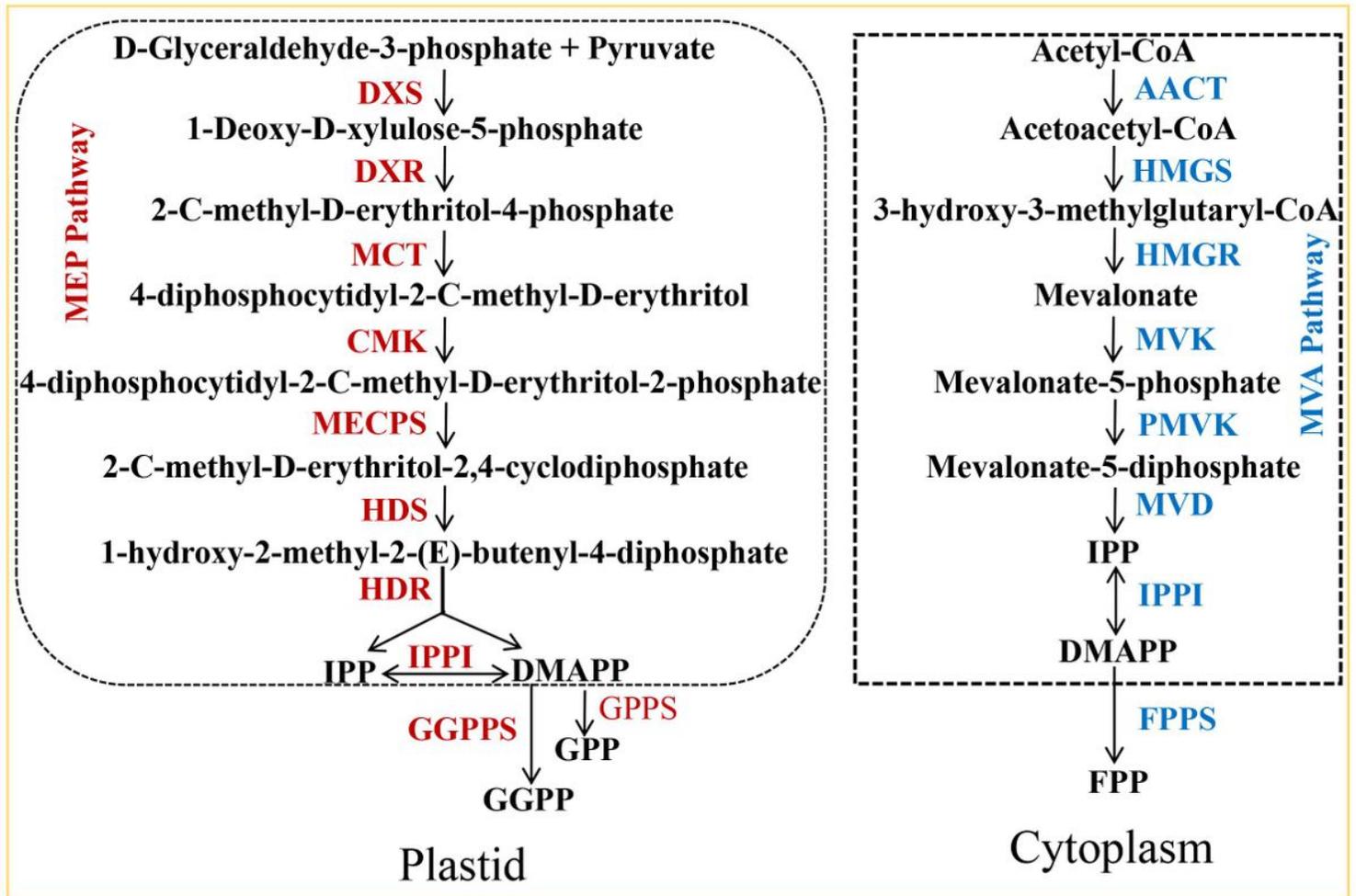


Figure 1

MVA and MEP pathways in plant terpenoid biosynthesis

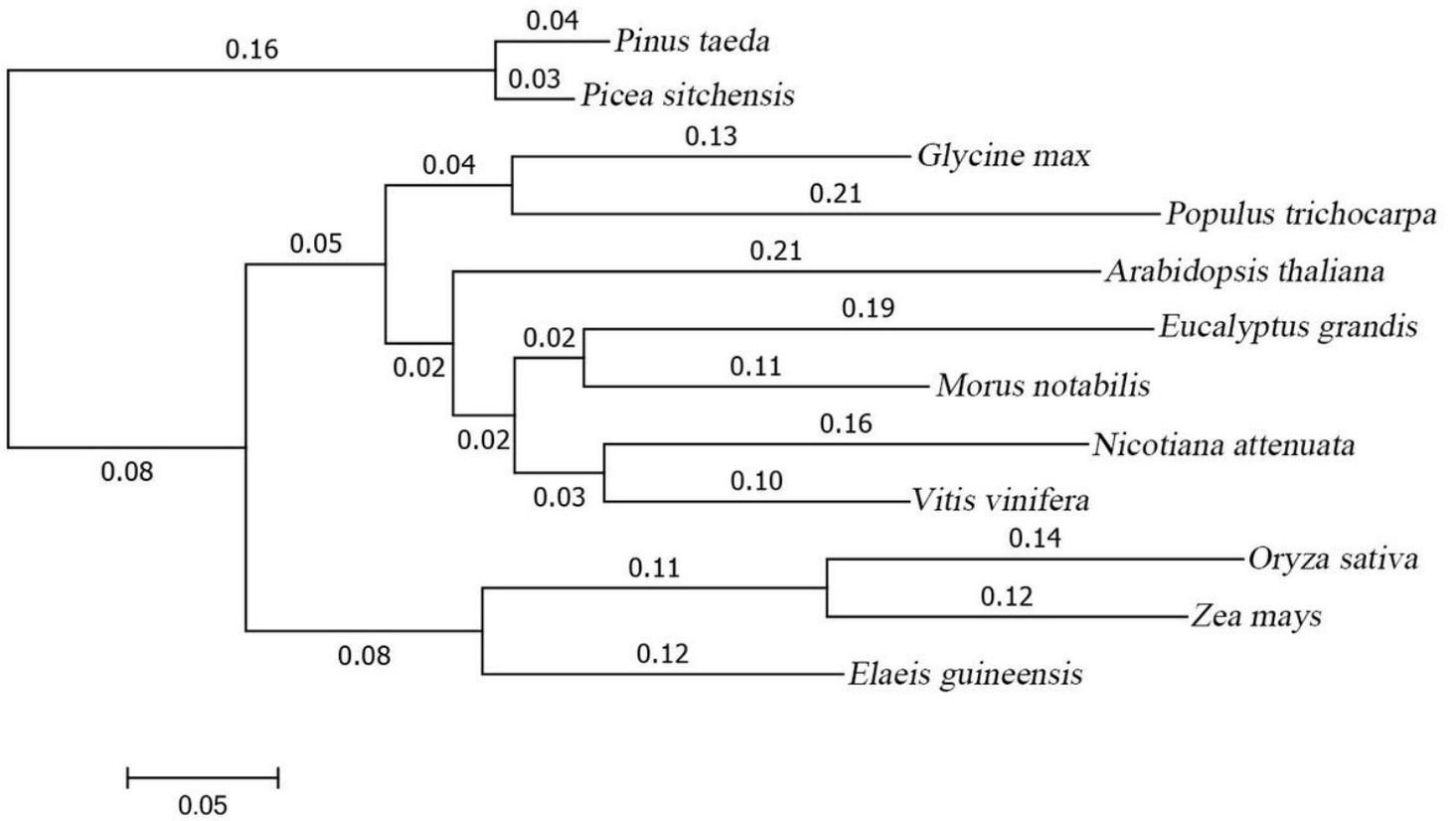


Figure 2

The maximum likelihood tree obtained using combined CDS of 16 terpenoid pathway genes and the GTR+I+G model. Numbers on branches are estimated branch lengths.

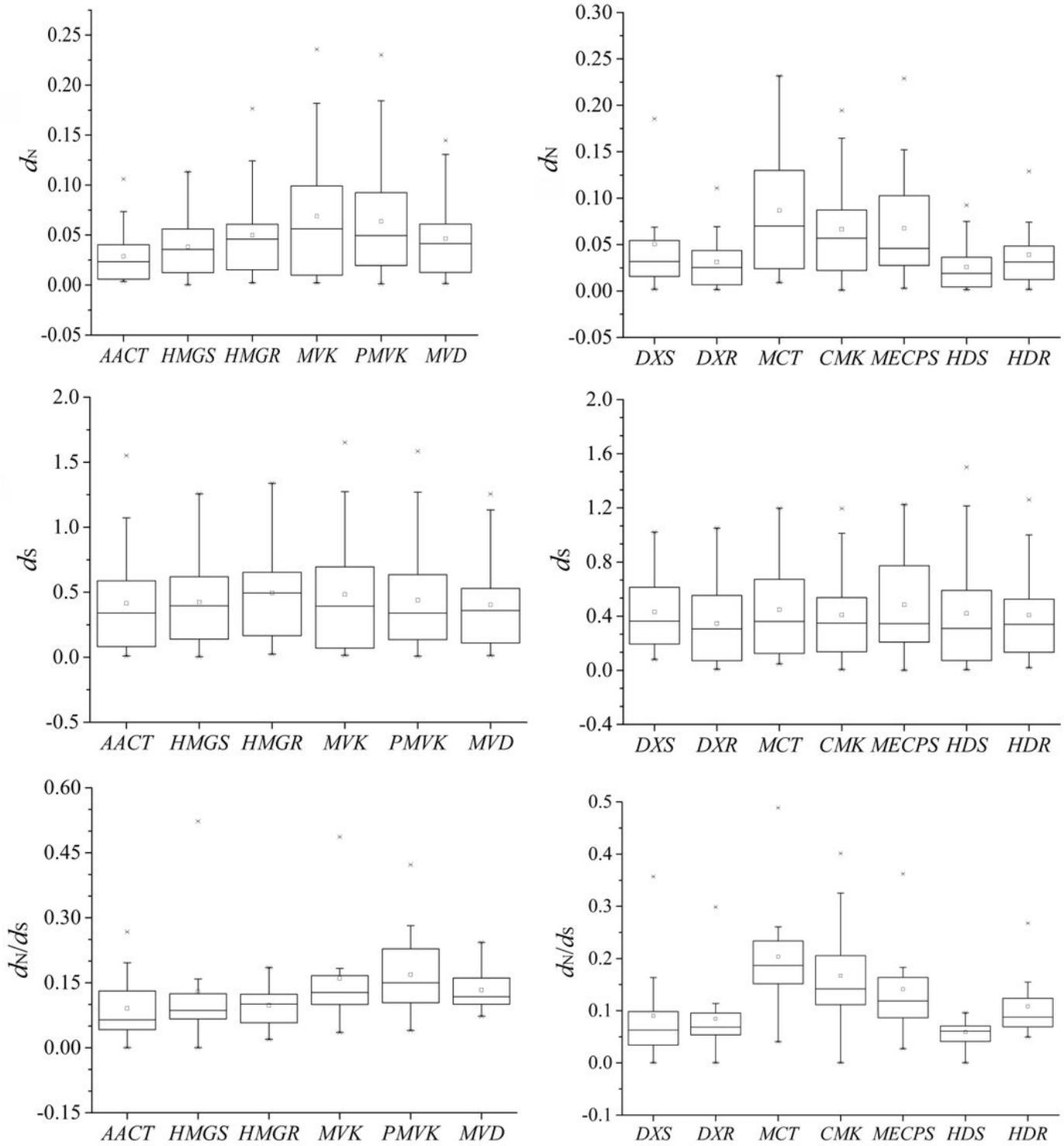


Figure 3

The evolutionary parameters of 21 branches pairs of MVA and MEP pathways genes. Average values of d_N , d_S and d_N/d_S ratio are represented by horizontal bars within the boxes at 95% confidence intervals.

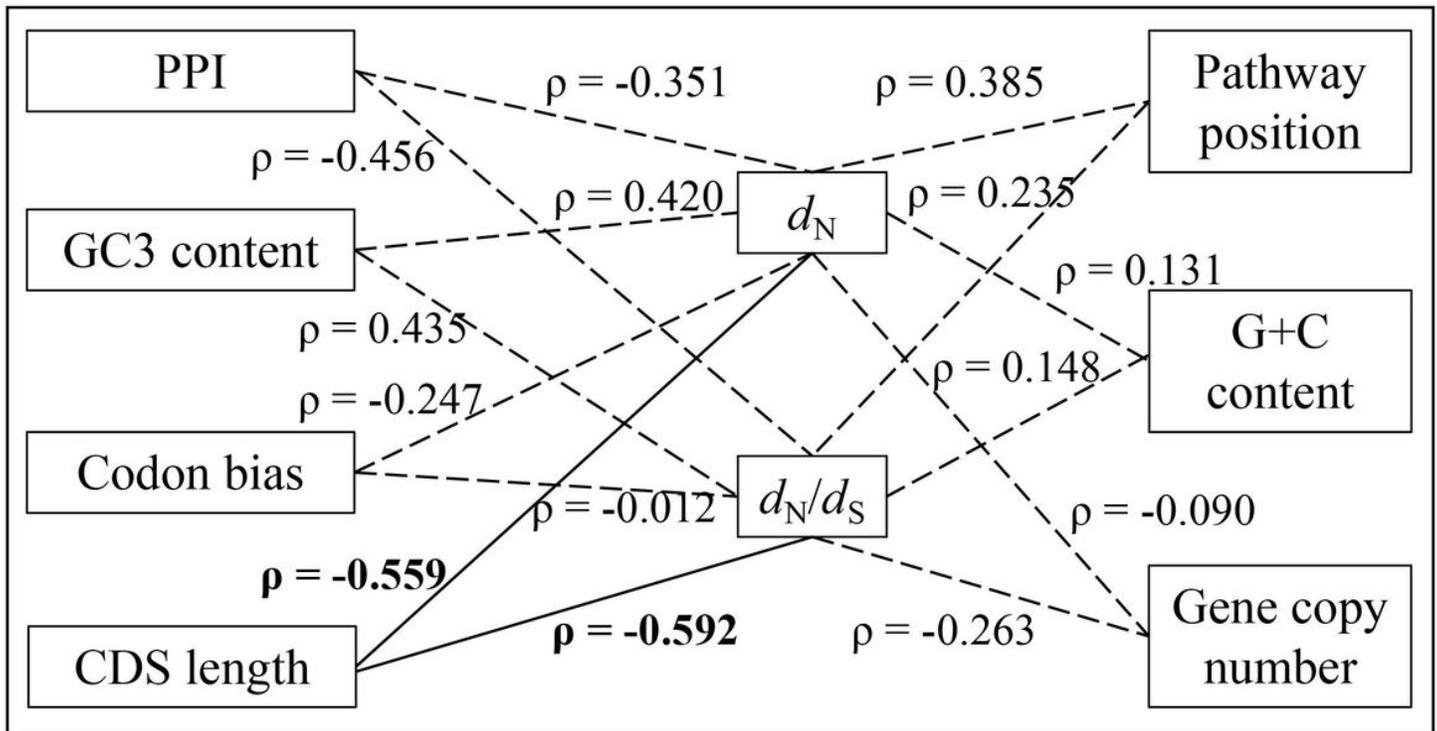


Figure 4

Spearman's rank correlation coefficient of d_N , d_N/d_S and pathway position, G+C, GC3, copy number, codon bias, PPI and CDS length. Solid and dotted lines represent significant and non-significant, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xls](#)
- [Additionalfile2.xls](#)
- [Additionalfile3.xls](#)
- [Additionalfile4.xls](#)
- [Additionalfile5.xls](#)