



# Computational Identification of Amino-Acid Mutations that Further Improve the Activity of a Chalcone–Flavonone Isomerase from *Glycine max*

Hui Yuan<sup>1</sup>, Jiaqi Wu<sup>1</sup>, Xiaoqiang Wang<sup>2</sup>, Jiakuan Chen<sup>1</sup>, Yang Zhong<sup>1,3</sup>, Qiang Huang<sup>4\*</sup> and Peng Nan<sup>1\*</sup>

<sup>1</sup> Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai, China, <sup>2</sup> Department of Biological Sciences, University of North Texas, Denton, TX, USA, <sup>3</sup> Institute of Biodiversity Science and Geobiology, Tibet University, Lhasa, China, <sup>4</sup> State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

## OPEN ACCESS

### Edited by:

Santosh Kumar Upadhyay,  
Panjab University, Chandigarh, India

### Reviewed by:

Ashutosh Pandey,  
National Agri-Food Biotechnology  
Institute, India  
Prashant Misra,  
Indian Institute of Integrative Medicine  
(CSIR), India

### \*Correspondence:

Peng Nan  
nanpeng@fudan.edu.cn  
Qiang Huang  
huangqiang@fudan.edu.cn

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Plant Science

Received: 01 November 2016

Accepted: 09 February 2017

Published: 24 February 2017

### Citation:

Yuan H, Wu J, Wang X, Chen J,  
Zhong Y, Huang Q and Nan P (2017)  
Computational Identification  
of Amino-Acid Mutations that Further  
Improve the Activity of a  
Chalcone–Flavonone Isomerase from  
*Glycine max*. *Front. Plant Sci.* 8:248.  
doi: 10.3389/fpls.2017.00248

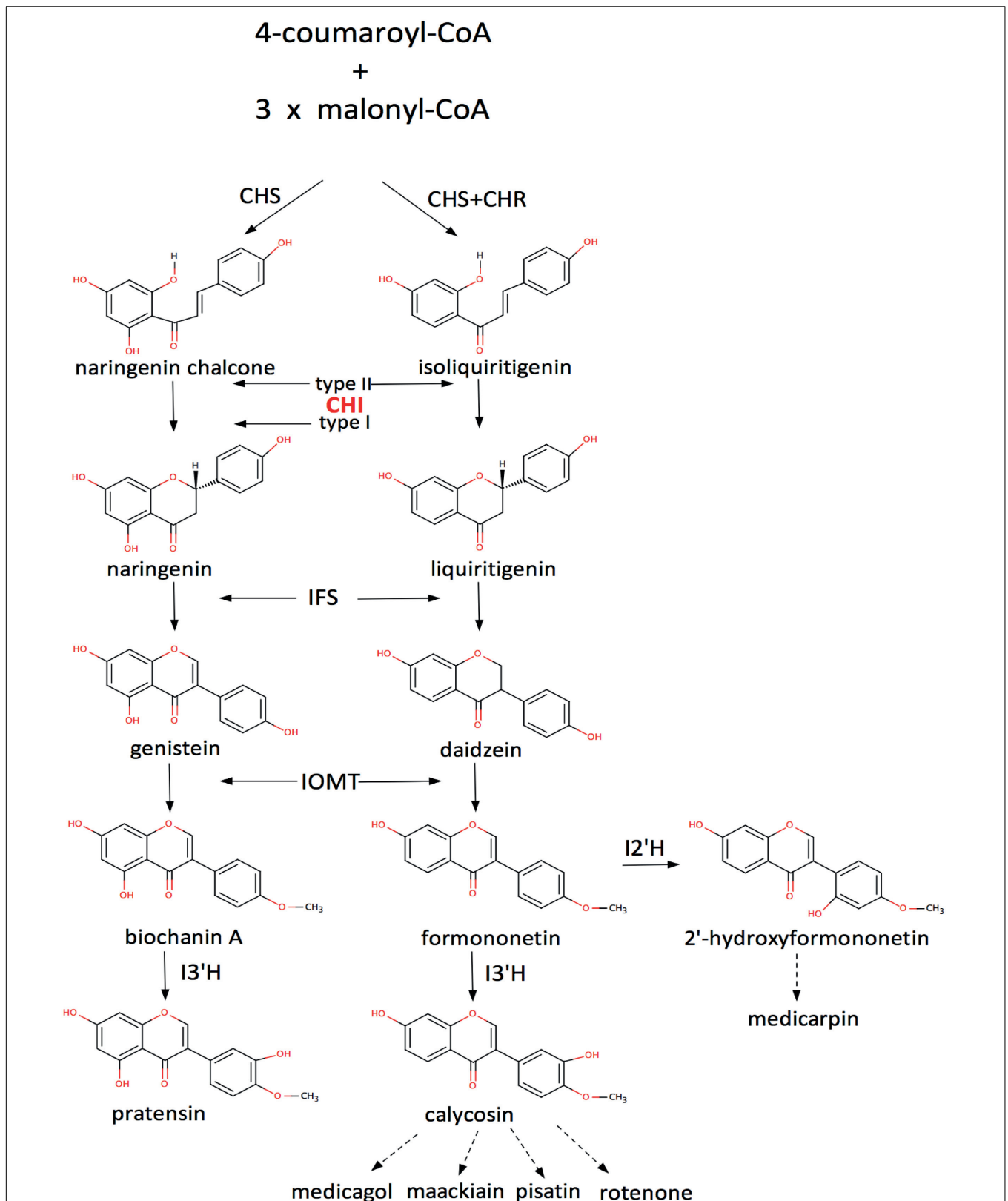
Protein design for improving enzymatic activity remains a challenge in biochemistry, especially to identify target amino-acid sites for mutagenesis and to design beneficial mutations for those sites. Here, we employ a computational approach that combines multiple sequence alignment, positive selection detection, and molecular docking to identify and design beneficial amino-acid mutations that further improve the intramolecular-cyclization activity of a chalcone–flavonone isomerase from *Glycine max* (GmCHI). By this approach, two GmCHI mutants with higher activities were predicted and verified. The results demonstrate that this approach could determine the beneficial amino-acid mutations for improving the enzymatic activity, and may find more applications in engineering of enzymes.

**Keywords:** enzyme engineering, positive selection, protein design, molecular modeling, chalcone–flavonone isomerase

## INTRODUCTION

Flavonoids are widespread secondary products in plants, especially in leguminous plants. They play important roles in plant physiology and ecology (Mol et al., 1998; Mahajan and Yadav, 2014), and are also important source of medicine and drug development (Dixon and Steele, 1999; Martens and Mithofer, 2005). Thus enzymes in the flavonoid biosynthetic pathways are of considerable value in biotechnological practices (Liu and Dixon, 2001; Liu et al., 2002, 2003). Of them, chalcone–flavonone isomerase (CHI) is an important enzyme in the biosynthetic pathway that catalyzes the intramolecular cyclization of a chalcone into a (2S)-flavonone (Figure 1, Supplementary Figure S1). According to their catalytic features, CHIs could be divided into two groups: type-I and type-II, respectively (Shimada et al., 2003). The type-II CHIs exist only in legumes and have broader substrate acceptability than the type-I enzymes (Figure 1), which are found in both non-legumes and legumes (Kimura et al., 2001; Shimada et al., 2003).

Because of their unique properties, there is a great demand of the flavonoids and their derivatives in biotechnology and medicine. However, the flavonoid production from natural plants could not fulfill such a demand. Therefore, it is desirable to develop biochemical and

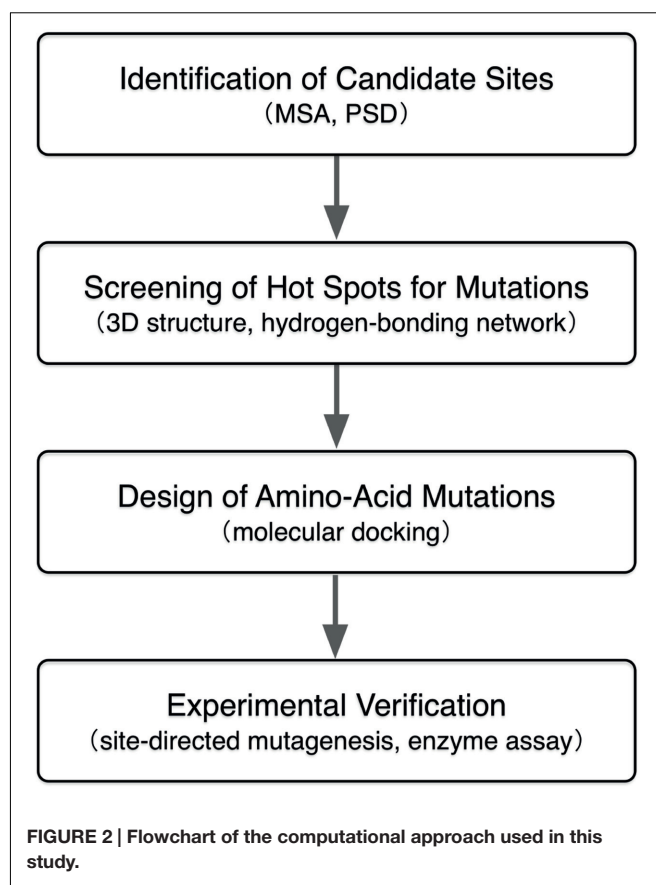


**FIGURE 1 | The flavonoid pathway.** CHS, chalcone synthase; CHR, chalcone reductase; IFS, isoflavone synthase; IOMT, 2'-hydroxyisoflavanone 4'-O-methyltransferase; I2'H, isoflavone 2'-hydroxylase; I3'H, isoflavone 3'-hydroxylase (Liu and Dixon, 2001; Liu et al., 2002, 2003; Shimada et al., 2003).

biotechnological methods to synthesize novel derivatives and increase their production by improving the corresponding biosynthetic enzymes, such as the CHIs. Thus, the improvement of the CHI activity is not only important for understanding the molecular determinants of the enzymatic activity, but also significant for biotechnological applications. Currently, rational design and directed evolution are two major strategies (Barak et al., 2008; Damian-Almazo and Saab-Rincon, 2013). Based on structural information, rational design usually identifies target amino acids near the active site, and then carries out site-directed mutagenesis to obtain protein mutants with enhanced activity. As an alternative, directed evolution mimics the process of nature evolution or/and recombination to obtain better mutants of the enzymes (Barak et al., 2008). Although both methods have proven to be useful in protein design, they also have certain limitations (Grove et al., 2003; Chica et al., 2005; Chen et al., 2009). Therefore, it is necessary to use combined approaches in order to overcome such limitations (Funke et al., 2005; Damian-Almazo and Saab-Rincon, 2013).

In practice, it is very laborious and costly to experimentally test a large number of candidate mutants. Therefore, it is very important to accurately identify residue sites for the mutagenesis. To the end, both sequence-based and structure-based methods were used. A common sequence-based approach is the multiple sequence alignment (MSA) that is effective for identification of conserved sites, some of which are the target mutation sites toward better activity. For example, one could build a correlation between the sequence pattern observed in the MSA and enzymatic property (Damian-Almazo and Saab-Rincon, 2013), e.g., the analysis of subfamily specific positions (SSPs) (Suplatov et al., 2012). Another way is to examine the ancestral relationship among the homologous sequences by combing the MSA with phylogenetic information (Di Giulio, 2001, 2003). On the other hand, structure-based approaches usually focus on those sites that are in the vicinity of the catalytic residues or the substrates (Morley and Kazlauskas, 2005; Park et al., 2006; Paramesvaran et al., 2009). However, because these approaches usually identify too many mutagenesis sites, it is still hard to experimentally test all those sites. In addition, mutants with reduced activity are often generated using these approaches, and it is very difficult to design and generate mutants with enhanced activity, especially for those which wild-type enzymes exhibit high efficiency. Therefore, to accurately identify the mutagenesis sites and predict proper amino-acid types on those sites, it is necessary to develop new strategies.

In this study, we employed a computational approach that combines the MSA, evolutionally positive selection detection (PSD), and structure-based molecular docking to identify the beneficial amino-acid mutations for enhancing the intramolecular-cyclization activity of a CHI from *Glycine max* (GmCHI, GI: 351723101), which is a type-II CHI and possesses high catalytic proficiency ( $k_{cat}/K_m$  is about  $5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ). In the present study, as shown in **Figure 2**, candidate mutation sites in CHI enzyme were firstly identified using the MSA and PSD. Then, those candidate sites were



investigated and further screened by analyzing the structural information and reaction mechanism. Next, the selected sites were further studied using molecular docking that determines the lowest-energy binding poses of the substrate in the active sites of the mutant enzymes. Finally, by *in vitro* assay using recombinant mutant enzymes, we identified beneficial amino-acid substitutions that improve the activity of GmCHI enzyme. Taken together, we demonstrate that our approach should be useful in designing of enzymes with improved enzyme activity.

## MATERIALS AND METHODS

### Expression and Enzyme Assay of GmCHI Proteins

The ORF of GmCHI gene was cloned into an expression vector pET28a+ (Novagen<sup>1</sup>) and then expressed in *Escherichia coli* strain BL21 (DE3). The protein expression was induced by IPTG (1 mM) at 20°C, 180 rpm for 6–12 h. After expression, the cells were harvested and the protein was purified with Ni<sup>2+</sup>-NTA agarose (Bio-Rad<sup>2</sup>). The activities of the GmCHI proteins were measured according to the reaction kinetics of CHI

<sup>1</sup><http://www.novagen.com>

<sup>2</sup><http://www.bio-rad.com>

enzymes. The substrate was incubated at 25°C, 90 s with total 500  $\mu$ l reaction buffer (50 mM Tris, 500 mM NaCl, 1.0 mM DTT, pH 7.8) containing 5 ng of purified GmCHI protein. We performed the enzyme assays in a gradient concentration of 2–100  $\mu$ M for isoliquiritigenin. After the reaction, the reaction mixtures including isoliquiritigenin and liquiritigenin were analyzed on an Agilent HP1100 HPLC with eclipse plus C-18 column. The eluents, consisting of 35% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water, were monitored at 276 and 372 nm (at constant flow rate of 1 ml per minute). The UV absorption values at 276 and 372 nm were used for quantifying liquiritigenin and isoliquiritigenin, respectively (He and Dixon, 2000; Liu and Dixon, 2001; Liu et al., 2002, 2003).

### Site-Directed Mutagenesis of GmCHI

The site-directed mutagenesis was performed using a mutagenesis kit from SBS Genetech<sup>3</sup> and by following the manufacturer's instructions. The primers used for the site-directed mutagenesis are listed in Supplementary Table S1. The mutants were confirmed by sequencing and then expressed in *Escherichia coli* according to the methods described above for the wild-type enzyme.

### Homology Modeling and Molecular Docking

We used the MODELLER program (Eswar et al., 2007) to build the all-atom structural models of the wild-type GmCHI protein and its mutants, with the crystal structure of MsCHI (PDB code: 1F7M) (Jez et al., 2000) as the template. Then, 2,000 independent, standard high-resolution refinement runs with the Rosetta program (Das and Baker, 2008) were carried out to generate refined atomic models with low free-energies. For each GmCHI protein (the wild-type or the mutant), the 3D structure with the lowest free-energy from the 2,000 refined models was selected as the receptor structure for the subsequent molecular docking.

Hydrated ligand docking using the program AutoDock 4.2 (Forli and Olson, 2012) was conducted to predict the binding poses of the substrate in the active-sites of the wild-type GmCHI protein and its mutants. The hydration state of the substrate for the docking was determined and treated according to the reported method (Forli and Olson, 2012). All other docking parameters for the proteins and the substrate were set to the default values of AutoDock (Morris et al., 2009), with a size of the grid box around the active-site as 70 Å  $\times$  70 Å  $\times$  70 Å. The Lamarckian genetic algorithm was employed to search for the native-like binding pose, with a population number of 150, a maximum of 27,000 generations, and a maximum of 1,500,000 energy evaluations. To construct the binding energy landscape of the substrate in the active site of a GmCHI protein, 2,000 independent docking runs were performed, and thereby 2,000 binding poses were obtained for analysis and generating the RMSD-binding free energy plot.

<sup>3</sup><http://www.sbsgene.com>

## RESULTS

### Identification of Amino-Acid Sites for Site-Directed Mutagenesis Multiple Sequence Alignment (MSA)

Identification of the amino-acid sites for the site-directed mutagenesis is the first step toward the improvement of the GmCHI activity. To this end, we collected the sequences of 14 homologous CHIs (Supplementary Table S2), including seven type-II CHIs (group 1) and seven type-I CHIs (group 2), respectively. Then, we carried out the MSA for these CHIs, and thereby identified nine SSPs (Figure 3). On each of these sites, the wild-type amino acids are conserved within either the type-I or type-II group, but different between the type-I and type-II groups. Because the type-I and type-II CHIs are different in the catalysis (Kimura et al., 2001; Shimada et al., 2003), some of these sites might directly affect the reaction mechanisms and thus the enzymatic activity. Therefore, we considered them as the candidate sites for introducing beneficial amino-acid mutations.

### Positive Selection Detection (PSD)

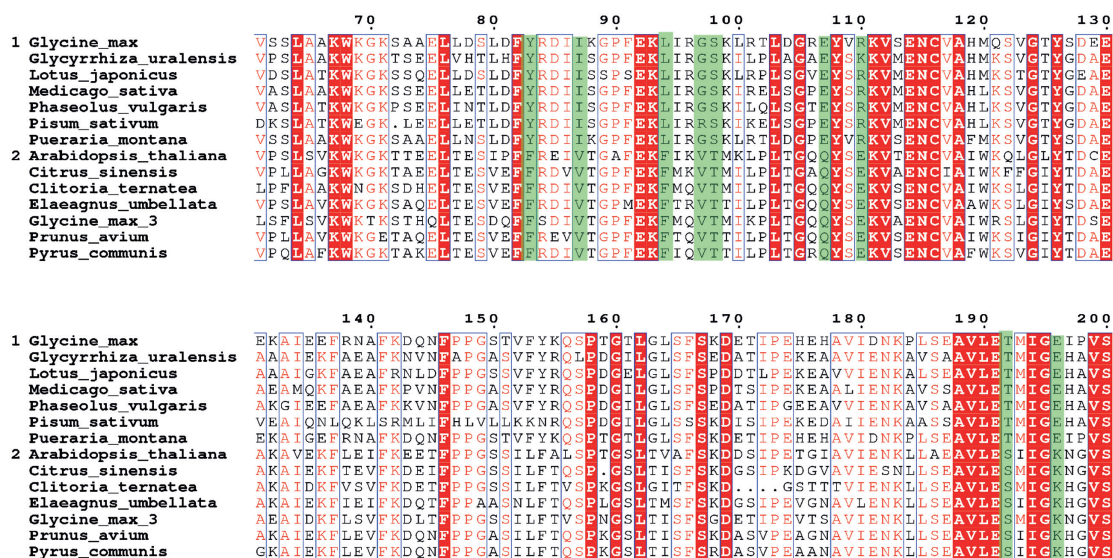
Besides the moderate conserved sites identified in the MSA, some particular unconserved sites under selection pressure (i.e., the positively selected sites) may also play a critical role in the evolution of protein function, because positive selection was considered to drive the fixation of advantageous mutations (Yang, 2006). Therefore, PSD was performed to further identify candidate sites for the mutagenesis. Firstly, we used RaxML 7.04 (Stamatakis, 2006; Stamatakis et al., 2008) under the GTR+ $\Gamma$ +I model to infer a phylogenetic tree of 15 CHI genes from 13 species, including eight type-I and seven type-II CHI genes. Then, based on this tree topology (Supplementary Figure S2) and branch-site model (Yang and Nielsen, 2002), the posterior probability of every site under positive selection was calculated with PAML 4.4 (Yang, 1997, 2007; Yang et al., 2005). Two sites (Val109 and Ile197) were identified as the positively selected sites, and then considered as the candidate mutation sites for improving the activity.

### Screening by Structural and Reaction Information

By the above sequence-based methods, we have identified 11 potential sites for the mutagenesis. We further analyzed their spatial positions in the 3D structure obtained by homology modeling, and their effects on the hydrogen-bonding network in the active site required for the catalysis (Supplementary Figure S3). By manual inspection with the structural model (Supplementary Figures S4–S7), four target mutation sites were chosen, namely, Glu107, Ala110, Glu196 and Ile197, including conserved and unconserved sites (Supplementary Figure S8).

### Amino-Acid Mutation Design by Molecular Docking

For the candidate mutation sites identified above, we further performed molecular docking studies to determine the types of amino acids for the substitution toward improving activity. For each site, we predicted the activities of the mutants with about



**FIGURE 3 | The MSA of the CHI core regions.** The seven sequences in group 1 are type- II CHIs and another seven sequences in group 2 are type-I CHIs. The sequence positions are numbered according to GmCHI. The subfamily specific positions (SSPs) are marked in green.

five representative amino acids according to the physicochemical properties of their side chains, namely, a non-polar, an aromatic, a non-charged polar, an alkaline, and an acidic amino acid (for Glu196 and Ile197 on the loop, we predicted the activity of the proline mutant). To predict the best amino-acid type at a given candidate site, we firstly constructed the 3D structure of a GmCHI mutant by the same homology modeling method for the wild-type enzyme. Then, we used the program AutoDock to dock the substrate (i.e., isoliquiritigenin) into the active site of the mutant by the protocols as described in Section “Materials and Methods.” Two thousand independent docking runs were conducted for each amino-acid mutation. Eventually, the RMSD values of the 2000 docking poses with respect to the product (i.e., liquiritigenin) (Figure 4A) were calculated and analyzed to identify the most likely amino-acid mutation.

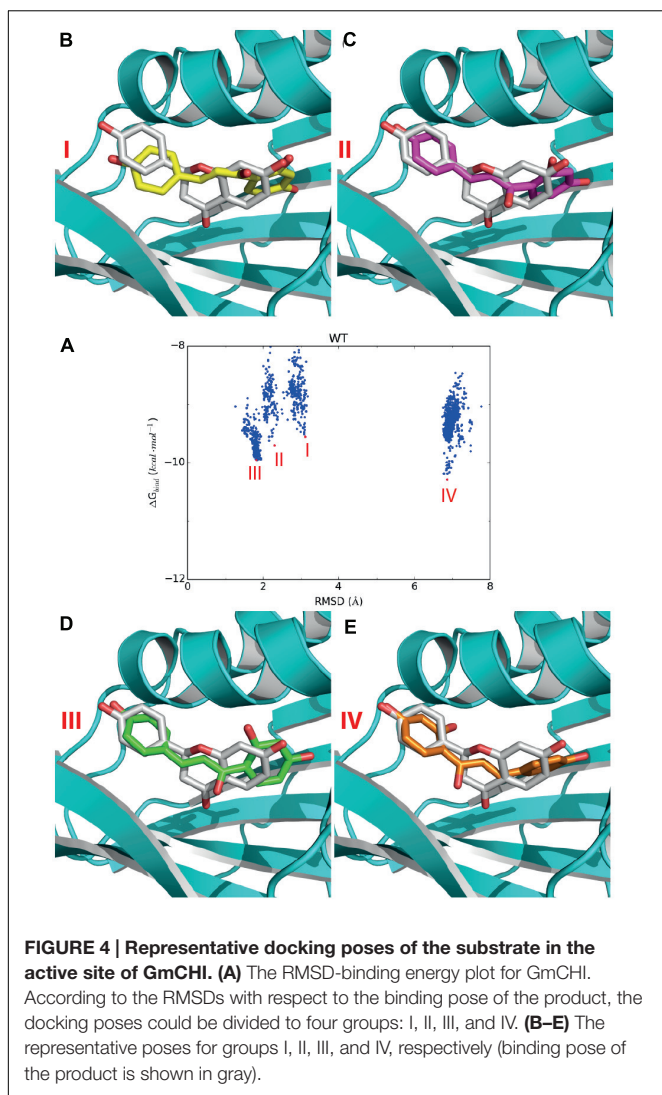
To select the amino-acid type at the candidate sites, we firstly analyzed the RMSD-energy plot of the wild-type GmCHI obtained by molecular docking. As shown in Figure 4A, we found that the lowest-energy docking poses in the active site could be clustered into four main groups: groups I, II, III, and IV, respectively. For a given group, we selected the docking pose with the lowest binding energy in the group as the representative pose of the group. As indicated by the RMSD values, the representative binding poses of groups I, II, and III are very similar to that of the product revealed by the crystal structure (Figures 4B–D), whereas that of group IV possess an almost opposite orientation with a RMSD > 6.0 Å (Figure 4E). Moreover, the binding energies of the representative poses in groups II, III, and IV are lower than that of group I. Therefore, considering the binding to the active site is the initial step of the reaction, the most likely binding conformation of the substrate in the initial phase of the reaction is the representative pose of group I. Because of their similarity in the binding conformation

of the product, the representative poses of groups II and III could be considered as intermediate conformations from the substrate to the product. Considering the order of the binding energy in three groups: group I > group II > group III, we may regard them as a transition states from the reactant to the product (Figures 4B–D). Thus, we hypothesized that, if an amino-acid mutation could further lower the binding energy of the representative pose of group III, the enzymatic activity of the mutant may be improved with respect that of the wild-type GmCHI.

Based on the above hypothesis, we generated the RMSD-binding free energy plots for the wild-type enzyme and all possible GmCHI mutants. Interestingly, we found that the binding energy of the I197P mutant with the representative pose of group III is significantly lower than that of the wild-type enzyme (Figure 5A), suggesting that this mutant might have higher activity than the wild-type enzyme. On the contrary, the corresponding energy of the R110H mutant is higher than that of the wild-type (Figure 5B), suggesting a decrease in the enzymatic activity with respect to the wild-type. Similarly, three mutants (E107D, R110A, and I197P) were also predicted to have higher activities than the wild-type (Supplementary Figures S9–S11).

## Molecular Cloning, Mutagenesis, and Enzyme Assay

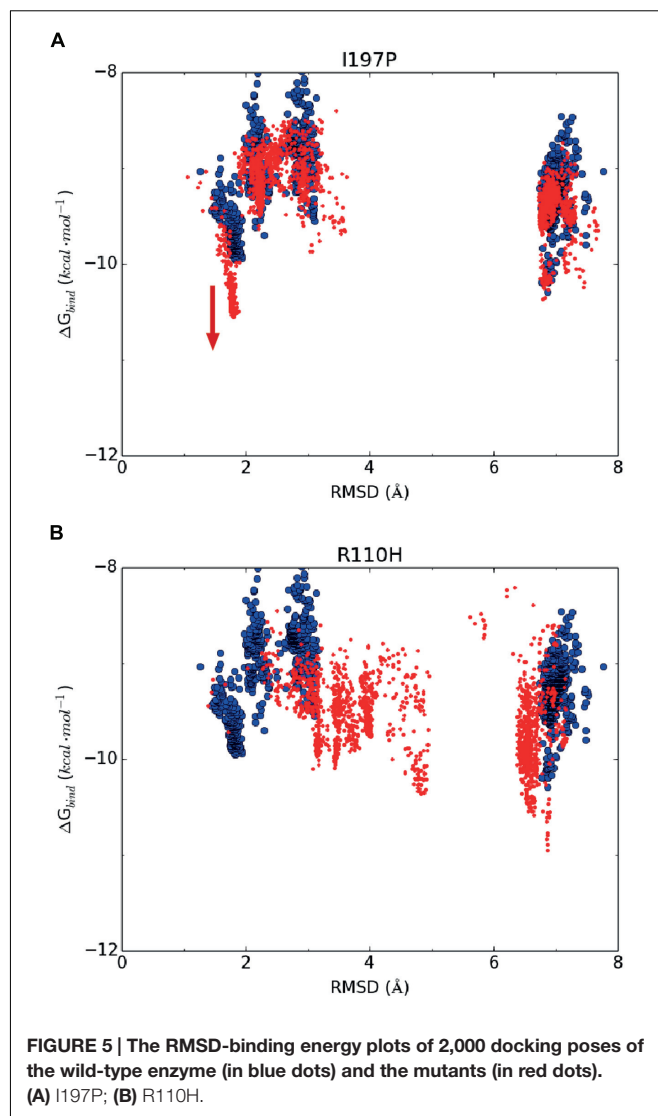
According to the computational predictions, we carried out experiments to verify the activities of three beneficial GmCHI mutants (E107D, R110A, and I197P), as shown in Supplementary Figure S12. To the end, the wild-type GmCHI was cloned, and the site-directed mutagenesis was conducted on the target binding sites. To examine the accuracy of the docking results, we also tested other mutants that represent various amino-acid types. We expressed the wild-type and mutant enzymes in



*E. coli* expression system with BL21 (DE3) cells and purified them for the activity measurement (see Materials and Methods). The activities of all tested mutants are listed in **Table 1** (for more details see Supplementary Figures S13–S15). Compared with the wild-type enzyme, two mutants (R110A and I197P) do possess relatively higher activities, other mutants do not. This is consistent with the computational predictions (**Figure 5**, Supplementary Figures S9–S11). Significantly, the increase in the activity by the I197P mutation is about 53.3%, in good agreement with the prediction (**Figure 5A**). Thus, by using the computational approach in **Figure 2**, we identified two GmCHI mutants with higher activities.

## DISCUSSION

In this study, we have used a computational approach that integrates sequence-based analysis (MSA, PSD) and structure-based docking to identify the amino-acid hot spots for the site-directed mutagenesis, and



then to predict the beneficial amino-acid mutations at those hot spots. The results demonstrate that the used approach could identify the beneficial amino-acid

**TABLE 1 | Activities of the GmCHI mutants with respect to that of the wild-type.**

Enzyme	$k_{cat}/K_m$ ( $10^6 \text{ M}^{-1} \text{ s}^{-1}$ )	Change in activity (%)	<i>p</i> -value*
Wild-type	$4.963 \pm 0.306$	0.0	–
E107D	$4.773 \pm 0.336$	–3.8	0.686
E107Q	$3.715 \pm 0.221$	–25.1	$8.82 \times 10^{-4}$
R110A	$7.0477 \pm 0.412$	42.0	$3.24 \times 10^{-3}$
R110E	$4.142 \pm 0.316$	–16.6	0.046
R110H	$3.345 \pm 0.173$	–32.6	$2.98 \times 10^{-6}$
I197P	$7.608 \pm 0.574$	53.3	$5.90 \times 10^{-4}$

\*Calculated by covariance analysis using R software (<http://www.R-project.org>).

mutations that further improve the intramolecular-cyclization activity of GmCHI. Usually, when the catalytic proficiency of an enzyme ( $k_{\text{cat}}/K_m$ ) reaches about  $10^7 \text{ M}^{-1} \text{ s}^{-1}$ , diffusion rate becomes the main limitation factor in the catalysis (Nelson and Cox, 2000). Then, it becomes difficult to further improve the catalytic proficiency. As mentioned, though the wild-type GmCHI has evolved to possess high catalytic proficiency ( $k_{\text{cat}}/K_m = 5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ), the two GmCHI mutants predicted by our approach could further display a 50% increase in catalytic proficiency. Earlier researchers have employed bioinformatics methods for identifying potential target sites to be used for introducing mutations. These methods were very successful in the identification of those sites that stabilize enzymes (Damborsky and Brezovsky, 2014), e.g., thermostability (Sullivan et al., 2011; Anbar et al., 2012; Blum et al., 2012). However, they led to limited success in the identification of activity-related sites, because it is difficult to find out the functionally related sites based on sequence information alone.

Typically, to identify such sites, bioinformatics approaches are used for the analysis of moderately conserved sites, such as the SSPs (Suplatov et al., 2012). In our study, nine SSPs were identified in the MSA analysis, but only one of them was found to improve the enzymatic activity. Due to the natural selection, usually the wild-type amino acid at the conserved sites possesses the best activity, for example, mutations at site Glu107 do not improve the activity. Therefore, it remains very interesting to develop new methods that could identify functionally related sites from the unconserved sites.

The positively selected sites (Yang, 2006) are probably the most interesting unconserved sites, because they drive the functional divergence of many enzyme families during the evolution (Barkman et al., 2007; Lan et al., 2009, 2013; Huang et al., 2012), and could affect the activity or specificity of an enzyme (Kapralov and Filatov, 2007; Hao et al., 2010; Huang et al., 2012; Lan et al., 2013). Thus, PSD offers a way to predict the functionally related sites. In this study, two positively selected sites were detected, and the mutation at one of them was found to improve the enzymatic activity. Indeed, PSD is also able to identify functionally related sites that are not restricted in the vicinity of the active site, but distant away from the active site (Gillespie, 1991; Yang, 2006), e.g., Ile197.

On the other hand, to predict beneficial amino-acid mutations at the hot spots for the mutagenesis, here we used molecular docking as a fast method to search for such mutations, instead of using the computation intensive QM/MM methods. As demonstrated by the results, the predictions for almost all the

six mutants are consistent with the experimental results (Table 1; Supplementary Materials see Enzyme assay). No doubt, this design strategy could also be used to improve other enzymes whose complex structures with the products have already been solved.

## CONCLUSION

To improve the GmCHI activity, we used a computational approach that combines sequence-based analysis with structure-based docking to identify the hot spots for amino-acid mutations and design beneficial mutations at those sites. We successfully discovered two GmCHI mutants display higher activities than that of the wild-type enzyme. Because of its simplicity and low computational cost, this approach may find more applications in the design and engineering of enzymes.

## AUTHOR CONTRIBUTIONS

PN and QH conceived and designed the study, and revised the manuscript; HY prepared samples, analyzed data, drafted the manuscript, and performed experiment verifications. JW performed the PSD; XW and QH directed the analysis of CHI structural models and molecular docking; JC and YZ provided constructive advices on the study. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the grants from, the Natural Science Foundation of China (No. 81373963, 91430112, 31671386), the Shanghai Natural Science Foundation (No. 13ZR1402400), the National High-tech R&D Program of China (2012AA02A602), and the Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00248/full#supplementary-material>

## REFERENCES

- Anbar, M., Gul, O., Lamed, R., Sezerman, U. O., and Bayer, E. A. (2012). Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. *Appl. Environ. Microbiol.* 78, 3458–3464. doi: 10.1128/Aem.07985-11
- Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008). Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J.* 2, 171–179. doi: 10.1038/ismej.2007.100
- Barkman, T. J., Martins, T. R., Sutton, E., and Stout, J. T. (2007). Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. *Mol. Biol. Evol.* 24, 1320–1329. doi: 10.1093/molbev/msm053
- Blum, J. K., Ricketts, M. D., and Bommarius, A. S. (2012). Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method. *J. Biotechnol.* 160, 214–221. doi: 10.1016/j.jbiotec.2012.02.014
- Chen, C. Y., Georgiev, I., Anderson, A. C., and Donald, B. R. (2009). Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3764–3769. doi: 10.1073/pnas.0900266106
- Chica, R. A., Doucet, N., and Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and

- rational design. *Curr. Opin. Biotechnol.* 16, 378–384. doi: 10.1016/J.Copbio.2005.08.004
- Damborsky, J., and Brezovsky, J. (2014). Computational tools for designing and engineering enzymes. *Curr. Opin. Chem. Biol.* 19, 8–16. doi: 10.1016/j.cbpa.2013.12.003
- Damian-Almazo, J. Y., and Saab-Rincon, G. (2013). “Site-directed mutagenesis as applied to biocatalysts,” in *Biochemistry, Genetics and Molecular Biology*, ed. D. Figurski (Rijeka: In Tech), doi: 10.5772/53330
- Das, R., and Baker, D. (2008). Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* 77, 363–382. doi: 10.1146/annurev.biochem.77.062906.171838
- Di Giulio, M. (2001). The universal ancestor was a thermophile or a hyperthermophile. *Gene* 281, 11–17. doi: 10.1016/S0378-1119(01)00781-8
- Di Giulio, M. (2003). The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.* 221, 425–436. doi: 10.1006/Jtbi.2003.3197
- Dixon, R. A., and Steele, C. L. (1999). Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends Plant Sci.* 4, 394–400. doi: 10.1016/S1360-1385(99)01471-5
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., et al. (2007). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* Chapter 2, Unit2.9. doi: 10.1002/0471140864.ps0209s50
- Forli, S., and Olson, A. J. (2012). A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* 55, 623–638. doi: 10.1021/jm2005145
- Funke, S. A., Otte, N., Eggert, T., Bocola, M., Jaeger, K. E., and Thiel, W. (2005). Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Eng. Des. Sel.* 18, 509–514. doi: 10.1093/Protein/Gzi062
- Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford: Oxford University Press.
- Grove, J. I., Lovering, A. L., Guise, C., Race, P. R., Wrighton, C. J., White, S. A., et al. (2003). Generation of *Escherichia coli* nitroreductase mutants conferring improved cell sensitization to the prodrug CB1954. *Cancer Res.* 63, 5532–5537.
- Hao, D. C., Mu, J., and Xiao, P. G. (2010). Molecular evolution and positive Darwinian selection of the gymnosperm photosynthetic Rubisco enzyme. *Bot. Stud.* 51, 491–510.
- He, X. Z., and Dixon, R. A. (2000). Genetic manipulation of isoflavone 7-O-methyltransferase enhances biosynthesis of 4'-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. *Plant Cell* 12, 1689–1702. doi: 10.1105/tpc.12.9.1689
- Huang, R. Q., Hippauf, F., Rohrbeck, D., Hausteil, M., Wenke, K., Feike, J., et al. (2012). Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2966–2971. doi: 10.1073/Pnas.1019605109
- Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* 7, 786–791. doi: 10.1038/79025
- Kapralov, M. V., and Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* 7:73. doi: 10.1186/1471-2148-7-73
- Kimura, Y., Aoki, T., and Ayabe, S. (2001). Chalcone isomerase isozymes with different substrate specificities towards 6'-hydroxy- and 6'-deoxychalcones in cultured cells of *Glycyrrhiza echinata*, a leguminous plant producing 5-deoxyflavonoids. *Plant Cell Physiol.* 42, 1169–1173. doi: 10.1093/Pcp/Pce130
- Lan, T., Wang, X. R., and Zeng, Q. Y. (2013). Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family. *J. Biol. Chem.* 288, 24441–24451. doi: 10.1074/Jbc.M113.456863
- Lan, T., Yang, Z. L., Yang, X., Liu, Y. J., Wang, X. R., and Zeng, Q. Y. (2009). Extensive functional diversification of the populus glutathione S-transferase supergene family. *Plant Cell* 21, 3749–3766. doi: 10.1105/Tpc.109.070219
- Liu, C. J., Blount, J. W., Steele, C. L., and Dixon, R. A. (2002). Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14578–14583. doi: 10.1073/pnas.212522099
- Liu, C. J., and Dixon, R. A. (2001). Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents the formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *Plant Cell* 13, 2643–2658. doi: 10.1105/tpc.13.12.2643
- Liu, C. J., Huhman, D., Sumner, L. W., and Dixon, R. A. (2003). Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from *Medicago truncatula*. *Plant J.* 36, 471–484. doi: 10.1046/j.1365-313X.2003.01893.x
- Mahajan, M., and Yadav, S. K. (2014). Overexpression of a tea flavanone 3-hydroxylase gene confers stress tolerance to salt stress and *Alternaria solani* in transgenic tobacco. *Plant Mol. Biol.* 85, 551–573. doi: 10.1007/s11103-014-0203-z
- Martens, S., and Mithofer, A. (2005). Flavones and flavone synthases. *Phytochemistry* 66, 2399–2407. doi: 10.1016/j.phytochem.2005.07.013
- Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends Plant Sci.* 3, 212–217. doi: 10.1016/S1360-1385(98)01242-4
- Morley, K. L., and Kazlauskas, R. J. (2005). Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* 23, 231–237. doi: 10.1016/J.Tiblech.2005.03.005
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/Jcc.21256
- Nelson, D. L., and Cox, M. M. (2000). *Lehninger Principles of Biochemistry*, 3rd Edn. New York, NY: Worth Publishers.
- Paramesvaran, J., Hibbert, E. G., Russell, A. J., and Dalby, P. A. (2009). Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. *Protein Eng. Des. Sel.* 22, 401–411. doi: 10.1093/Protein/Gzp020
- Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., et al. (2006). Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311, 535–538. doi: 10.1126/Science.1118953
- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S. (2003). A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol.* 131, 941–951. doi: 10.1104/pp.004820
- Stamatakis, A. (2006). RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/Bioinformatics/Btl446
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAXML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642
- Sullivan, B. J., Durani, V., and Magliery, T. J. (2011). Triosephosphate Isomerase by consensus design: dramatic differences in physical properties and activity of related variants. *J. Mol. Biol.* 413, 195–208. doi: 10.1016/J.Jmb.2011.08.001
- Suplatov, D. A., Besenmatter, W., Svedas, V. K., and Svendsen, A. (2012). Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Eng. Des. Sel.* 25, 689–697. doi: 10.1093/Protein/Gzs068
- Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556. doi: 10.1093/bioinformatics/13.5.555
- Yang, Z. H. (2006). *Neutral and Adaptive Protein Evolution in Computational Molecular Evolution*. Oxford: Oxford University Press, 259–292.
- Yang, Z. H. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/Molbev/Msm088
- Yang, Z. H., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917. doi: 10.1093/oxfordjournals.molbev.a004148
- Yang, Z. H., Wong, W. S. W., and Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/Molbev/Msi097

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Yuan, Wu, Wang, Chen, Zhong, Huang and Nan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.