Molecular evolution of the ent-kaurenoic acid oxidase gene in Oryzeae

Yanhua Yang1,2 and Keping Chen1

1Institute of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu Province, China.
2State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China.

Abstract

We surveyed the substitution patterns in the ent-kaurenoic acid oxidase (KAO) gene in 11 species of Oryzeae with an outgroup in the Ehrhartoidaeae. The synonymous and non-synonymous substitution rates showed a high positive correlation with each other, but were negatively correlated with codon usage bias and GC content at third codon positions. The substitution rate was heterogeneous among lineages. Likelihood-ratio tests showed that the non-synonymous/synonymous rate ratio changed significantly among lineages. Site-specific models provided no evidence for positive selection of particular amino acid sites in any codon of the KAO gene. This finding suggested that the significant rate heterogeneity among some lineages may have been caused by variability in the relaxation of the selective constraint among lineages or by neutral processes.

Key words: codon usage bias, ent-kaurenoic acid oxidase (KAO), positive selection, rate heterogeneity, substitution rate.

Introduction

Gibberellins (GAs) are an important class of plant hormones involved in the regulation of various growth and developmental processes in higher plants (Appleford et al., 2006). The absence of GAs results in dwarfism in some plant species. ent-kaurenoic acid oxidase (KAO), a member of the CYP88A subfamily of cytochrome P450 enzymes, catalyzes a three-step reaction in the gibberellin biosynthetic pathway from ent-kaurenoic acid to GA12 (Helliwell et al., 2001). A primary goal of molecular evolutionary studies is to estimate the rate of DNA mutation and elucidate the mechanisms of molecular evolution. Such studies frequently involve a comparison of orthologous DNA fragments among species to determine evolutionary rates and an assessment of the evolutionary processes involved, e.g., natural selection, rate heterogeneity of lineages and mutational biases. Analysis of the molecular evolutionary patterns of different genes provides understanding of the evolutionary processes and pressures experienced by particular lineages.

The tribe Oryzeae (Poaceae) includes approximately 12 genera and more than 70 species distributed throughout tropical and temperate regions of the world (Clayton and Renvoize, 1986; Vaughan, 1994). In the genus Oryza, the Asian cultivated rice (Oryza sativa L.) is one of the world’s most important crops and a primary food source for more than one-half of the world’s population (Chandler and Wessler, 2001). This species has become a model monocotyledon in scientific research and its entire genome has been sequenced. Other members of the Oryzeae are also of economic importance, including wild species of Oryza that can be used in the genetic improvement of rice.

Analysis of the substitution patterns in the KAO gene can provide insights into the driving forces that have led to evolutionary change in this gene in Oryzeae. In addition, the identification of patterns of molecular evolution in the KAO gene can improve our understanding of the evolutionary history of some Oryzeae species. In this work, we examined the heterogeneity of the substitution rate in the KAO gene among various genera and species of Oryzeae and sought to identify the possible causes of such heterogeneity. We also sought for evidence of natural selection in the exon regions of the KAO gene.

Materials and Methods

Plant material

A portion of the KAO gene was isolated and sequenced from members of the rice tribe (Oryzeae) (Table 1). Eleven diploid species were selected to represent the major phylogenetic lineages of Oryzeae (Figure S1, Supplementary Material) (Guo and Ge, 2005). These consisted of seven Oryza species representing six diploid genome types, namely, Oryza sativa (AA), O. meridionalis (AA), O. punctata (BB), O. officinalis (CC), O. australiensis (EE), O. brachyantha (FF), O. granulata (GG), and one
species from each of four other genera in the tribe Oryzeae (Leersia tisserantii, Chikusichloa aquatica, Luziola leiocarpa, and Rhynchoryza subulata) (Table 1). *Ehrharta erecta*, a species in the tribe Ehrhartoideae, which is a sister tribe to the Oryzeae, was used as an outgroup (GPWG, 2001; Guo and Ge, 2005). Plastid, mitochondrial and nuclear gene sequences have been used to establish the phylogeny of the Oryzeae (Ge et al., 1999; Guo and Ge, 2005; Tang et al., 2010) and have provided an important framework for the study of molecular evolution in this group (Figure S1, Supplementary Material).

**DNA extraction, amplification and sequencing**

Total DNA was isolated from silica-gel dried or fresh leaves as described by Ge et al. (1999). A 1-2 kb fragment of the KAO gene containing several exons and introns was obtained by using the polymerase chain reaction (PCR) in conjuncition with the forward primer KAOF (5'-CCAGGA CGTTCATGTTCCAGCAG-3') and the reverse primers KAOR1 (5'-TCGTCGCCAAGCAGTGTTC-3') and KAOR2 (5'-GCCAAGCAGTTGGCACC-3') (Figure 1). The PCR was done in a total volume of 25 µL that contained 5-50 ng of template DNA, 0.2 µM of each primer, 200 µM of each dNTP, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂ and 0.75 U of ExTaq DNA polymerase (TaKaRa, Shiga, Japan). Amplifications were done in a T gradient 96 U thermocycler (Biometra, Göttingen, Germany) as follows: 3 min at 94 °C, followed by 33 cycles of 30 s at 94 °C, 30 s at 56 °C, 2.5 min at 72 °C and a final extension at 72 °C for 10 min. Further internal primers used for sequencing were: KAO707F 5'-ACCGTCTTCTCCAGGAGAAC-3' (Tm = 61.9 °C), KAO931F 5'-GATGCACCTTCTCCACAG-3' (Tm = 57.6 °C) and KAO1478F 5'-CGTCAACATCTCCCTGTC-3' (Tm = 60 °C) (Yang et al., 2009). All of the sequences were deposited in GenBank under accession numbers EF577665-EF577670 and EU179429-EU179435 (Table 2).

**Sequence analysis**

Sequences were aligned using ClustalX v.1.81 (Thompson et al., 1997) and refined by manual adjustment based on the predicted amino acid sequence. The amino acid sequences (excluding introns) were sufficiently conserved across the 12 species to provide unambiguous alignments. We examined the possibility of sequence saturation using DAMBE v.4.5.45 (Xia and Xie, 2001). Pairwise synonymous and non-synonymous substitutions per site (d₀ and dₛ) among the 11 species were estimated for the coding regions of the KAO gene.

The extent of codon usage bias often reflects the degree of selective constraint in a gene (Sharp, 1991; Sharp et al., 1986). To measure the extent of codon usage bias, we estimated the effective number of codons (ENC) and codon bias index (CBI) using DnaSP v.4.10.9 (Rozas and Rozas, 1999). The ENC values range from 20 (only one codon is used for each amino acid, i.e., the codon bias is maximal) to 61 (all synonymous codons for each amino acid are equally used, i.e., there is no codon bias) (Wright, 1990). The CBI values range from 0 (uniform use of synonymous codons) to 1 (maximum codon bias) (Morton, 1993). Variation in the rate of synonymous substitution among genes may be related to codon use (Sharp, 1991). Therefore, several parameters related to codon usage bias, such as the GC content at the first and second codon positions (GC1, 2), as well as third codon positions (GC3), were also estimated using DnaSP v.4.10.9 (Rozas and Rozas, 1999).

**Detecting rate heterogeneity among lineages**

The relative-rate test based on the method of Muse and Gaut (1994), as implemented in Hyphy (Pond et al., 2005), was used to detect variation in the synonymous and non-synonymous substitution rates along different lineages, with *Ehrharta erecta* as the reference sequence. This method examines substitution rates between two lineages with reference to a third outgroup lineage. In the first model, the two related taxa from the most recent common ancestor are constrained to have the same substitution rate. In the second model, the two lineages may have different substitution rates. A likelihood ratio test is used to test
which of the models best explains the data (Muse and Gaut, 1994).

Detection of positive selection

The ratio $\omega (d_S/d_\text{eq})$ provides an effective means of detecting selection or selective pressure on a gene or gene region, with $\omega < 1$, $= 1$ and $> 1$ indicating negative selection, neutral evolution and positive selection, respectively (Yang, 2006). We ran likelihood-based analyses using the CODEML program of PAML 4 (Yang, 2007) to explore the selective processes acting on the KAO gene. First, we used the branch models to examine whether the evolutionary rates differed among lineages within the gene tree. The one ratio model (M0) assumes a single $\omega$ for all branches and all sites. However, the free ratio model (Mf) postulates an independent $\omega$ ratio for each branch of the tree. A likelihood ratio test (LRT) was used to decide whether there was a significant difference between M0 and Mf. The model with the higher likelihood value was assumed to be the better model (Bielawski and Yang, 2003; Yang and Nielsen, 1998).

We next used site-specific models to detect whether particular amino acid residues were subject to positive selection (Yang, 2006). The neutral model (M1a) classifies all of the sites into two categories, i.e., strict constraint ($0 < \omega < 1$) (purifying selection) and neutral ($\omega = 1$). Based on M1a, the positive selection model (M2a) assumes a third category under positive selection ($\omega > 1$). The beta model (M7) assumes a beta distribution for the $\omega$ ratios over sites, and the beta and $\omega$ model (M8) increases the independent ratio estimated by the data. M8 and M2a assume positive selection and are compared with M7 and M1a, respectively. If the LRT is significant and there is a site with $\omega > 1$ then positive selection is invoked for the gene (Bielawski and Yang, 2003; Yang, 2006).

Results and Discussion

Previous studies showed that the KAO gene was a single-copy gene (Helliwell et al., 2001; Sakamoto et al., 2004; Yamaguchi, 2008) and the loss-of-function mutant exhibits a typical phenotype, indicating the functional importance of this enzyme in GA biosynthesis (Sakamoto et al., 2004). In view of the importance of comparing orthologous rather than paralogous genes when estimating substitution rates, we initially examined this issue and found that the KAO gene was orthologous in all of the species analyzed. The similarity of the aligned coding regions ranged from 87.5% to 99.5% (Figure S2, Supplementary Material). Sequences of the KAO gene were isolated from all of the Oryzeae species and from the outgroup, Ehrharta erecta. The sequenced regions ranged in size from 1772 bp to 2626 bp and their aligned coding regions varied from 1047 bp to 1053 bp (Table 2). The total GC content and the GC content of the third position of the codons (GC3) were

<table>
<thead>
<tr>
<th>Species</th>
<th>Length sequenced (bp)</th>
<th>Coding</th>
<th>Noncoding</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oryza sativa</td>
<td>2231</td>
<td>1053</td>
<td>35.02</td>
<td>AP004572</td>
</tr>
<tr>
<td>O. meridionalis</td>
<td>2189</td>
<td>1053</td>
<td>35.60</td>
<td>EU179429</td>
</tr>
<tr>
<td>O. punctata</td>
<td>1819</td>
<td>1053</td>
<td>35.13</td>
<td>EF577665</td>
</tr>
<tr>
<td>O. officinalis</td>
<td>1833</td>
<td>1053</td>
<td>35.13</td>
<td>EF577666</td>
</tr>
<tr>
<td>O. australis</td>
<td>1853</td>
<td>1053</td>
<td>35.13</td>
<td>EF577667</td>
</tr>
<tr>
<td>O. glauca</td>
<td>1867</td>
<td>1053</td>
<td>35.13</td>
<td>EF577668</td>
</tr>
<tr>
<td>O. brachyantha</td>
<td>1887</td>
<td>1053</td>
<td>35.13</td>
<td>EF577669</td>
</tr>
<tr>
<td>O. granulata</td>
<td>1908</td>
<td>1053</td>
<td>35.13</td>
<td>EF577670</td>
</tr>
<tr>
<td>O. brachyantha</td>
<td>1929</td>
<td>1053</td>
<td>35.13</td>
<td>EF577671</td>
</tr>
<tr>
<td>Leersia tisserantii</td>
<td>1950</td>
<td>1053</td>
<td>35.13</td>
<td>EF577672</td>
</tr>
<tr>
<td>Luziola leiocarpa</td>
<td>1970</td>
<td>1053</td>
<td>35.13</td>
<td>EF577673</td>
</tr>
<tr>
<td>Chikusichloa aquatica</td>
<td>2000</td>
<td>1053</td>
<td>35.13</td>
<td>EF577674</td>
</tr>
<tr>
<td>Rhynchoryza subulata</td>
<td>2030</td>
<td>1053</td>
<td>35.13</td>
<td>EF577675</td>
</tr>
<tr>
<td>Ehrharta erecta</td>
<td>2060</td>
<td>1053</td>
<td>35.13</td>
<td>EF577676</td>
</tr>
<tr>
<td>Mean ± SE</td>
<td>1962.83 ± 1.51</td>
<td>1049.50 ± 2.24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 - Information for the KAO gene sampled in this study.
similar across species. Table 2 summarizes the sequence data for this gene.

**Codon usage bias and its correlation with GC3 and substitution rates**

Codon usage bias has been important in studies of molecular evolution because it provides examples of weak selection at the molecular level. CBI and ENC were calculated to measure the degree of codon usage bias. CBI showed a marked negative correlation with ENC ($r^2 = 0.958, p < 0.0001$) (Figure 2A) such that both CBI and ENC could be used to measure the degree of codon usage bias. In this study, ENC was used to measure the degree of codon usage bias.

To determine the relative effects of mutation pressure versus natural selection on codon composition, we examined the relationship between the GC content at third codon positions (GC3) and the GC content at the first and second codon positions (GC1,2). The GC content of GC1,2 ranged from 48.9% to 50.3%, which there was a tendency of positive correlation with GC3 ($r^2 = 0.227$) but this was not significant ($p = 0.139$) (Figure 2F). This pattern of base composition suggests that the GC content is most likely the result of mutation pressure since natural selection acts differently on different codon positions (Shackleton et al., 2006). Interestingly, after excluding *L. tisserantii*, GC1,2 showed a significant positive correlation with GC3 ($r^2 = 0.604, p < 0.05$) (data not shown), which further confirmed that these changes were most likely the result of mutation pressure. $d_s$ was positively correlated with $d_N$ ($r^2 = 0.498, p < 0.05$) (Figure 2D), as also observed in other organisms (Bielawski et al., 2000; Dunn et al., 2001; Hurst and Williams, 2000; Kusumi et al., 2002), and negatively correlated with codon bias ($r^2 = 0.713, p < 0.05$) (Figure 2B) and GC3 ($r^2 = 0.796, p < 0.001$) (Figure 2E). The negative correlation between $d_s$ and codon usage bias may be explained by natural selection (Bielawski et al., 2000; Smith and Eyre-Walker, 2001; Urrutia and Hurst, 2001) since codon usage bias is a primary factor in $d_s$ variation among genes and is thought to be under natural selection, perhaps because of the need to maintain accuracy or speed in translation (Yang and Gaut, 2011). There was also a tendency for $d_N$ being negatively correlated with codon usage bias ($r^2 = 0.348$) but this was not significant ($p = 0.056$) (Figure 2C). The latter would be consistent with sites that are functionally constrained and consequently conserved at the amino acid level. Such sites are also likely to experience stronger selection for translation accuracy and hence have a higher codon bias (Akashi, 2003). This might explain the negative correlation between $d_N$ and codon bias observed here (though not significant), and by others in enteric bacteria (Rocha, 2004; Sharp, 1991), *Drosophila* (Betancourt and Presgraves, 2002), yeast (Drummond et al., 2005), and viruses (Duffy et al., 2008). The fact that $d_N$ is correlated to codon bias suggests that codon bias might be used as a measure of the level of constraint upon a site or gene (Plotkin et al., 2004, 2006; Stoletzki and Eyre-Walker, 2007).

**The driving forces governing evolution of the KAO gene in Oryzeae**

A codon-based approach showed that the free ratio model (Mf) had significantly higher likelihood scores (ln4103.38) than the one ratio model (M0) (ln4124.44) ($p < 0.001$) (Table 3). Although the $d_N/d_S$ ratios varied across lineages from 0.0001 to 0.358 (with one of the 21 lineages showing no predicted synonymous substitutions, i.e., the $d_N/d_S$ ratio was equal to 999.000), the estimated $d_N/d_S$ ratio for each lineage was less than 1. The $o_0$ values were estimated to be 0.079 under the M0 model, suggesting that purifying selection or selection constraint best explained the molecular evolution of the *KAO* gene, in agreement with the studies on anthocyanin pathway genes (Lu and Rausher, 2003; Rausher et al., 2008).

The branch model test is a very conservative test of positive selection because it averages the ratio across all sites. We therefore used site-specific codon models to examine whether there was positive selection on codon sites. The M2a and M8 models, which assume positive selection, were not significantly better than the null models M1a and

---

**Figure 2** - The relationships between effective number of codons (ENC) and codon bias index (CBI) (A), synonymous substitution rates ($d_s$) (B), and non-synonymous substitution rates ($d_N$) (C), between $d_s$ and $d_N$ (D) and third codon positions (GC3) (E), and between the first and second codon positions (GC1, 2) and GC3 (F).
M7 (for M1a vs. M2a, 2ΔL = 0, p = 1.0; for M7 vs. M8, 2ΔL = 0, p = 1.0) (Table 3). These results indicate that the KAO gene is under strong selective constraint, thus ruling out the possibility of past episodes of positive selection on this gene. Previous studies have shown that variation in the evolutionary rate among nucleotide sites may be attributed to differences in the frequency of positive selection (Yang et al., 2000; Gaut et al., 2011) or in the magnitude of selective constraints (Li, 1997; Rausher et al., 1999, 2008).

In this study, the branch and codon models failed to detect any sign of positive selection for any lineage and codon of the KAO gene, suggesting that the significant heterogeneity of some lineages was attributable mainly to the relaxed constraint among lineages or neutral processes rather than positive selection. However, the power to detect positive selection using the methods mentioned above may be low, especially when adaptive substitutions are spread across many amino acid sites (Pond et al., 2005; Rausher et al., 2008). Further investigations with alternative tests on intraspecific changes (Olsen et al., 2002; Whitt et al., 2002; Flowers et al., 2007; Rausher et al., 2008) would be necessary to detect evidence of positive selection.

Rate variation among lineages

There was significant heterogeneity in the synonymous and non-synonymous substitution rates of the KAO gene among lineages of the rice tribe (Table 4), especially in C. aquatica and L. leiocarpa. Among 55 relative-rate tests for synonymous substitutions, 11 comparisons were significant at the 5% or 1% level. At the same time, among 55 relative-rate tests for non-synonymous substitutions, the null hypothesis of rate homogeneity was rejected for 18 comparisons. In C. aquatica and L. leiocarpa, dN appeared to be decelerated, and dS in C. aquatica. The significant slowdown in the rate of synonymous and non-synonymous substitutions in C. aquatica and L. leiocarpa lineages may reflect differences in the intensity of selection, i.e., the KAO gene may be under different functional constraints in different lineages.

Several mechanisms could explain the observed rate heterogeneity, including life history traits such as genera-

Table 3 - Log likelihood values, ω ratios and parameter estimates for the KAO gene in models with variable ω ratios among codon sites.

<table>
<thead>
<tr>
<th>Model</th>
<th>ω0</th>
<th>ln Parameter estimates</th>
<th>Models compared</th>
<th>2ΔL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mf</td>
<td>3.1</td>
<td>-4103.38</td>
<td>ω0 = 0.0001 – 999.000, tree length = 2.140, kappa (ts/tv) = 1.103</td>
<td>M0-Mf</td>
<td>42.12</td>
</tr>
<tr>
<td>M0</td>
<td>2.3</td>
<td>-4124.44</td>
<td>ω0 = 0.079, tree length = 2.181, kappa (ts/tv) = 1.082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1a</td>
<td>2.4</td>
<td>-4067.60</td>
<td>ω0 = 0.049, ω1 = 0.921, ω2 = 1.000, p0 = 0.079</td>
<td>M1a-M2a</td>
<td>0</td>
</tr>
<tr>
<td>M2a</td>
<td>2.6</td>
<td>-4067.60</td>
<td>ω0 = 0.921, ω1 = 0.053, ω2 = 0.026, p2 = 1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>2.4</td>
<td>-4061.19</td>
<td>p = 0.282, q = 2.548</td>
<td>M7-M8</td>
<td>0</td>
</tr>
<tr>
<td>M8</td>
<td>2.6</td>
<td>-4061.19</td>
<td>p0 = 1.000, p = 0.282, q = 2.548; p1 = 0.000, ω = 8.931</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 p – number of parameters, ln – log-likelihood values of the data in each model. 2 Parameter estimates in different models. 3 Tree length is the sum of branch lengths.

Table 4 - Results of 110 relative-rate tests for dS (lower triangle) and dN (upper triangle). Rejection of rate equality is indicated by * at the 0.05 level, ** at the 0.01 level, or *** at the 0.001 level. Ehrharta erecta was used as the outgroup in all comparisons. Species names that were inferred to have evolved more quickly in each pairwise comparison are indicated in the table by the first letter of the genus name and the first three letters of the species name.

<table>
<thead>
<tr>
<th></th>
<th>Osat</th>
<th>Omer</th>
<th>Opun</th>
<th>Ooff</th>
<th>Oaus</th>
<th>Obra</th>
<th>Ogra</th>
<th>Ltis</th>
<th>Llei</th>
<th>Caqu</th>
<th>Rsub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***Osat</td>
<td></td>
</tr>
<tr>
<td>Omer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***Omer</td>
<td>***Omer</td>
</tr>
<tr>
<td>Opun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***Opun</td>
<td>***Opun</td>
</tr>
<tr>
<td>Ooff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**Ooff</td>
<td>**Ooff</td>
</tr>
<tr>
<td>Oaus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**Oaus</td>
<td>**Oaus</td>
</tr>
<tr>
<td>Obra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*Obra</td>
<td></td>
</tr>
<tr>
<td>Ogra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*Ogra</td>
<td></td>
</tr>
<tr>
<td>Ltis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***Ltis</td>
<td>*Ltis</td>
<td></td>
</tr>
<tr>
<td>Llei</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caqu</td>
<td>***Osat</td>
<td>***Omer</td>
<td>***Opun</td>
<td>**Ooff</td>
<td>**Oaus</td>
<td>**Obra</td>
<td>***Ogra</td>
<td>***Ltis</td>
<td>***Llei</td>
<td>*Rsub</td>
<td></td>
</tr>
<tr>
<td>Rsub</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*Ogra</td>
<td>*Rsub</td>
</tr>
</tbody>
</table>

tion time, biochemical features such as efficiency of DNA repair machinery, and environmental variables such as energy and temperature (Eyre-Walker and Gaut, 1997; Li, 1997; Brown et al., 2005; Soria-Hernanz et al., 2008). Rate heterogeneity may also result from differences in population size since variation in population size can alter evolutionary rates within a lineage (Eyre-Walker and Gaut, 1997; Lynch and Conery, 2003) and vice versa. Variation in the nucleotide substitution rates of the \(KAO\) gene significantly changed the \(\omega\) ratios of the respective lineages. These features of the \(KAO\) gene in Oryzeae resulted from the influence of various factors that affected the evolution of these species and their ancestors. A detailed knowledge of these factors will help us to understand the evolutionary history of Oryzeae species.

Conclusions

The results of this study showed that codon usage bias was negatively correlated with synonymous and non-synonymous substitution rates, a finding consistent with the importance of codon usage. CBI was positively correlated with ENC, thus confirming the similarity of CBI and ENC as parameters for measuring the degree of codon usage bias. There was considerable heterogeneity in the nucleotide substitution rates of the \(KAO\) gene and this significantly affected the \(\omega\) ratios of the respective lineages. There was no positive selection and no positively selected codons in this gene, a finding indicative of substantial selective constraint. These features of nucleotide substitutions in the \(KAO\) gene reflected the influence of various factors on the evolution of many Oryzeae species and their ancestors.

Acknowledgments

YYH thanks his instructor, Prof. Song Ge, for help with the design and execution of this work. We thank Fu-Min Zhang, Hong-Zheng Sun and other members of Dr. Ge’s group for their help during the experimental work and data analysis. We thank the International Rice Research Institute (Los Banos, Philippines) for providing the seed samples. This work was supported by the National Basic Research Program of China (2007CB815704), the Ministry of Agriculture Transgenic Major Project (2009ZX08012-018B) and the Scientific Research Promotion Fund for the Talents of Jiangsu University (11JDG049).

References


O. punctata_kao
GACGACCACCGCCGCATCCGCAAGCTCACCGCCGCCCCCATCAACGGCTTCGACGCCCTC
O. officinalis_kao
GACGACCACCGCCGCATCCGCAAGCTCACCGCCGCCCCCATCAACGGCTTCGACGCCCTC
O. australiensis_kao
GACGACCACCGCCGCATCCGCAAGCTCACCGCCGCCCCCATCAACGGCTTCGACGCCCTC
O. brachyantha_kao
GACGACCACCGCCGCATCCGCAAGCTCACCGCCGCCCCCATCAACGGCTTCGACGCCCTC
O. granulata_kao
GACGACCACCGCCGCATCCGCAAGCTCACCGCCGCCCCCATCAACGGCTTCGACGCCCTC
L. tisserantti_kao
GACGATCACCGGCGAATCCGCAAGCTGACGGCGGCGCCGATCAACGGCTTCGACGCCCTC
C. aquatica_kao
GAGGAACACCGGCGCTTGCGCAAGCTCACGGCGGCGCCCATCAACGGCTTCGAGGCCCTC
R. subulata_kao
GAGGATCACCGGCGGCTTGCGCAAGCTCACGGCGGCTCCCATCAACGGCTTCGAGGCCCTC
L. leiocarpa_kao
GACGACCACCGCCGCTCCGCAAGCTCACGGCGGCGCCCATCAACGGCTTCGACGCCCTC
E. erecta_kao
GACGACCACCGCCTCTGCGCAAGCTGACCCGCGGCGCCAATCAACGGCTTTGATTCCTC
G
Clustal Consensus

130 140 150 160 170 180

130

O. sativa_kao
ACCACCTACCTCTCCTTCATCGACCAGACCGTCGTCGCCCTCCCTCCGCCGCTGGTCCTCG

O. meridionalis_kao
ACCACCTACCTCTCCTTCATCGACCAGACCGTCGTCGCCCTCCCTCCGCCGCTGGTCCTCG

O. punctata_kao
ACCACCTACCTCTCCTTCATCGACCACACCGTCGTCTCCACGCTCCGCCGCTGGTCCTCG

O. officinalis_kao
<table>
<thead>
<tr>
<th></th>
<th>O.australiensis_kao</th>
<th>O.brachyantha_kao</th>
<th>O.granulata_kao</th>
<th>L.tisserantti_kao</th>
<th>A.aquatica_kao</th>
<th>R.subulata_kao</th>
<th>L.leiocarpa_kao</th>
<th>E.erecta_kao</th>
<th>Clustal Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCAAACCCGTCGTCGCCACTCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
<td>ACCACGACCGTCGTCGCTACGCTCCGCCGCTGGTCGGAG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
<tr>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACAACCTA CCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
<td>ACCACCTACCTCGGCTTCATCG</td>
<td>ACCACCTACCTGGGCTTCATCG</td>
<td>ACCACCTACCTCTCCTTCATCG</td>
<td>ACCACCTACCTCGCCTTCATCG</td>
</tr>
</tbody>
</table>

Clustal Consensus

<table>
<thead>
<tr>
<th>190</th>
<th>200</th>
<th>210</th>
<th>220</th>
<th>230</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>....</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

O.sativa_kao
CCGGAGTCCGGCCAGGTCGAGTTCCTCACCGAGCTCAGGCGCATGACCTTCAAGATCATC

O.meridionalis_kao
CCGGAGTCCGGCCAGGTCGAGTTCCTCACCGAGCTCAGGCGCATGACCTTCAAGATCATC

O.punctata_kao
CCGGACCTCCTCGGCGAGTTCGAGTTCCTCACCGAGCTCAGGCGCATGACCTTCAAGATCATC

O.officinalis_kao
CCGGACCTCCTCGGCGAGTTCGAGTTCCTCACCGAGCTCAGGCGCATGACCTTCAAGATCATC

O.australiensis_kao
CCGGAGTCCGGCCAGGTCGAGTTCCTCACCGAGCTCAGGCGCATGACCTTCAAGATCATC

O.brachyantha_kao
GAGCCAGCCGGGAGGTCGAGTTCCTCACCGAGCTGCGGCGGATGACCTTCAAGATCATC
O.granulata_kao
CCCCGCCAGGGCGAGGTCGAGTTCCTCACCGAGCTCAGGCAGCGCATGACGTTCAAGATCATC
L.tisserantti_kao
TCCGGGAGGTGGAGTTCCTCACGGAGCTGCGGCGGATGACCTTCAAGATCATC
C.aquatica_kao
E.erecta_kao
GGGGGGAAGGAGATGGAGTTGATCCTCACGGAGATGCGGCGGATGAACTTCAAGGATCATC
R.subulata_kao
GCCGGCGAGGTGGAGTTCCTCACGGAGCTGCGGCGGATGACGTTCAAGATCATC
L.leiocarpa_kao
GGCGCCGGCGAGGTCGAGTTCCTCACCGAGCTCCGCCGCATGACGTTCAAGATCATC
E.erecta_kao
GGGGGGAAGGAGATGGAGTTGATCCTCACGGAGATGCGGCGGATGAACTTCAAGGATCATC
Clustal Consensus

<table>
<thead>
<tr>
<th></th>
<th>250</th>
<th>260</th>
<th>270</th>
<th>280</th>
<th>290</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

O.sativa_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.meridionalis_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.punctata_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.officinalis_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.australiensis_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.brachyantha_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
O.granulata_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
L.tisserantti_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
C.aquatica_kao
GTCCAGATTTCATGAGCGGCGCCGCAGACCGACCATTGGAGGCTCATGGAGCCGGAGCTAC
**Clustal Consensus**

```
<table>
<thead>
<tr>
<th></th>
<th>550</th>
<th>560</th>
<th>570</th>
<th>580</th>
<th>590</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td>O. sativa_kao</td>
<td>CTCATCATGTAACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. meridionalis_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. punctata_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. officinalis_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. australiensis_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. brachyantha_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. granulata_kao</td>
<td>CTCATCATGTACCTCAACGCGGCCAACGAAGTCCTCCGGCCACATCACCATGTGGGCA</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
**Clustal Consensus**  
*** ******** ************** ** ** ***  ** ************** **

610 620 630 640 650 660
...|...|...|...|...|...|...|...|...|...|...|...|

**O. sativa_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. meridionalis_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. punctata_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. officinalis_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. australiensis_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. brachyantha_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA TC

**O. granulata_kao**
GTCTT CCTCGAGAGAAACCCGACATCCTCGCAAGAGAGGCTGAGCAAGAGGAGAGA ATC

**L. tisserantti_kao**
GTCTT CCTCGAGAGAAATCCGACATCCTGCAAGAGGCAAGGCTGAGCAAGAGGAGAGA TC

**C. aquatica_kao**
GTCTT CCTCGAGAGAAACCCGAAATCCTTGCAAGGCAAGGCGAGCAAGAGGAGAGA TC
\textbf{R. subulata\_kao}

\texttt{GTCTTCCTCCAGGAGAACCCCGAAATCTTCGCAAGGGCAAAGGCCGAGCAAGAGGAAA}

\texttt{TC}

\textbf{L. leiocarpa\_kao}

\texttt{GTCTTCCTGCGAGGAACCCCGACATCTTCGCGAGGGCAAAGGCCGAGCAAGAAGAGATC}

\textbf{E. erecta\_kao}

\texttt{TTTTTTCCCTGCAAGGAACCCCGGACGTATTTAGCAAAAGGGAAGGCCGAGCAAGAGAGA}

\texttt{TC}

\textbf{Clustal Consensus} \hspace{1cm} * ***** ** ***** ** **  * ** ** *  ** ** ** *  ***** ** ***

\begin{tabular}{cccccccc}
670 & 680 & 690 & 700 & 710 & 720 \\
\texttt{....|....|....|....|....|....|....|....|....|....|....|....|} \\
\end{tabular}

\textbf{O. sativa\_kao}

\texttt{ATGAGAAGCATTTCCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AC}

\textbf{O. meridionalis\_kao}

\texttt{ATGAGAAGCATTTCCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AC}

\textbf{O. punctata\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AT}

\textbf{O. officinalis\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AC}

\textbf{O. australiensis\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{GC}

\textbf{O. brachyantha\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AG}

\textbf{O. granulata\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AG}

\textbf{L. tisseranti\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AC}

\textbf{C. aquatica\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AC}

\textbf{R. subulata\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGCA}

\texttt{T}

\textbf{L. leiocarpa\_kao}

\texttt{ATGAGAAGCATACCAGCAACGCAGAACGGATTAACCCTTCAAGGACTTTAAGAAGATGC}

\texttt{AG}
E. erecta_kao  ATGAGAAGC---------------------------CTGAGGGATTTCAGGAAGATGGAG
Clustal Consensus ****** *                        ** ***** **** ******

730 740 750 760 770 780

O. sativa_kao
TTCTCTCACAGGTTTGTCAAGGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
O. meridionalis_kao
TTCTCTCACAGGTTTGTCAAGGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
O. punctata_kao
TTCTCTCACAGGTTTGTCAAGGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
O. officinalis_kao
TTCTCTCACAGGTTTGTCAAGGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
O. australiensis_kao
TTCTCTCACAGGTTTGTCAAGGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
O. brachyantha_kao
TTCTCTCCCAGGTTATCGACGAGACACTTCGCTGCGTCAACATCTCGTTTGTGTCCTT
C
O. granulata_kao
TTCTCTCGCAGGTTATCGACGAGACGCTTCGCTGCGTCAACATCTCGTTTGTGTCCTT
C
L. tisseranti_kao
TTCTCCAGCAGGTTATCGACGAGACGCTTCGCTGCGTCAACATCTCGTTTGTGTCCTT
C
C. aquatica_kao
TACCTTCACAGGTTTATCGACGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
R. subulata_kao
TACCTTCACAGGTTTATCGACGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
L. leiocarpa_kao
TACCTTCACAGGTTTATCGACGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
E. erecta_kao
TACCTTCACAGGTTTATCGACGAGACACTTTGCTGCGTCAACATCTCCTTTGTGTCCTT
C
Clustal Consensus  * *****  ** *** ***** ****** ********** ** **** ***

790 800 810 820 830 840

O. sativa_kao
CGTCAGGCCAACAGAGACATCTTTGTGAACGGTATCTTATCCCCAAGGGGTGGAAGGTT
Clustal Consensus

910  920  930  940  950  960

O.sativa_kao
AACCCTTCAGATGGGAGGGACCACCTCCGAAAGCCGGAACATTCCTTCCATTTGGAC

O.meridionalis_kao
AACCCTTCAGATGGGAGGGACCACCTCCGAAAGCCGGAACATTCCTTCCATTTGGAC

O.punctata_kao
AACCCTTCAGATGGGAGGGACCACCTCCGAAAGCCGGAACATTCCTTCCATTTGGAC

O.officinalis_kao
AACCCTTCAGATGGGAGGGACCACCTCCGAAAGCCGGAACATTCCTTCCATTTGGAC

O.australiensis_kao
AACCCTTCAGATGGGAGGGACCACCTCCGAAAGCCGGAACATTCCTTCCATTTGGAC
O. brachyantha_kao
AACTCCATCAAGATGGGAGGGTCCCCCGCCGAGAGCCGGAACATTCCTTCCATTTGGAC
TG
O. granulata_kao
AACCCCATCAAGATGGGAGGGTCCCCCGCCGAGAGCTGGAACATTCCTTCCATTTGGAC
TG
L. tisseranti_kao
AACCCCATCAAGATGGGAGGGTCCTCCGCCGAGAGCCGGAACATTCCTTCCATTTGGAC
TG
C. aquatica_kao
AACCCCTTCAAGATGGGAGGGTCCCCCGCCGAGAGCCGGAACATTCCTTCCATTTGGAC
TG
R. subulata_kao
AACCCCTTCAAGATGGGAGGGTCCCCCGCCGAGAGCCGGAACATTCCTTCCATTTGGAC
TG
L. leiocarpa_kao
AACCCCTTCAAGATGGGAGGGTCCCCCGCCGAGAGCCGGAACATTCCTTCCATTTGGAC
TG
E. erecta_kao
AAGCCTTCAAGATGGGATGGTTCTTCACCGAGAGCCGGAACGTTCCTTCCATTCGGACT
G

Clustal Consensus

970       980       990       1000      1010    1020

O. sativa_kao
GGAGCAAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
O. meridionalis_kao
GGAGCAAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
O. punctata_kao
GGATCGAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
O. officinalis_kao
GGAGCAAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
O. australiensis_kao
GGAGCGAGGCTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
AT
O. brachyantha_kao
GGAGCGAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
O. granulata_kao
GGAGCGAGACTGTGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCTCCA
T
Figure S2 Alignment of coding sequences of KAO gene in twelve species. Highly conserved sites are indicated with asterisks in the bottom.