Molecular Cloning and Bioinformatic Analysis of Endosperm Specific Promoter, α-Globulin (AsGlo1)

Abu Bakar Sadiq Mustaffa, Alina Wagiran
*Department of Biotechnology and Medical Engineering, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
*Corresponding author: alina@utm.my

Abstract
The α-globulin promoter (AsGlo1) have been successfully amplified from pmCACA:GFP using PCR. The construction of the recombinant plasmid, pMRGpro containing α-globulin promoter was achieved by ligation at HindIII and BamHI of pMR104a. Cassette containing α-globulin promoter and nos terminator was successfully cloned in pCAMBIA1305.2 and designated as pCAMGpro. Digestion methods were used to confirm the recombinant plasmid. The resulting plasmid was then sequenced and analyzes using bioinformatics tools. The analysis showed a 99% homology with A. sativa α-globulin gene promoter (Gene Bank: AY795082.1) and further confirmed using phylogenetic analysis. The result showed a successful cloning of the correct fragment Avena sativa α-globulin promoter into an expression vector.

Keywords: Rice endosperm; recombinant protein; endosperm specific promoter; α-globulin promoter

1.0 INTRODUCTION
Endosperm specific promoters play an important role in the genetic modification of endosperm composition and in using endosperm as bioreactors. Although several endosperm specific promoters have been cloned and used in expression studies, their origin of gene is similar with the transformed plant species. Since α-globulin promoter (AsGlo1) from oat has been shown to drive high expression in barley, it would be interesting to know its effect in rice.

In rice, globulin promoter has been shown to increase targeted protein in endosperm (Hwang et al., 2002). The use of tissue-specific promoters have several advantages such as potential to increase recombinant protein stability and deliver targeted gene expression to only specific parts of the plant (Facy, 2009), cost effective and easy agricultural scale-up, high rate of protein synthesis and easy long term storage of recombinant protein (Takaiwa et al., 2007). Different research outcomes have illuminated the importance and role of several endosperm-specific promoter to engineer transgenic crops with improved endosperm-specific biosynthesis (Furtado et al., 2009, Kawakatsu and Takaiwa, 2010, Vickers et al., 2006). The use of endosperm-specific promoters presents an approach that could solve the potential detrimental effect of non-specific constitutive promoters to the host plant since the foreign gene is continuously being expressed in the entire plant tissues (Qu et al., 2008). The
endosperm-specific expression promoters of other cereals have also received attention from various researchers such as from maize (Zea mays L.) (Hu et al., 2011), barley (Hordeum vulgare L.) (Choi et al., 2003), rice (Qu and Takaiwa, 2004, Rasmussen and Donaldson, 2006) and wheat (Triticum aestivum L.) (Lamacchia et al., 2001). Therefore, the present study focused on the development of an expression cassette containing endosperm specific promoter (α-globulin) from Avena sativa and analyzing the sequences through bioinformatics analysis and the construction of phylogenetic tree.

### 2.0 MATERIALS AND METHODS

A α-globulin promoter, AsGlo1 coding region from Avena sativa was amplified from 4.5 kb plasmid (pmCACA:GFP) was obtained from Dr. Claudia Vickers (Queensland University, Australia) and cloned into an intermediate vector, pMR104a (UKM) containing a NOS terminator. The isolated α-globulin promoter were excised using HindIII and BamHI by T4 DNA ligase was used to replace the CaMV35S coding region in pMR104a. Amplification of the promoter was achieved by PCR (Promega) using primers F: ‘5'-CACAAACGTGCAAAGCTTAAATTCG-3' and primers R: 5'-GACGGATCCGAGATTGTAGAAGG-3'. The forward primers contained engineered HindIII site whereas reverse primers contained BamHI site. PCR was performed under the following conditions; initial denaturation at 94°C at 3 min; (30 cycles), denaturation at 94°C for 1 min; annealing at 55°C for 1 min; extension at 72°C for 1 min and final extension at 72°C at 10 min. The amplified α-globulin promoter was excised using HindIII and BamHI, and ligated by T4 DNA ligase to replace CaMV35S coding region in pMR104a, producing a new recombinant plasmid called pMRGpro (Figure 1). The new recombinant plasmid contained AsGlo1 fragment was then transformed into E.coli competent cells and Qiagen Spin Miniprep kit was used to isolate the plasmid. The cassette containing AsGpro:NOS cassette (~1,111 bp) were then digested with the HindIII and EcoRI and ligated into the expression vector (pCAMBIA1305.2) at the same restriction enzymes. The final expression vector was designated as pCAMGpro. The confirmation of recombinant plasmid was obtained through digestion and PCR. All the PCR mixture was obtained from Promega while restriction enzymes from NEB or otherwise stated. Finally, the AsGlo1 fragment was sequenced and analyzed by bioinformatics. The forward sequence was aligned with the reverse complement of the reverse sequence using the ClustalX alignment mode of Bioedit software. Phylograms of the selected sequences with highest homology to AsGpro was constructed using MEGA5 software (Tamura et al., 2011) for construction of phylogenetic tree.

![Figure 1](image_url)
subjected to double digestion with enzyme HindIII and EcoRI generated 2 fragments; pMRGpro (lane 6) and pMR104a (lane 7). From Figure 2A lane 6 and lane 7 it can be observed that 2 fragments were generated for both pMRGpro and pMR104a. From the double digestion, pMRGpro produced the upper fragment (2,418 bp) representing the pMRGpro backbone without the AsGpro:NOS cassette while ~1111 bp for AsGlo1:NOS cassette (lane 6). While for pMR104a, the upper fragment (2,418 bp) represented the pMR104a backbone while ~1174 bp represented CaMV35S:NOS cassette (lane 7). The cassette difference in AsGlo1:NOS (1,111 bp) and the CaMV35Spro:NOS (1,174 bp), confirms the replacement of the CaMV35Spro with the AsGlo1.

**Figure 2** Gel electrophoresis (1% w/v) shows (A) confirmation of recombinant plasmid, pMRGpro. Lane 1 & 2: Single fragment of pMRGpro and pMR104a digested with HindIII respectively, Lane 3: Undigested pMR104a, Lane 4: Fragment AsGlo1 promoter (below) from pMRGpro when digested with HindIII and BamHI, Lane 5: Two fragments show pMR104a without CaMV35S (upper) and fragment CaMV35S (below) when digested with HindIII and BamHI, Lane 6: pMRGpro digested with HindIII and EcoRI, Lane 7: pMR104a digested with HindIII and EcoRI, Lane 8: Undigested pMRGpro and M: 1 Kb ladder. (B) Digestion of 2 colonies pCAMGpro with HindIII and EcoRI. Lane 1 and 2: pCAMBIA1305.2Gpro digested with HindIII and EcoRI, Lane 3 and 4: Undigested pCAMBIA1305.2Gpro. M1: Lambda DNA-HindIII digest and M2: 1 Kb DNA marker.

The construction of final expression vector was achieved by ligation of AsGlo1:NOS cassette and pCAMBIA1305.2 at site of HindIII and EcoRI. From Figure 2B, digestion with HindIII and EcoRI generated two bands (lane 1 and 2) with upper band (~11,870 bp) representing the pCAMGpro backbone without the cassette (pCAMBIA1305.2) while the lower band is the AsGlo1:NOS cassette (~1,111 bp). The amplification of the GUSPlus gene was performed using the forward primer, GUS_F (5’CGCCGATGCAGATATTCGTA3’) and the reverse primer, GUS_R (5’ ATTAATGCGTGGTCGTGCAC 3’) at an annealing temperature of 58°C for 1 min and gave the expected band size of approximately 789 bp. This confirms that the pCAMGpro recombinant plasmid was formed from the pCAMBIA1305.2 backbone containing the GUS Plus gene (Figure 3).
3.2 Bioinformatic Analysis and Phylogenetic Tree

The present study shows amplified sequence was at highest query coverage (98%) and sequence homology to *A. sativa* (AsGlo1) gene, promoter region (Acc: AY795082.1) and *A. sativa* 12S α-globulin seed storage protein gene, complete cds (Acc: J05485.1) (Table 1).

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Name of sequences</th>
<th>Query Coverage</th>
<th>Maximum Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY795082.1</td>
<td><em>A. sativa</em> (Glo1) gene, promoter region</td>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td>J05485.1</td>
<td><em>A. sativa</em> 12S α-globulin seed storage protein gene, complete cds</td>
<td>98%</td>
<td>96%</td>
</tr>
<tr>
<td>X68648.1</td>
<td><em>A. sativa</em> pseudogene for 12S seed α-globulin</td>
<td>61%</td>
<td>86%</td>
</tr>
</tbody>
</table>

The optimal tree with the sum of branch length = 0.53005192 is shown in Figure 4. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsentein, 1985).

From Table 2, the pairwise distance alignment shows AsGlo1 promoter is a closest relative to the *A. sativa* α-globulin (AsGlo1) gene promoter region (Gene Bank Accession Number: AY795082.1) with a pairwise distance of 0.003. The sequences were divided into two broad groups, 11S and 12S globulin families. This confirms that *A. sativa* α-globulin (AsGlo1) gene promoter (AY795082.1) region is the source of the targeted promoter and shows that our present studies successfully managed to clone the right fragment of AsGlo1 promoter. The alignment also confirms the AsGlo1 promoter belongs to the 12S
globulin protein family, which is the closest relative of the AsGlo1. The second relative of the α-globulin promoter is Avena sativa 12S α-globulin seed storage protein gene complete cds (Gene Bank Accession Number: J05485.1) with a pairwise distance of 0.034. This further confirms the targeted AsGlo1 fragment formed one cluster, indicating their close relationship.

Table 2 The distance pairwise alignment of AsGlo1 and the six relatives

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. sativa GLAVI gene for 11S globulin (X74740.1)</td>
<td>0.007</td>
<td>0.005</td>
<td>1.586</td>
<td>0.661</td>
<td>0.662</td>
<td>0.686</td>
<td></td>
</tr>
<tr>
<td>A. sativa (GLAVI2) gene partial cds (DQ388881.1)</td>
<td>0.026</td>
<td>0.005</td>
<td>1.547</td>
<td>0.693</td>
<td>0.694</td>
<td>0.718</td>
<td></td>
</tr>
<tr>
<td>A. sativa GLAV3 gene for 11S globulin (X74740.1)</td>
<td>0.013</td>
<td>0.015</td>
<td>1.573</td>
<td>0.667</td>
<td>0.668</td>
<td>0.692</td>
<td></td>
</tr>
<tr>
<td>A. sativa pseudogene for 12S seed globulin (X86848.1)</td>
<td>0.687</td>
<td>0.676</td>
<td>0.683</td>
<td>1.837</td>
<td>1.841</td>
<td>1.909</td>
<td></td>
</tr>
<tr>
<td>AsGpro</td>
<td>0.314</td>
<td>0.325</td>
<td>0.320</td>
<td>0.797</td>
<td>0.002</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>A. sativa (Glo1) gene promoter region (AY795082.1)</td>
<td>0.316</td>
<td>0.325</td>
<td>0.321</td>
<td>0.795</td>
<td>0.003</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>A. sativa 12S globulin seed storage protein gene complete cds (J05485.1)</td>
<td>0.317</td>
<td>0.326</td>
<td>0.323</td>
<td>0.816</td>
<td>0.034</td>
<td>0.031</td>
<td></td>
</tr>
</tbody>
</table>

#### 4.0 CONCLUSION

The new recombinant plasmid, pCAMGpro (~12,870 bp) together with any gene of interest can be used for high expression of any protein of interest in the endosperm of cereal crops. Endosperm is the most valuable plant-based production system for the production of recombinant proteins over other tissues because it is more cost-effective, it is easier to scale-up agricultural yield, provides a larger storage ability and safe long-term storage.

Acknowledgement

Special thanks to Universiti Teknologi Malaysia (UTM) for financial support in this research. Special thanks also to Dr. Claudia Vickers from Queensland University, Australia for the plasmid, pmCACA:GFP.

References


