Isolation and characterization of the aconitate hydratase 4 (Aco4) gene from soybean

Z. Coleman, J. Boelter, K. Espinosa, A.S. Goggi, R.G. Palmer, and D. Sandhu

Abstract: Aconitase catalyzes the reversible isomerization of two tricarboxylic acids, citrate and isocitrate, during the Krebs cycle. Five aconitase genes, namely, Aco1, Aco2, Aco3, Aco4, and Aco5, have been identified in soybean. Previously, Aco4 was mapped on chromosome 11. The purpose of this investigation was to isolate and sequence the candidate gene for Aco4. We mapped the Aco4 gene to a 148 kb region on chromosome 11 that contained 19 predicted genes. One of these, Glyma.11G080600, codes for aconitate hydratase. Sequencing of two isozyme variants (A-line and B-line) for Glyma.11G080600 revealed three synonymous and two non-synonymous substitutions. Perhaps, the two non-synonymous substitutions resulted in a variable isozyme pattern between the variants. Glyma.11G080600 contains a catalytic domain and a swivel domain that are known to catalyze isomerization of citrate to isocitrate and swiveling conformational change in the enzyme mechanism, respectively. Conservation of both the domains implies the role of Glyma.11G080600 in the interconversion of citrate and isocitrate. Glyma.11G080600 is expressed in most tissues, with maximum expression in leaves. Phylogenetic analysis of 25 genes from different species displayed three major clusters. Glyma.11G080600 (Aco4) and Glyma.01G162800, which are conserved in Glycine max (L.) Merr. and Glycine soja Siebold. & Zucc., may have common ancestry before G. max and G. soja split.

Key words: soybean, aconitase, linkage mapping, isozyme, Aco4.


Mots-clés : soja, aconitase, cartographie des liaisons, isozyme, Aco4.

Introduction

The aconitase gene produces an essential enzyme that plays a key role in the Krebs cycle (Gangloff et al. 1990; Zhou and Ragan 1995), which is an important process that is required by all aerobic organisms to produce adenosine triphosphate, or energy, which is required to...
power cells. Aconitase catalyzes the reversible isomerization of two tricarboxylic acids, citrate and isocitrate (Chenicke and Hart 1987; Gangloff et al. 1990; Peyret et al. 1995; Zhou and Ragan 1995). Citrate is converted into isocitrate via an intermediate product, cis-aconitate. The enzyme is a single cluster or labile made up of iron and sulfur (Rouault and Klausner 1996). The inactive form of aconitase is 3Fe–4S, which prohibits the enzyme from converting tricarboxylic acids (Zhou and Ragan 1995; Sadka et al. 2000). The active form has a single Fe added that changes the structure to 4Fe–4S and allows the conversion of citrate to isocitrate (Zhou and Ragan 1995; Rouault and Klausner 1996; Sadka et al. 2000). Aconitase protein is one of the few proteins that contains an iron–sulfur cluster but does not contain an electron transport function (Zhou and Ragan 1995).

Aconitases are present in a wide variety of living organisms and are highly conserved (Gangloff et al. 1990). Considerable work has been done on studying the multitude of purposes and uses of aconitase genes. There are two different isoforms of aconitase found in cells: mitochondrial aconitase and cytosolic aconitase. Mitochondrial aconitases are a fundamental part of the Krebs cycle (Sadka et al. 2000; Schnarrenberger and Martin 2002; Moeder et al. 2007; Terol et al. 2010). Although cytosolic aconitases are under-researched as compared with mitochondrial aconitases, they are thought to be involved in numerous processes such as cytosolic citrate metabolism and the glyoxylate cycle (Hayashi et al. 1995; Eprintsev et al. 2015). Mitochondrial and cytosolic aconitases in plants have similar kinetic properties and are believed to be encoded by the same genes (Carrari et al. 2003; Moeder et al. 2007).

Isozymes are enzymes that differ in sequences but catalyze the same chemical reaction. Aconitase isoforms are codominant markers and have been employed in mapping studies in soybean (Griffin and Palmer 1987; Espinosa et al. 2015). Several aconitase genes have been isolated and characterized in various plant species (Peyret et al. 1995; Zhou and Ragan 1995; Sadka et al. 2000; Carrari et al. 2003; Moeder et al. 2007). The knockout mutants for the three aconitase genes, At4g35830, At4g26970, and At2g05710 in Arabidopsis thaliana (L.) Heynh. did not show a morphological phenotype, but caused several noticeable alterations, such as the darker pigmentation of leaves due to an increase in chlorophyll content, increase in total starch and starch accumulation, less root matter, higher sucrose content, and higher rate of electron transport (Carrari et al. 2003).

Five different aconitase genes have been identified in soybean and are named as Aco1, Aco2, Aco3, Aco4, and Aco5 (Doong and Kiang 1987; Griffin and Palmer 1987; Rennie et al. 1987; Kiang and Bult 1991; Espinosa et al. 2015). Inheritance studies showed that they assort independently of each other (Griffin and Palmer 1987; Rennie et al. 1987; Kiang and Bult 1991). Of these five genes, Aco3 and Aco4 have been mapped to soybean chromosomes 6 and 11, respectively (Griffin and Palmer 1987; Espinosa et al. 2015). Aco4 was flanked by two simple sequence repeat (SSR) markers, BARCSOYSSR_11_323 and BARCSOYSSR_11_336, to a ~292 kb region (Espinosa et al. 2015), which made it an excellent candidate for map-based cloning. The objectives of this research were to fine map the chromosomal region containing Aco4 and to identify and sequence the candidate gene for Aco4.

Materials and Methods

Plant material

Two diverse parents, BSR 101 (PI 548519; Aco4-aa) and Noir I (PI 290136; Aco4-bb), that differ in their aconitase pattern were crossed to generate a mapping population. The F1 plants were individually threshed and the F2 seeds were planted at the Bruner Farm near Ames, IA. The F2 mapping population consisted of 94 individuals. Samples were taken from the individual F2 plants for aconitase activity and for DNA analysis.

Aconitase isozyme analysis

Aconitase activity [aconitase hydratase, enzyme commission (EC) 4.2.1.3] was determined as previously described (Espinosa et al. 2015). Isozyme patterns at the Aco4 locus were determined for the parents and the F2 population using starch gel electrophoresis as previously described (Espinosa et al. 2015). After electrophoresis, gels were sliced horizontally into 1.5 mm thick pieces to allow analysis of several isozymes from one gel. For visualization of the aconitase activity, the gel slices were incubated at 37 °C in 100 mL 0.2 mol L\(^{-1}\) tris(hydroxymethyl)aminomethane (Tris) – HCl (pH 8.0), 200 mg cis-aconitic acid, 100 mg MgCl\(_2\), 20 mg β-nicotinamide adenine dinucleotide phosphate, 20 mg methyl thiazolyl tetrazolium bromide, 4 mg phenazine methosulfate, and 40 units isocitrate dehydrogenase.

Molecular mapping

In a previous study, the Aco4 gene was located on chromosome 11 (Espinosa et al. 2015). Twelve SSR markers from chromosome 11 were used on the parents to test for polymorphism. SSR primers were synthesized based on the sequence information from http://soybase.org/resources/ssr.php (Song et al. 2004, 2010). For the polymerase chain reactions (PCR), 50 ng DNA was used in 10 μL reaction containing 1x reaction buffer (10 mol L\(^{-1}\) Tris–HCl, 50 mol L\(^{-1}\) KCl, pH 8.3), 0.25 μM of each primer, 200 μM of each deoxynucleotide, 2.0 mol L\(^{-1}\) MgCl\(_2\), and 0.25 units of Biolase DNA polymerase enzyme (Bioline USA, Inc., Tauton, MA). Thermocycler conditions were as follows: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 46 °C for 30 s, and 72 °C for 1 min, and a final cycle of 72 °C for 10 min. The PCR products were visualized on 4% agarose gel containing ethidium bromide.
Polymorphic markers were tested on the entire F2 population and genetic distances were calculated. The mapping software Mapmaker 2.0 was used to determine genetic linkage, utilizing the Kosambi mapping function (Kosambi 1944; Lander et al. 1987). A logarithm of the odds threshold of 3.0 was used to determine order of markers. Maps were drawn using the “MapChart” software (Voorrips 2002).

Long-range PCR

A long-range PCR approach was used to capture the entire gene for sequencing. Five forward and 4 reverse primers were designed to capture the gene of interest. The gene was amplified using the long-range PCR approach. The PCR products were sub-cloned and sequenced to confirm the identity of the gene.
primers were developed (Supplementary Table S1). The forward and reverse long-range primers were designed within 1000 base pairs (bp) upstream and downstream of the gene, respectively. For long-range PCR, a LongAmp™ Taq PCR kit (New England Biolabs, Ipswich, MA) was used. The PCR program consisted of initial denaturation at 94 °C for 30 s, followed by 40 cycles of 94 °C for 30 s, 53 °C for 30 s, and 65 °C for 7 min, and elongation was set at 65 °C for 10 min. PCR products were separated on 0.8% agarose gel and desired fragment bands were

---

Fig. 2. The amino acid sequence comparison of **Glyma.11G080600** between the A-line (BSR 101) and B-line (Noir I). The differences are due to two non-synonymous substitutions between A-line and B-line. The red font represents non-synonymous substitutions and the green font represents synonymous substitutions. [Colour online.]

| A-Line | MATENPFSNILTLEKFGGAGEFKEQKYSFLPALNDRIRIDLPYRLLSRILLESAIRNCDEFQV 60 |
| B-Line | MATENPFSNILTLEKFGGAGEFKEQKYSFLPALNDRIRIDLPYRLLSRILLESAIRNCDEFQV 60 |
| A-Line | KSNDEVIKIIWTSPLKEIGFIFKQPARVLLQDFTQGPKQVAPVLDALCMRDANMNLGDSNKN 120 |
| B-Line | KSNDEVIKIIWTSPLKEIGFIFKQPARVLLQDFTQGPKQVAPVLDALCMRDANMNLGDSNKN 120 |
| A-Line | PLVFDVLIDHDSQVQDVARASENANOQANAMEMLEFPQRKNERGFQIKWGSNAFNNMVLVFPSSG 180 |
| B-Line | PLVFDVLIDHDSQVQDVARASENANOQANAMEMLEFPQRKNERGFQIKWGSNAFNNMVLVFPSSG 180 |
| A-Line | IVHQVNLGCLGRVFNTNVLIPDSVSSTHTMDLGLVAQWQVIGEEAAALMLQPMM 240 |
| B-Line | IVHQVNLGCLGRVFNTNVLIPDSVSSTHTMDLGLVAQWQVIGEEAAALMLQPMM 240 |
| A-Line | SMLFPGVQFLKLLFKLRLKDRGVTADLVLTVTQMLRKHGQVTGVQVFYEGYQGEMSELSADRT 300 |
| B-Line | SMLFPGVQFLKLLFKLRLKDRGVTADLVLTVTQMLRKHGQVTGVQVFYEGYQGEMSELSADRT 300 |
| A-Line | IANMSEPGATMGFVFDHVLQYRLRTGRSDETVMIESLYRNKMFDVDYSQFVERVY 360 |
| B-Line | IANMSEPGATMGFVFDHVLQYRLRTGRSDETVMIESLYRNKMFDVDYSQFVERVY 360 |
| A-Line | SYLEINLEDEVCVSIPKRPKFRQHVRPLREMVKDHACLVNQKVGKFGAFFVQESQNKVAEF 420 |
| B-Line | SYLEINLEDEVCVSIPKRPKFRQHVRPLREMVKDHACLVNQKVGKFGAFFVQESQNKVAEF 420 |
| A-Line | TFQGTPAHLRHGDVQVIAITCSTNTSANVMGALALAVAKGQCELQGQPKWIWLTSALPGS 480 |
| B-Line | TFQGTPAHLRHGDVQVIAITCSTNTSANVMGALALAVAKGQCELQGQPKWIWLTSALPGS 480 |
| A-Line | GVVTKYLRQSLQKYLNELFGIVGCTCTCINSQDSINEAVASITENDVAAVLSGN 540 |
| B-Line | GVVTKYLRQSLQKYLNELFGIVGCTCTCINSQDSINEAVASITENDVAAVLSGN 540 |
| A-Line | RNVEGBVHPIITRNLASPLPVAYLAALTVGQDFTEDIGIHKGTVLFRIDWPSSE 600 |
| B-Line | RNVEGBVHPIITRNLASPLPVAYLAALTVGQDFTEDIGIHKGTVLFRIDWPSSE 600 |
| A-Line | IANVQVSSTLAPMFVTDNNTIAQONGPWNNLSVPTGLYAWDPTSTYHEPFPFRDMMS 660 |
| B-Line | IANVQVSSTLAPMFVTDNNTIAQONGPWNNLSVPTGLYAWDPTSTYHEPFPFRDMMS 660 |
| A-Line | PPGSHGVDAYCCLNLNGFSITDTHISPSIHKDSPAARYLIERGVRDFDNSYGRSGGN 720 |
| B-Line | PPGSHGVDAYCCLNLNGFSITDTHISPSIHKDSPAARYLIERGVRDFDNSYGRSGGN 720 |
| A-Line | DEVMARGTTFVANIRYVKNKFLNGEVEPKTHISPQGESKLGFVDAEKEYKSEGSHMILLAGAY 780 |
| B-Line | DEVMARGTTFVANIRYVKNKFLNGEVEPKTHISPQGESKLGFVDAEKEYKSEGSHMILLAGAY 780 |
| A-Line | GSGSRRWAKGMLLNGVAKIAFEKIERHRSNLVGMIIPCLCOPGPDADSLGTLHGER 840 |
| B-Line | GSGSRRWAKGMLLNGVAKIAFEKIERHRSNLVGMIIPCLCOPGPDADSLGTLHGER 840 |
| A-Line | YITDLPSVNEIRPGQTVTVDTAGKSFVSTLRFDEVELAYNYGGGLQYVIRNVMNK 900 |
| B-Line | YITDLPSVNEIRPGQTVTVDTAGKSFVSTLRFDEVELAYNYGGGLQYVIRNVMNK 900 |

excised. DNA was purified from the gel fragments using an extraction kit (Clontech Laboratories, Inc., Mountain View, CA).

**Sequencing**

Sequencing was performed using 15 forward and 15 reverse primers along the length of the candidate Aco4 gene using the long-range product as a template (Supplementary Table S2). PCR conditions used for amplification were as follows: 94 °C for 3 min, 40 cycles of 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 4 min, with final extension at 72 °C for 10 min. Sequencing was performed at Functional Biosciences, Inc. (Madison, WI) using Sanger sequencing. Vector NTI Express (Life Technologies Corporation, Carlsbad, CA) was used to align the DNA fragments.

**Expression analysis of Glyma.11G080600**

The soybean RNA sequencing dataset was used to compare the expression of Glyma.11G080600 during different developmental stages (Severin et al. 2010). Fragments per kilobase of transcript per million mapped reads (FPKM) for different development stages of soybean were compared (https://phytozome.jgi.doe.gov/pz/portal.html#!gene?search=1&detail=1&method=4433&searchText=transcriptid:30528935). FPKM, fragments per kilobase of transcript per million mapped reads.

**Phylogenetic analysis**

The protein sequence for Glyma.11G080600 was used in a protein Basic Local Alignment Search Tool (BLAST®) search using the National Center for Biotechnology Information BLAST web service (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to find homologous sequences from several plant species. The protein sequences were aligned in SeaView 4.5.4, a phylogenetic alignment software program (Gouy et al. 2010). The phylogenetic relationship between Aco4 (Glyma.11G080600) and 24 homologous proteins was then constructed using the neighbor-joining method. The evolutionary distances were computed using the Poisson correction method and are in units of the number of amino acid substitutions per site.

**Results and Discussion**

**Fine mapping**

In a previous study, the Aco4 gene was mapped on chromosome 11 between SSR markers BARCSOYSSR_11_332 and BARCSOYSSR_11_336 that flanked a 292 kb region (Espinosa et al. 2015). In this investigation, we wanted to narrow down the region, so the putative candidate gene for Aco4 can be identified. We selected eight SSR markers between BARCSOYSSR_11_323 and BARCSOYSSR_11_336 and tested those for polymorphism between the parents. Only BARCSOYSSR_11_332 showed polymorphism; we then generated a new map including BARCSOYSSR_11_332. The Aco4 gene mapped between BARCSOYSSR_11_332 and BARCSOYSSR_11_336 (Fig. 1). This region is physically ~148 kb and there are only 19 predicted genes in this region (Table 1). Of these 19 genes, Glyma.11G080600 is annotated to code for acionitate hydratase (Table 1; www.soybase.org), which makes Glyma.11G080600 a perfect candidate for Aco4.

**Isolation of the Aco4 gene**

To confirm if Glyma.11G080600 is a viable candidate for Aco4, we sequenced the Glyma.11G080600 locus in the parents BSR 101 (Aco4-aa; A-line) and Noir I (Aco4-bb; B-line) and homozygous F2 plants Aco4-aa (pooled sample from 10 plants) and Aco4-bb (pooled sample from 10 plants). The genomic sequence of Glyma.11G080600 is 8366 bp long with 21 exons and codes for a 901 amino acid-long protein (Supplementary Fig. S1). Sequence comparison between the A-line and B-line showed 12 base substitutions and 7 indel variations in the introns (Supplementary Fig. S2). In the exons, there were only five differences. Of these, three were synonymous substitutions of G with A, C with T, and G with C at the positions 766 (G766A), 1718 (C1718T), and 2750 (G2750C), respectively (Fig. 2; Supplementary Fig. S2). Positions of the substitutions are described based on the sequence of the A-line. The substitutions G766A, C1817T, and G2750C did not affect the amino acids glutamic acid, histidine, and serine, respectively, in the protein sequence. Synonymous substitutions most likely were not the cause of variation in the isozyme pattern in the A-line and B-line. The other two differences between the A-line and B-line were non-synonymous substitutions. The substitution of T with A at position 1197 (T1197A) resulted in a change from valine to glycine and substitution G5424T caused a change from alanine to serine in the protein sequence (Supplementary Fig. S2; Fig. 2). The new amino acids create slightly
different proteins in the A- and B-lines, which possibly will have a different structure and lead to a different isozyme pattern.

Glyma.11G080600 contains two major domains including aconitase A catalytic domain (amino acids 87–571) and aconitase A swivel domain (amino acids 675–845) (Supplementary Fig. S1). Aconitase catalytic domain catalyzes isomerization of citrate to isocitrate and vice versa during the Krebs cycle (Marchler-Bauer et al. 2015). The swivel domain is known to play an important role in swiveling conformational change in the enzyme mechanism (Marchler-Bauer et al. 2015). Conservation of both domains suggests that the Glyma.11G080600 protein is able to catalyze the interconversion of isocitrate and citrate.
Expression of Glyma.11G080600 in different soybean tissues

Publicly available RNA sequencing data were used to compare the expression of Glyma.11G080600 in different soybean tissues (Severin et al. 2010). Glyma.11G080600 is expressed in most plant tissues with maximum expression in leaves and minimum expression in the stem (Fig. 3). Expression in leaves is more than 4-fold higher than stem, which may be due to active involvement of leaves in metabolism.

Phylogenetic analysis of aconitate hydratase in different plant species

To analyze the evolutionary history of the aconitate hydratase family members in different plant species, the Glyma.11G080600 protein sequence was used in BLAST analysis to identify homologous proteins in different plant species. Twenty-five different proteins were compared in a multiple alignment and phylogenetic analysis was performed. The phylogenetic tree displayed three major clusters (Fig. 4). The Aco family in soybean displays multiple sequences associated with a specific aconitate gene, perhaps due to the paleopolyploid nature of the soybean genome. Four soybean genes, Glyma.12G193800, Glyma.13G308700, Glyma.06G305700, and Glyma.12G089800, cluster together with the Arabidopsis Aco3 gene, At2g05710 (Fig. 4). In a previous study, the soybean Aco3 gene was mapped on chromosome 6 in the vicinity of Glyma.06G305700 (Griffin and Palmer 1987; Kiang and Bult 1991). It is likely that Glyma.06G305700 is an ortholog of Arabidopsis Aco3. The other three soybean genes may be paralogs of Glyma.06G305700, which may have developed due to duplication events. Two soybean genes, Glyma.11G080600 and Glyma.01G162800, fall in the same cluster with Arabidopsis Aco1. In this investigation, we have shown that Glyma.11G080600 is the most suitable candidate for Aco4. It seems that the Aco1 gene has evolved into two closely associated genes, Glyma.11G080600 (Aco4) and Glyma.01G162800 (Aco1), in soybean. The short evolutionary distance between the genes suggests an occurrence of a recent duplication event. Each of these two Glycine max (L.) Merr. genes has a closely associated orthologous gene in G. soja Siebold. & Zucc., which indicates that the duplication event happened prior to the G. max and G. soja split. Further research on the characterization of the Aco genes in soybean may shed light on the need of several variants in plant development, survival, and reproductive success.

Conclusion

We have mapped the Aco4 gene to about a 148 kb region on chromosome 11 in soybean. Of the 19 predicted genes present in this region, Glyma.11G080600 codes for aconitate hydratase. Sequencing of Glyma.11G080600 in the lines that vary for their aconitate pattern displayed five differences in their exon sequences, of which, only two were non-synonymous substitutions, which most likely are the cause of different isozyme patterns in the two variants. These observations indicate that Glyma.11G080600 is the most suitable candidate for Aco4. Phylogenetic analysis suggested that soybean Aco4 and Aco1 are very closely associated and may have emerged from a single progenitor gene in the recent past.

Acknowledgements

We gratefully acknowledge funding from the University of Wisconsin – Stevens Point Student Research Fund. Dr. K. Espinosa’s graduate research was supported by the American Seed Trade Association’s Seed Research Fund and Iowa Crop Improvement Association.

References


