Identification and Evolutionary Analysis of the Cotton DUR3 Gene (Postprint)

Authors: Gong Yuanyong, Zhao Lihua, Yan Fei, Sun Yichen, Wang Hui, Liu Laihua

Date: 2021-04-29T00:00:00+00:00

Abstract

Plant DUR3 homologs are high-affinity urea transporters belonging to the sodium ion/solute symporter family, which play important roles in the active uptake of exogenous urea and the redistribution of endogenous urea in plants. To elucidate the structure and evolutionary status of cotton DUR3 genes, bioinformatics approaches were employed to identify DUR3 genes from Gossypium hirsutum and Gossypium raimondii at the whole-genome level, and to analyze their gene structure, transmembrane domains, motif distribution, and evolutionary relationships. The results indicated that one DUR3 gene was identified from each of the A and D subgenome chromosomes of Gossypium hirsutum, and one DUR3 gene was identified from the genome of Gossypium raimondii. These three cotton DUR3 homologous proteins, like other plant DUR3 homologs, possess 15 transmembrane domains and three highly conserved motifs at consistent positions. Gene structure analysis revealed that the number of exons in DUR3 genes of dicotyledonous plants is significantly greater than that in monocotyledonous plants, and these three cotton DUR3 genes exhibit the same pattern. The phylogenetic tree constructed from DUR3 amino acid sequences of various species demonstrated that these genes clustered according to phylogenetic relationships, with cotton genes grouping with those of other dicotyledonous plants. The Ka/Ks ratios of both orthologous and paralogous DUR3 genes were generally greater than 1, suggesting that these genes have been primarily subjected to positive selection during evolution. The findings of this study will provide a theoretical foundation for further research on cotton DUR3 homologous proteins.

Full Text

Preamble

Identification and Evolutionary Analysis of Cotton DUR3 Genes

Yuanyong Gong[1], Lihua Zhao[1], Fei Yan[1], Yichen Sun[2], Hui Wang[2], Laihua Liu[2]*

[1]Biological and Chemical Engineering College, Panzhihua University, Panzhihua, Sichuan 617000, China
[2]Key Laboratory of Plant-Soil Interactions, Ministry of Education; Center for Resources, Environment and Food Security, China Agricultural University, Beijing 100193, China

Abstract

Plant DUR3 homologous proteins are high-affinity urea transporters belonging to the sodium/solute symporter family that play crucial roles in the active absorption of exogenous urea and redistribution of endogenous urea in plants. To elucidate the structure and evolutionary characteristics of cotton DUR3 genes, we employed bioinformatics approaches to identify DUR3 genes from the whole genomes of Gossypium hirsutum and Gossypium raimondii, and systematically analyzed their gene structure, transmembrane domains, motif distribution, and evolutionary relationships. The results identified one DUR3 gene from each of the A and D subgenome chromosomes of upland cotton, and one DUR3 gene from the G. raimondii genome. These three cotton DUR3 homologous proteins, like other plant DUR3 homologs, possess 15 transmembrane domains and three highly conserved motifs at consistent positions. Gene structure analysis revealed that the number of exons in dicotyledonous plant DUR3 genes is significantly higher than that in monocotyledonous plants, a pattern also observed for the three cotton DUR3 genes. Phylogenetic analysis of DUR3 amino acid sequences from different species showed clustering according to phylogenetic relationships, with cotton DUR3 genes grouping with other dicotyledonous plants. The Ka/Ks ratios for both orthologous and paralogous DUR3 genes were generally greater than 1, indicating that these genes have predominantly experienced positive selection during evolution. These findings provide a theoretical foundation for further investigation of cotton DUR3 homologous proteins.

Keywords: Gossypium hirsutum, Gossypium raimondii, DUR3 gene, bioinformatics, evolution

Introduction

Nitrogen is an essential nutrient for plant growth, serving as a fundamental component for synthesizing proteins, amino acids, chlorophyll, and other nitrogen-containing compounds (Marschner, 1995). Urea is one of the most widely used fertilizers worldwide; however, urease—an enzyme secreted by ubiquitous soil bacteria—rapidly degrades most applied urea into ammonia and carbon dioxide, resulting in typically low soil urea concentrations insufficient to meet plant growth requirements as a sole nitrogen source (Kojima et al., 2006). Nevertheless, research has demonstrated that urea applied to soil exhibits a half-life of

1–8 days (Liu et al., 2003a), during which relatively high urea concentrations persist, providing a valuable window for root absorption.

To date, only two classes of urea transporters have been identified in plants. The first class comprises Major Intrinsic Proteins (MIPs), which facilitate low-affinity, passive urea transport and localize to both plasma and vacuolar membranes (Liu et al., 2003b). The second class consists of DUR3 homologous proteins—high-affinity urea transporters belonging to the sodium/solute symporter (SSS) family that mediate active urea transport across membranes. DUR3 genes are induced under nitrogen-deficient conditions and encode transmembrane proteins localized to the plasma membrane (Liu et al., 2003a; Liu et al., 2003b; Kojima et al., 2007). The physiological role of DUR3 extends beyond nitrogen-deficient roots; it also functions in senescing leaves (Bohner et al., 2015). Plants can obtain urea not only from external soil sources but also through endogenous nitrogen metabolism, as evidenced by increased urea content in aging leaves (Bohner et al., 2015). Studies in Arabidopsis revealed that AtDUR3 is highly expressed in vascular tissues of senescing leaves, where it transports urea generated from nitrogen compound degradation out of mesophyll cells into the apoplast (Bohner et al., 2015), suggesting that DUR3 serves as a critical channel for nitrogen redistribution within plants. Reverse genetics studies in rice demonstrated that OsDUR3 insertion lines exhibited 26.2% reduced grain yield under nitrogen-deficient field conditions due to poor grain filling. During seed development, nitrogen accumulated primarily in leaves rather than panicles, and urea content in old leaves of insertion lines was lower than in wild-type plants, confirming DUR3' s involvement in nitrogen transport and rice yield under nitrogen-limiting conditions (Beier et al., 2015).

Currently, only three DUR3 genes from higher plants have been cloned and functionally characterized: Arabidopsis thaliana (AtDUR3) (Liu et al., 2003a), Oryza sativa (OsDUR3) (Wang et al., 2003a), and Zea mays (ZmDUR3) (Zamin et al., 2014; Liu et al., 2015). These studies have focused on DUR3-mediated active urea transport. Yeast studies have shown that Saccharomyces cerevisiae ScDUR3 and Candida albicans CaDUR3 can transport not only urea but also polyamines (Uemura et al., 2006; Kumar et al., 2011). Polyamines are aliphatic small molecules that participate extensively in plant growth and development and play crucial roles in plant responses to abiotic stress (Groppa and Benavides, 2008). However, no polyamine transporters have been identified in plants, making it worthwhile to investigate whether plant DUR3 homologs possess polyamine transport functions.

Cotton is a major crop for plant fiber production, with upland cotton (Gossypium hirsutum) being the most important cultivated species worldwide, accounting for over 90% of global cotton acreage. The upland cotton genome is an allotetraploid (AADD), making it large and complex, which has limited progress in cotton molecular biology research. However, the completion of genome sequencing for diploid D-subgenome G. raimondii (Wang et al., 2012) and A-subgenome G. arboreum (Li et al., 2014), along with the publication of the whole genome

sequence of the upland cotton genetic standard line TM-1 (Li et al., 2015; Zhang et al., 2015), has made both G. hirsutum and G. raimondii genomes available through JGI, accelerating functional gene research in cotton. This study identifies cotton DUR3 genes from the G. hirsutum and G. raimondii genomes available at JGI and conducts comprehensive bioinformatics analyses to provide a theoretical basis for further cloning and functional validation of these genes.

Materials and Methods

1.1 Materials

DUR3 gene sequences from the following species were obtained from NCBI (https://www.ncbi.nlm.nih.gov/): Arabidopsis thaliana (AtDUR3, NP_{199351}), Oryza sativa (OsDUR3, NP_{001065513}), Zea mays (ZmDUR3, KJ652242), Sorghum bicolor (SbDUR3, XP_{002438118}.1), Setaria italica (SiDUR3, XP_{004965066}.1), Solanum lycopersicum (SlDUR3, XP_{004245999}.1), Vitis vinifera (VvDUR3, XP_{002263043}.1), Glycine max (GmDUR3, XP_{003523904}.1), Hordeum vulgare (HvDUR3, BAJ94433.1), Brachypodium distachyon (BdDUR3, XP_{003571687}.1), Medicago truncatula (MtDUR3, XP_{003612583}.1), Populus trichocarpa (PtDUR3, XP_{002303472}.2), Saccharomyces cerevisiae (ScDUR3, L19875.1), Aspergillus nidulans (AnDUR3, ACZ62639.1), Pyropia yezoensis (PyDUR3, BAU04114.1), Crassostrea gigas (CgDUR3, XP_{019929725}.1), and Tridacna squamosa (TsDUR3, MF073181.1). Coding region and CDS sequences for the aforementioned plant DUR3 genes were retrieved from Phytozome at JGI (https://phytozome.jgi.doe.gov/pz/portal.html).

1.2.1 Identification of Cotton DUR3 Genes

Using the Arabidopsis AtDUR3 protein sequence (NCBI accession: NP_{199351}) as a query, BLAST searches were performed against the Gossypium hirsutum and Gossypium raimondii genomes in Phytozome to retrieve homologous sequences with E-values less than e^{-10}. The obtained sequences (cotton and Arabidopsis) were analyzed for functional domain presence using InterProscan5 (http://www.ebi.ac.uk/Tools/pfa/iprscan/) to finalize target sequence identification.

1.2.2 Acquisition of Basic Information for Plant DUR3 Genes

Information on coding region length, CDS sequence length, and exon number for cotton and other plant DUR3 genes required for this study was obtained from Phytozome. Basic information including molecular weight and isoelectric point of plant DUR3 amino acid sequences was analyzed using the ProtParam tool (http://web.expasy.org/protparam/).

1.2.3 Bioinformatics Analysis of DUR3 Genes

Multiple sequence alignments of DUR3 amino acid sequences from different plants were performed using DNAMAN software. Phylogenetic trees were constructed from DUR3 protein sequences of various species using MEGA6 software with the Neighbor-Joining method and bootstrap validation parameter set to 1000. Amino acid sequence similarity among different plant DUR3 genes was analyzed using MegAlign in the Lasergene software suite. Conserved motifs in plant DUR3 protein sequences were identified using MEME (http://meme.sdsc.edu/meme/meme.html). Transmembrane domains of plant DUR3 protein sequences were predicted online using TMHMM (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0). Gene structure diagrams for different plant DUR3 genes were generated online using GSDS9 (Gene Structure Display Server, http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). Non-synonymous (Ka) and synonymous (Ks) substitution rates for cotton DUR3 paralogous and orthologous genes were calculated using DnaSPv5 to determine Ka/Ks ratios.

Results and Analysis

2.1 Identification and Basic Sequence Information of Cotton DUR3 Genes

Two DUR3 homologous genes were identified from the allotetraploid standard line TM-1 genome of Gossypium hirsutum: Gohir.A03g179800 located on the A03 subgenome and Gohir.D08g001500 located on the D08 subgenome. For convenience, these were designated as GhDUR3.1 and GhDUR3.2, respectively. One DUR3 homologous gene, Gorai.005G228700, was identified from the Gossypium raimondii genome and named GrDUR3. InterProscan5 analysis of the amino acid sequences of these three DUR3 genes revealed that they all belong to the Urea active transporter (IPR031155) family under the Sodium/solute symporter (IPR001734) protein family, and possess the cd11476 functional domain (Na+/urea-polyamine cotransporter DUR3 and related proteins; solute-binding domain).

As shown in Table 1 , the coding region lengths of plant DUR3 genes vary considerably, ranging from 2,358 bp for OsDUR3 to 5,567 bp for ZmDUR3—a difference of 3,209 bp. However, the CDS sequence lengths show minimal variation, with the shortest being 2,034 bp for GhDUR3.1 and the longest being 2,214 bp for BdDUR3, differing by only 180 bp. Plant DUR3 genes exhibit a clear dichotomy in exon number: dicotyledonous plants generally have 9–10 exons, while monocotyledonous plants have only 3–4 exons. The deduced polypeptide lengths (677–737 amino acids) and molecular weights (72.905–77.608 kDa) are relatively consistent across plant DUR3 proteins, with isoelectric points mostly above 7 (alkaline range), except for SiDUR3 which is below 7.

2.2 Amino Acid Sequence Alignment of Plant DUR3 Genes

Multiple sequence alignment was performed on 15 DUR3 protein amino acid sequences from various model plants including Arabidopsis thaliana, Oryza sativa, Zea mays, Sorghum bicolor, Setaria italica, Solanum lycopersicum, Vitis vinifera, Glycine max, Hordeum vulgare, Brachypodium distachyon, Medicago truncatula, Populus trichocarpa, and the three cotton DUR3 proteins identified in this study. As shown in Figure 1 [Figure 1: see original paper], these 15 sequences share 79.81% identity, with divergent regions primarily at the N- and C-termini while the central functional domains show high conservation. The lowest sequence identity (71.8%) was observed between BdDUR3 and MtDUR3, while the highest (99.3%) was between GhDUR3.2 and GrDUR3.

2.3 Analysis of Conserved Motifs in Plant DUR3 Proteins

MEME online analysis identified motif information for 15 DUR3 proteins from 14 different plant species. All 15 protein sequences contain three distinct motifs. As shown in Figure 2 [Figure 2: see original paper], these three motifs exhibit consistent distribution patterns across the amino acid sequences, appearing in the order: motif 3 at the N-terminus, followed by motif 1, and motif 2 near the middle of the sequence. Figure 3 [Figure 3: see original paper] reveals that motifs 1 and 2 each consist of 50 amino acids, while motif 3 comprises 33 amino acids, with all three motifs showing high conservation.

2.4 Prediction of Transmembrane Domains in Plant DUR3 Proteins

Since DUR3 proteins are known to be transmembrane proteins, analysis of their transmembrane domains is essential. The TMHMM Server v. 2.0 was used to predict transmembrane structures for all 15 DUR3 proteins. As shown in Table 2 , all 15 DUR3 proteins contain 15 transmembrane domains (illustrated in Figure 4 [Figure 4: see original paper]), with these domains occupying similar positions in the amino acid sequences across different species.

2.5 Gene Structure Analysis of Plant DUR3 Genes

Based on coding region and CDS sequences obtained from JGI for 15 DUR3 genes from different plants, gene structure diagrams were generated using GSDS2.0 software to analyze exon-intron organization. Table 1 shows that dicotyledonous plant DUR3 genes average 9.2 exons, while monocotyledonous plants average 3.5 exons, with this difference visualized more intuitively in Figure 5 [Figure 5: see original paper]. Notably, genes with identical exon numbers also show similar exon lengths, as observed for the nine-exon genes (GhDUR3.2, GmDUR3, GrDUR3, MtDUR3, PtDUR3, SlDUR3, VvDUR3), three-exon genes (BdDUR3, HvDUR3, OsDUR3), and four-exon genes (SbDUR3, SiDUR3, ZmDUR3).

2.6 Phylogenetic Analysis of DUR3 Genes

To investigate the evolutionary relationships of DUR3 genes across species, a phylogenetic tree was constructed from 20 DUR3 protein sequences representing plants, fungi, algae, and mollusks using MEGA6 software. The results (Figure 6 [Figure 6: see original paper]) show that all plant DUR3 proteins cluster in one branch, two fungal proteins (Saccharomyces cerevisiae ScDUR3 and Aspergillus nidulans AnDUR3) group together, two mollusk proteins (Crassostrea gigas Cg-DUR3 and Tridacna squamosa TsDUR3) form another cluster, the algal protein (Pyropia yezoensis PyDUR3) occupies a separate branch, and monocotyledonous and dicotyledonous plants further segregate into distinct sub-branches.

2.7 Analysis of Ka/Ks Ratios for Plant DUR3 Homologous Genes

CDS sequences of all plant DUR3 homologous genes were aligned using ClustalW in MEGA6, followed by calculation of non-synonymous (Ka) and synonymous (Ks) substitution rates using DnaSPv5 to determine Ka/Ks ratios. Table 3 shows that Ka/Ks ratios for DUR3 orthologous gene pairs are generally greater than 1 (except for GhDUR3.2 and GrDUR3), indicating that DUR3 genes have undergone positive selection during evolution across different plant species. The Ka/Ks ratio for the two paralogous genes in upland cotton is 4.12, also greater than 1, suggesting that upland cotton DUR3 genes have similarly experienced positive selection during intragenomic evolution.

Discussion and Conclusion

In diploid plant genomes, DUR3 homologous genes exist as singletons (Cao et al., 2009). Our results confirm that, except for the allotetraploid upland cotton (AADD), all other plants examined are diploid and harbor single DUR3 genes in their genomes. As an allotetraploid (AADD), upland cotton contains A and D subgenomes, and the two identified genes, GhDUR3.1 and GhDUR3.2, belong to the A and D subgenomes, respectively. Gossypium raimondii is a D-subgenome diploid cotton, and its GrDUR3 gene shows greater than 99% similarity to GhDUR3.2 (also located in the D subgenome) at both the CDS and polypeptide levels. Except for minor differences in coding region length and polypeptide molecular weight, these two genes share identical basic information, including identical predicted transmembrane domain positions, indicating highly similar molecular and physiological functions.

Overall, the plant DUR3 amino acid sequences analyzed in this study share nearly 80% identity. Analyses of isoelectric points, transmembrane domains, and motifs all demonstrate high conservation among plant DUR3 proteins. Only SiDUR3 has an isoelectric point slightly below 7 (6.89), while all others are above 7, with SlDUR3 reaching 9.0, indicating that plant DUR3 proteins are generally basic (Table 1). All plant DUR3 proteins contain three highly conserved motifs with consistent arrangement patterns across amino acid sequences (Figures 2 and 3). They also share the same number of transmembrane domains (15) with

similar positions in the amino acid sequences (Table 2, Figure 4), a feature consistent in both upland cotton GhDUR3.1 and GhDUR3.2 and G. raimondii GrDUR3.

Phylogenetically, DUR3 amino acid sequences cluster according to species relationships, with plant, fungal, and mollusk DUR3 genes forming separate clades, and algae occupying a distinct branch. Within plants, monocotyledonous and dicotyledonous species further segregate into separate sub-branches (Figure 6). As a dicotyledonous plant, cotton groups within the dicot clade, with its three DUR3 genes forming a closely related subcluster.

The Ka/Ks ratios for both orthologous and paralogous DUR3 genes are generally greater than 1 (Table 3), indicating that these genes have primarily experienced positive selection during evolution. Additionally, plant DUR3 gene structure follows a distinct pattern, with dicotyledonous plants possessing significantly more exons than monocotyledonous plants (Table 1, Figure 5), providing valuable insights for studying the structural evolution of plant DUR3 genes.

References

Beier MP, Fujita T, Sasaki K, et al., 2019. The urea transporter DUR3 contributes to rice production under nitrogen-deficient and field conditions. *Physiol Plantarum*, 167(1): 75-89.

Bohner A, Kojima S, Hajirezaei M, et al., 2015. Urea retranslocation from senescing Arabidopsis leaves is promoted by DUR3-mediated urea retrieval from leaf apoplast. *Plant J*, 81(3): 377-387.

Cao FQ, Liu GW, Wang WH, et al., 2009. Molecular Processes of Urea Metabolism and Transport in Higher Plants. *Chin Bull Bot*, 44(3): 273-282.

Groppa MD, Benavides MP, 2008. Polyamines and abiotic stress: recent advances. *Amino Acids*, 34(1): 35-45.

Hu B, Jin J, Guo AY, et al., 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31(8): 1296-1297.

Kojima S, Bohner A, Gassert B, et al., 2007. AtDUR3 represents the major transporter for high-affinity urea transport across the plasma membrane of nitrogen-deficient Arabidopsis roots. *Plant J*, 52(1): 30-40.

Kojima S, Bohner A, Nicolaus von Wirén. 2006. Molecular Mechanisms of Urea Transport in Plants. *J Membrane Biol*, 212(2): 83-91.

Kumar R, Chadha S, Saraswat D, et al., 2011. Histatin 5 Uptake by Candida albicans Utilizes Polyamine Transporters Dur3 and Dur31 Proteins. *J Biol Chem*, 286(51): 43748-43758.

Li F, Fan G, Wang K, et al., 2014. Genome sequence of the cultivated cotton Gossypium arboreum. *Nat Genet*, 46(6): 562-574.

Li F, Fan Z, Lu C, et al., 2015. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nat Biotechnol*, 33(5): 524-530.

Liu GW, Sun AL, Li DQ, et al., 2015. Molecular identification and functional analysis of a maize (Zea mays) DUR3 homolog that transports urea with high affinity. *Planta*, 241(4): 861-874.

Liu LH, Ludewig U, Wolf B Frommer, et al., 2003a. AtDUR3 Encodes a New Type of High-Affinity Urea/H^+ Symporter in Arabidopsis. *Plant Cell*, 15(3): 790-800.

Liu LH, Ludewig U, Gassert B, et al., 2003b. Urea Transport by Nitrogen-Regulated Tonoplast Intrinsic Proteins in Arabidopsis. *Plant Physiol*, 133(3): 1220-1228.

Marschner H, 1995. *Mineral Nutrition of Higher Plants*. London: Academic Press, pp. 889.

Uemura T, Kashiwagi K, Igarashi K, 2006. Polyamine Uptake by DUR3 and SAM3 in Saccharomyces cerevisiae. *J Biol Chem*, 282(10): 7733-7741.

Wang K, Wang Z, Li F, et al., 2012. The draft genome of a diploid cotton Gossypium raimondii. *Nat Genet*, 44(10): 1098-1103.

Wang WH, Köhler B, Cao FQ, et al., 2012. Rice DUR3 mediates high-affinity urea transport and plays an effective role in improvement of urea acquisition and utilization when expressed in Arabidopsis. *New Phytol*, 193(2): 432-444.

Zamin L, Tomasi N, Wirdnam C, et al., 2014. Isolation and functional characterization of a high affinity urea transporter from roots of Zea mays. *BMC Plant Biol*, 14(1): 222.

Zhang T, Hu Y, Jiang W, et al., 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*, 33(5): 531-537.

Note: Figure translations are in progress. See original paper for figures.