

Case study of the Cellulose synthase (CesA) and cellulose synthase-like (Csl) superfamily

In this document, we provide an overview of the characterization of the cellulose synthase family using the Banana Genome Hub.

To identify Musa CESA and CSL sequences, the Musa proteome was searched using the protein combination tool of GreenPhylDB (<http://www.greenphyd.org/cgi-bin/ipr2genomes.cgi>) with InterPro domain IPR005150 corresponding to the cellulose synthase family (CESA, CSLB, CSLD-J) and IPR001173 corresponding to the Glycosyl transferase family 2 (CSLA, CSLC).

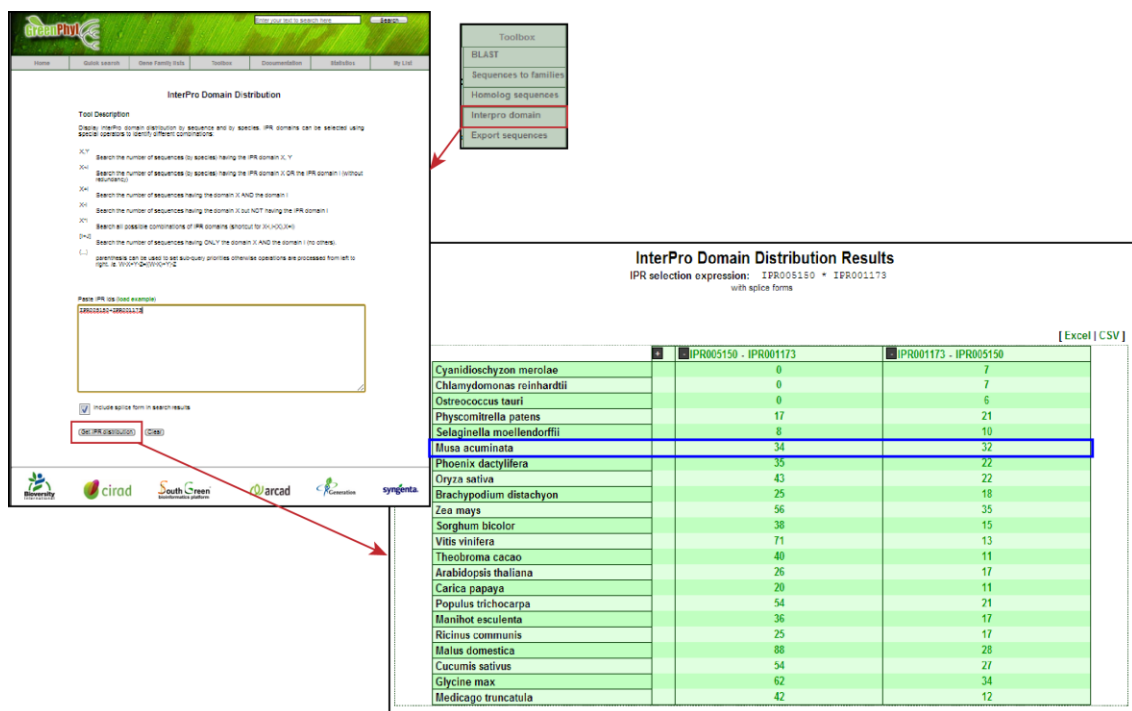


Figure 1 - Interpro domain research

The search using IPR005150 identified 34 sequences that clustered within the “Cellulose synthase” family in GreenPhylDB. The search using IPR001173 in GreenPhylDB identified 32 sequences of which 26 clustered within the “Putative glycosyl transferase” family that comprised CSLA and CSLC homologs from other species.

The sequences have also been identified through their function with pathway tools. (<http://banana-genome.cirad.fr/musacyc>)

Search Database *Musa acuminata* [change](#)

Musa acuminata Query Results

The query 2.4.1.12 matched the following objects:

[Proteins \(34\)](#) | [Gene Ontology Terms \(1\)](#) | [Reactions \(1\)](#)

Proteins Gene/Gene Product pages contain: chromosomal location of gene; depiction of its operon; link to genome browser; detailed summaries and citations; subunit structure (for protein complexes); cofactors, activators, and inhibitors (for enzymes); depiction of regulon (for transcriptional regulators); protein features.

- Cellulose synthase A catalytic subunit 4 [UDP-forming] (polypeptide) - GSMUA_Achr6P31810_001
- Cellulose synthase A catalytic subunit 4 [UDP-forming] (polypeptide) - GSMUA_Achr9P05580_001
- Cellulose synthase A catalytic subunit 6 [UDP-forming]
- Cellulose synthase A catalytic subunit 7 [UDP-forming]
- Cellulose synthase A catalytic subunit 9 [UDP-forming] (polypeptide) - GSMUA_Achr4P26500_001
- Cellulose synthase A catalytic subunit 9 [UDP-forming] (polypeptide) - GSMUA_Achr7P19410_001
- Cellulose synthase-like protein D1
- Cellulose synthase-like protein D2 (polypeptide) - GSMUA_Achr1P23960_001
- Cellulose synthase-like protein D2 (polypeptide) - GSMUA_Achr2P15300_001
- Cellulose synthase-like protein D2 (polypeptide) - GSMUA_Achr10P26260_001
- Cellulose synthase-like protein D2 (polypeptide) - GSMUA_Achr11P23440_001
- Cellulose synthase-like protein D2 (polypeptide) - GSMUA_AchrUn_randomP07850_001
- Cellulose synthase-like protein D3
- Cellulose synthase-like protein D4
- Cellulose synthase-like protein D5
- Cellulose synthase-like protein E6 (polypeptide) - GSMUA_Achr4P13080_001
- Cellulose synthase-like protein E6 (polypeptide) - GSMUA_Achr9P18830_001
- Cellulose synthase-like protein E6 (polypeptide) - GSMUA_Achr9P28000_001
- Cellulose synthase-like protein G3

Figure 2 - Search by function on MusaCyc

Predicted enzymes with the associated E.C. number E.C. 2.4.1.12 for cellulose synthase function that were retrieved from MusaCyc(Fig.2), corresponded to the same set of 34 sequences.

Homologous sequences from the rice, sorghum, Arabidopsis and grapevine genomes were retrieved from GreenPhylDB (http://www.greenphyl.org/cgi-bin/family.cgi?p=id&family_id=73#tab-famstruct) and were assigned to the different subfamilies following published data (Fig3.)

The image shows the GreenPhylDB interface for family GP000073. On the left, there is a sidebar with family details and a 'Build your own display' section. The main area on the right shows a table of homologous sequences.

Family GP000073 Details:

- Family ID: GP000073
- Family name: Cellulose synthase family
- Family Ontology:
 - molecular_function: GO:000723 (11%) - RNA binding; GO:000748 (11%) - translation initiation factor activity; GO:000818 (20%) - protein binding; GO:0008270 (17%) - protein binding; GO:0018780 (79%) - cellulose synthase (UDP-forming) activity
 - cellular_component: GO:0006737 (11%) - cytoplasm; GO:0016020 (79%) - membrane
 - biological_process: GO:0006415 (11%) - translational initiation; GO:0005244 (79%) - cellulose biosynthetic process

Table of Homologous Sequences:

Sequence	Species	Gene name	UniProt	InterPro	Annotation	Gene Ontology
At1g02730.1	Arabidopsis thaliana		Q9KXW9	IPR005150	Symbol: ATCLSD5, CSLD5 ATCLSD5; L4-beta-D-xylofan synthase/ cellulose synthase chr1:594697-596473 REVERSE	GO:0016020 GO:0008044 GO:0016760
At1g29550.1	Arabidopsis thaliana		Q9KCP6	IPR001340 IPR019770	Symbol: eucaryotic translation initiation factor 4E, putative / eIF4E, putative / eIF4E, putative / mRNA cap-binding protein, putative chr1:10330514-10552094 FORWARD	GO:0007223 GO:0008737 GO:0003740 GO:0006413
At1G29590.2	Arabidopsis thaliana		Q9KCP2	IPR001340 IPR019770	Symbol: eIF4E3 Eukaryotic initiation factor 4E protein chr1:10340072-10342750 FORWARD	GO:0007223 GO:0008737 GO:0003740 GO:0006413
At1g22180.1	Arabidopsis thaliana		Q9KFR3	IPR005150	Symbol: ATCLSD6, CSLD6 ATCLSD6; cellulose synthase/ transferase, transferring glycosyl group chr1:11505516-11509651 REVERSE	GO:0016020 GO:0008044 GO:0016760

Figure 3 - Homologous sequences from rice, sorghum, Arabidopsis and grapevine genomes on GreenPhylDB

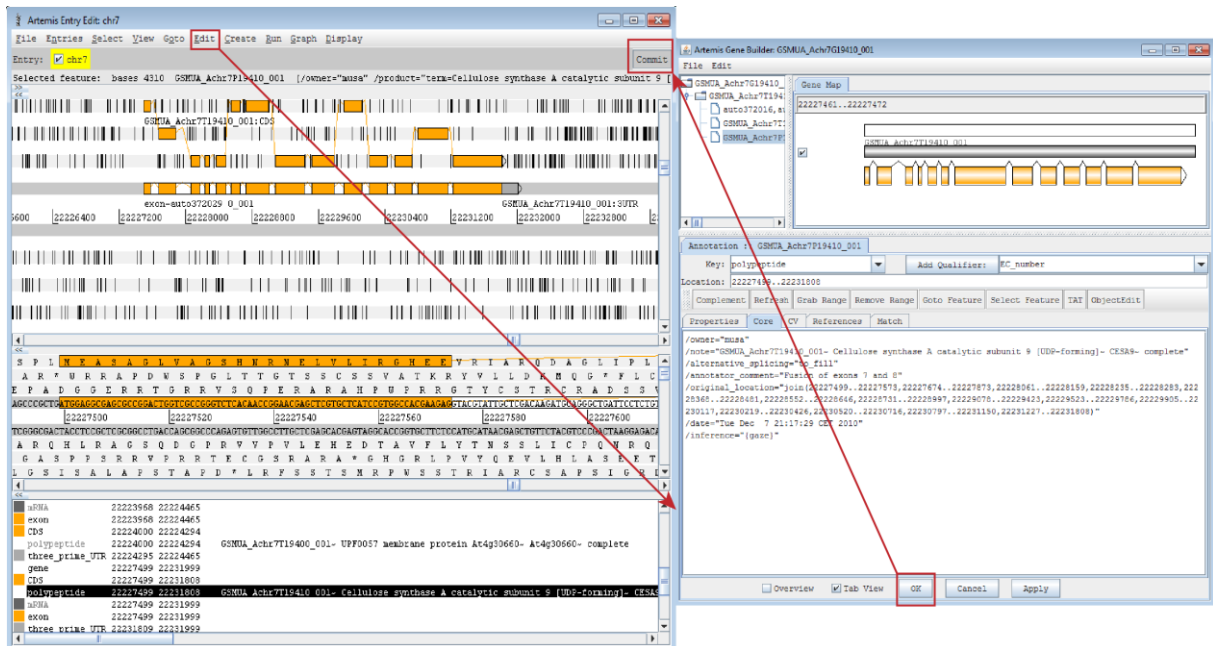


Figure 4 - Manual Curation with Artemis

Manual curation of the compiled 60 *Musa* genes (Fig. 4) was performed using information from different *de novo* prediction programs, expression data and BLASTP data available on the banana GBrowse. This resulted in a final set of 59 genes.