

Case Study of the Banana SPS Gene Family

Table of contents

- 1. Search for the *Musa* SPS Gene Family2
- 2. Structural Annotation4
 - 2.1. Genes.....4
 - 2.2. Molecular Markers.....8
- 3. Functional Annotation10
 - 3.1. Gene Report 10
 - 3.2. Gene Families 13
 - 3.3. Metabolic Pathways 15
 - 3.1. Transcriptome Expression..... 17
- 4. Phylogenetic and Paralogous Region Analyses19
 - 4.1. Phylogenomics System 19
 - 4.2. Phylogeny Refinement..... 23
 - 4.3. Synteny of Paralogous Musa Regions..... 30

Figure 1: Genes search (Advance Search).....	4
Figure 2: Genome browser (GBrowse).....	6
Figure 3: Annotation editor (Artemis).....	8
Figure 4: Genetic Map (CMap).....	10
Figure 5: Gene report.....	11
Figure 6: Sucrose-phosphate synthase family (GreenPhylDB).....	15
Figure 7: Enzyme Sucrose-phosphate synthase, reaction 2.4.1.14 and pathway sucrose biosynthesis (Pathway Tools / MusaCyc).....	17
Figure 8: EST report (ESTtik).....	19
Figure 9: Phylogenomics system (GreenphylDB).....	21
Figure 10: Phylogenetic trees (GreenphylDB)	23
Figure 11: Upload data to the workflow manager (Galaxy).....	25
Figure 12: Editing a workflow (Galaxy).....	27
Figure 13: Running a workflow (Galaxy)	29
Figure 14: Phylogenetic trees (Galaxy).....	29
Figure 15: Dotplot of paralogous <i>Musa</i> regions coloured according to <i>Musa</i> ancestral blocks (PGDD)	32

In this document, we provide an overview of the functionalities of the Banana Genome Hub. For this purpose we selected a specific gene family to support the description. We studied the Sucrose phosphate synthase (SPS) family of the *Musa acuminata* double-haploid Pahang proteome using several information systems available from the Banana Genome Hub. *M. acuminata* is a giant herb of the monocotyledon class.

To start with the Banana Genome Hub, go to <http://banana-genome.cirad.fr/>

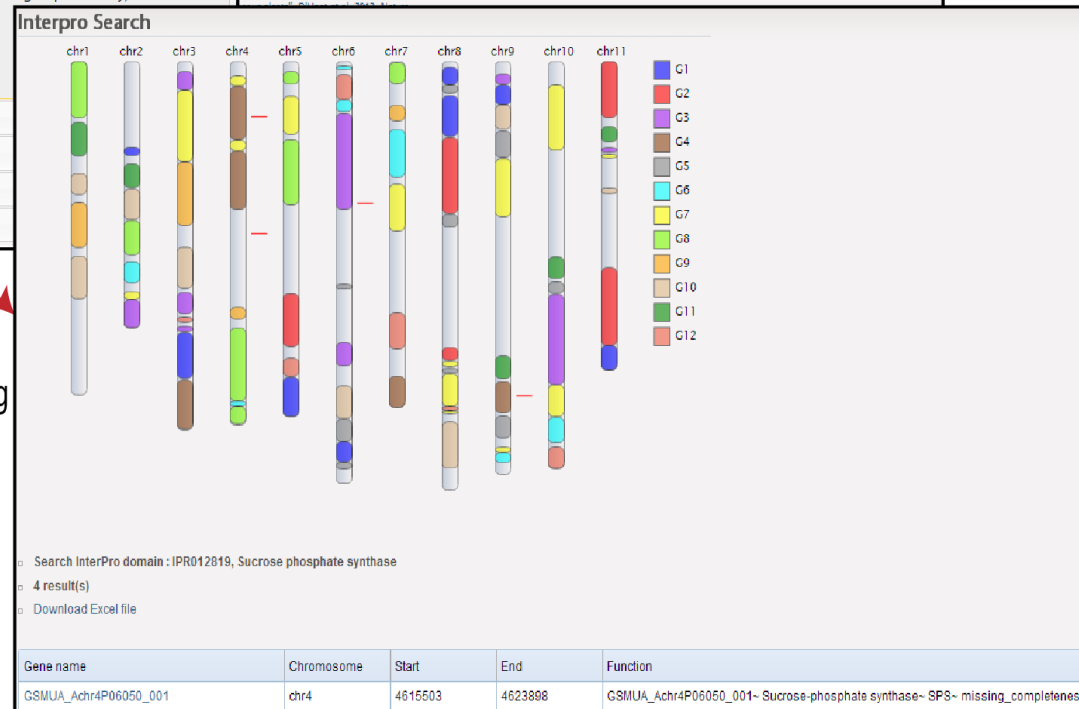
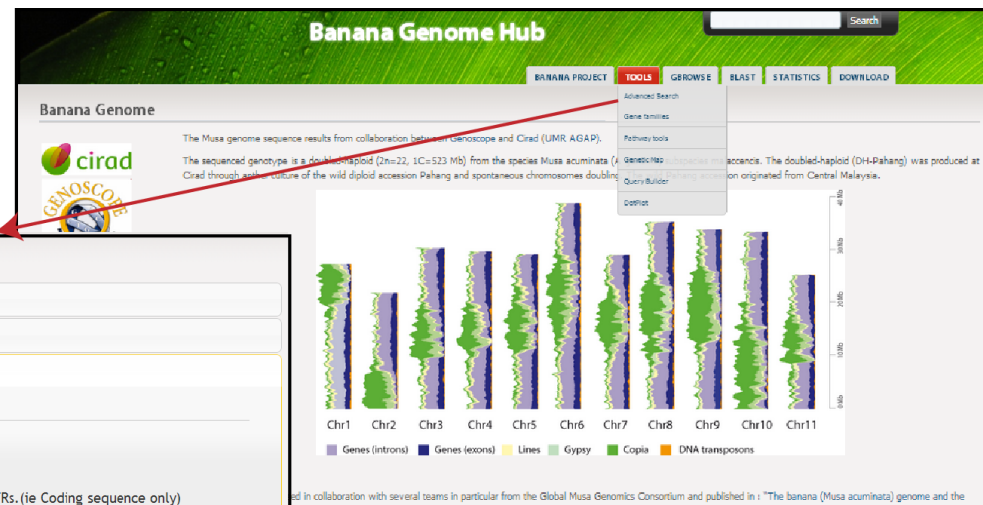
1. Search for the *Musa* SPS Gene Family

The advanced research tool is available in the "Tools" tab of the BananaGeome Hub.

The research can be made from several ways.

Search By

- Putative Function Search
- Locus Search
- InterPro Domain Search
 - Example : IPR000009 or try typing "GRA" to get a list of InterPro domain
 - Entry an InterPro domain
 - Output
 - FASTA file
 - ☐ CDS sequences. Does not include intron sequences or UTRs. (ie Coding sequence only)
 - ☐ Translated protein sequences.
 -
- GO Term Search
- Primer Designer
- Primer Blaster
- Location Search



Results are given as a list of gene with a karyotype showing the syntenic blocks and the position of gene

Figure 1: Genes search (Advance Search).

Method

Click on Advanced Search located in the dropdown menu TOOLS of the main menu.

<http://banana-genome.cirad.fr/advanced>

Gene retrieval can be performed using keywords, Locus, InterPro domains, GO Terms or genome location. Access the search field by clicking desired type of search. The loci of the resulting genes are located on the karyotypes and listed in a table with their genome location along the chromosome as well as their polypeptide product. To study the SPS, we searched for the InterPro IPR012819 corresponding to the Sucrose-phosphate synthase domain. By clicking on the name of one of the resulting gene (*e.g.* GSMUA_Achr4P06050_001) it leads to the genome browser, GBrowse.

Results

The four genes corresponding to this search are:

- GSMUA_Achr4P16070_001
- GSMUA_Achr4P06050_001
- GSMUA_Achr6P17480_001
- GSMUA_Achr9P22510_001

2. Structural Annotation

2.1. Genes

To analyze the structure of these genes, it is possible through the Community Annotation System (CAS) accessible via the dropdown menu TOOLS (http://banana-genome.cirad.fr/gene_annotation).

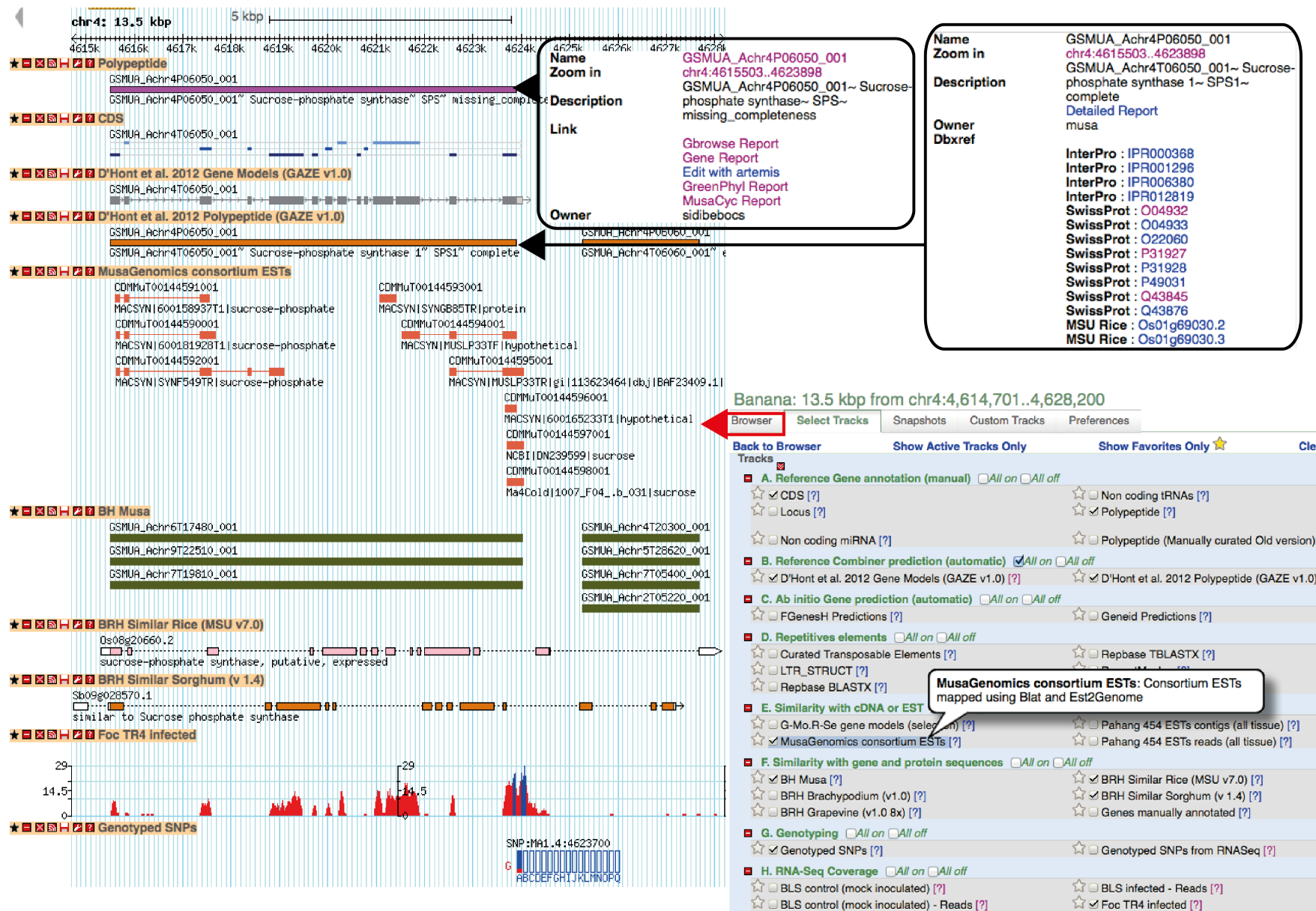


Figure 2: Genome browser (GBrowse).

Method

The Banana Genome Browser is accessible through links in various locations, (i) the most obvious is from the GBROWSE tab on the bar menu of the Banana Genome Hub, but (ii) it can be accessed by clicking on the name of a gene resulting from the Advanced Search.

(i) Search by name: you go to http://banana-genome.cirad.fr/cgi-bin/gbrowse/musa_acuminata/ then you search for GSMUA_Achr4P06050_001.

(ii) Search by region: you go directly to

http://banana-genome.cirad.fr/cgi-bin/gbrowse/musatract_tripal/?name=chr4:4615503..4623898

The choice of displayed tracks is available on the «Select Tracks» tab. Information about each feature is available via the question mark. By clicking on the tracks you can display more information and external links.

Looking at GSMUA_Achr4P06050_001 in the genome browser allows checking its intron/exon structure. Indeed, GBrowse provides information about the genomic region (chr4:4,615,503..4,623,898) and can aggregate various types of data such as similarities with EST, polypeptide sequences. The available information can be customized by checking/unchecking in track boxes. By clicking on any track feature, a pop up displays more information about the feature and provides external links such as cross references to other databases.

<http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR012819>

<http://www.uniprot.org/uniprot/P31927>

http://rice.plantbiology.msu.edu/cgi-bin/ORF_infopage.cgi?orf=LOC_Os08g20660.1

<http://www.uniprot.org/uniprot/Q6ZHZ1>

In the case of Reference Gene annotation (manual) track, the feature pop up provides a link to several tools of the Banana Genome Hub such as Artemis annotation editor (Figure 3) and gene reports (GBrowse; Tripal, Figure 5; GreenPhylDB, Figure 6 and Musacyc, Figure 7).

Results

In this case, we compared the automatic prediction of genes (D'Hont et al. 2012 Gene Models (GAZE v1.0) box of the Reference Combiner prediction (automatic) track) with the manual gene annotation (CDS box of the Reference Gene annotation (manual) track) and similarities with other genes and proteins (Best Hits, Reciprocal BH). The annotation comparison will allow to quickly checked whether the structure of the predicted gene (GSMUA_Achr4P06050_001 chr4:4,615,503..4,623,898) is correct. Structural modifications of size large enough should be visible in the Reference Gene annotation (manual) track. We can see by comparing those tracks, that exon 11 and 12 of GSMUA_Achr4P06050_001 predicted by GAZE were manually merged. These exons were merged for all of the four *Musa* SPS genes. The visualization of the GBrowse report shows that the annotation has been made by our team (e.g. owner: sidibebocs).

Artemis Entry Editor: chr4

File Entries Select View Goto Edit Create Run Graph Display

Entry: ☒ chr4

Selected feature: bases 3903 GSMUA_Achr4G05940_001 (/isObsolete=false /Name=GSMUA_Achr4G05940_001 /ID=GSMUA_Achr4G05940_001 /feature_id=210677 /tin

4613600 4614400 4615200 4616000 4616800 4617600 4618400 4619200

4623200 4624000 4624800 4625600

Exon-auto210796 0_001

GSMUA_Achr4T06050_001:CDS

GSMUA_Achr4T06050_001:3'UTR auto2104

4623200 4624000 4624800 4625600

Artemis Gene Builder: GSMUA_Achr4G06050_001

File Edit

GSMUA_Achr4G06050_001

GSMUA_Achr4T06050_001

auto210784, a

GSMUA_Achr4T06050_001

GSMUA_Achr4T06050_001

Annotation :: GSMUA_Achr4G06050_001

Key: polypeptide Add Qualifier: EC_number

Location: 4615503..4623898

Complement Refresh Grab Range Remove Range Goto Feature Select Feature TAI ObjectEdit

Properties Core CV References Match

GO terms	WITH/FROM	Dbxref	Evidence	Qualifier	Date
GO:0046524 [F]					----/--/--
GO:0009058 [P]					----/--/--
GO:0005985 [P]					----/--/--

Controlled Curation

Term	WITH/FROM	Dbxref	Evidence	Qualifier	Date
curated			IC :: ...		----/--/--
2.4.1.14					----/--/--
IC_2a					----/--/--
in_progress					----/--/--
SPS					----/--/--
complete					----/--/--

☐ Overview ☒ Tab View OK Cancel Apply

Artemis Entry Editor: chr4

File Entries Select View Goto Edit Create Run Graph Display

Entry: ☒ chr4

Selected feature: bases 3903 GSMUA_Achr4G05940_001 (/isObsolete=false /Name=GSMUA_Achr4G05940_001 /ID=GSMUA_Achr4G05940_001 /feature_id=210677 /tin

4613600 4614400 4615200 4616000 4616800 4617600 4618400 4619200

4623200 4624000 4624800 4625600

Exon-auto210796 0_001

GSMUA_Achr4T06050_001:CDS

GSMUA_Achr4T06050_001:3'UTR auto2104

4623200 4624000 4624800 4625600

Artemis Gene Builder: GSMUA_Achr4G06050_001

File Edit

GSMUA_Achr4G06050_001

GSMUA_Achr4T06050_001

auto210784, a

GSMUA_Achr4T06050_001

GSMUA_Achr4T06050_001

Annotation :: GSMUA_Achr4G06050_001

Key: polypeptide Add Qualifier: EC_number

Location: 4615503..4623898

Complement Refresh Grab Range Remove Range Goto Feature Select Feature TAI ObjectEdit

Properties Core CV References Match

GO terms	WITH/FROM	Dbxref	Evidence	Qualifier	Date
GO:0046524 [F]					----/--/--
GO:0009058 [P]					----/--/--
GO:0005985 [P]					----/--/--

Controlled Curation

Term	WITH/FROM	Dbxref	Evidence	Qualifier	Date
curated			IC :: ...		----/--/--
2.4.1.14					----/--/--
IC_2a					----/--/--
in_progress					----/--/--
SPS					----/--/--
complete					----/--/--

☐ Overview ☒ Tab View OK Cancel Apply

Artemis Entry Editor: chr4

File Entries Select View Goto Edit Create Run Graph Display

Entry: ☒ chr4

Selected feature: bases 3903 GSMUA_Achr4G05940_001 (/isObsolete=false /Name=GSMUA_Achr4G05940_001 /ID=GSMUA_Achr4G05940_001 /feature_id=210677 /tin

4613600 4614400 4615200 4616000 4616800 4617600 4618400 4619200

4623200 4624000 4624800 4625600

Exon-auto210796 0_001

GSMUA_Achr4T06050_001:CDS

GSMUA_Achr4T06050_001:3'UTR auto2104

4623200 4624000 4624800 4625600

Artemis Gene Builder: GSMUA_Achr4G06050_001

File Edit

GSMUA_Achr4G06050_001

GSMUA_Achr4T06050_001

auto210784, a

GSMUA_Achr4T06050_001

GSMUA_Achr4T06050_001

Annotation :: GSMUA_Achr4G06050_001

Key: polypeptide Add Qualifier: EC_number

Location: 4615503..4623898

Complement Refresh Grab Range Remove Range Goto Feature Select Feature TAI ObjectEdit

Properties Core CV References Match

GO terms	WITH/FROM	Dbxref	Evidence	Qualifier	Date
GO:0046524 [F]					----/--/--
GO:0009058 [P]					----/--/--
GO:0005985 [P]					----/--/--

Controlled Curation

Term	WITH/FROM	Dbxref	Evidence	Qualifier	Date
curated			IC :: ...		----/--/--
2.4.1.14					----/--/--
IC_2a					----/--/--
in_progress					----/--/--
SPS					----/--/--
complete					----/--/--

☐ Overview ☒ Tab View OK Cancel Apply

Enter Database Address

Server: medoc.cirad.fr

Port: 5432

Database: tripal_musa

User: banana_read_only

Password:

Available databases: Default

OK Annuler

Connecting ...

Figure 3: Annotation editor (Artemis).

Method

To edit a gene, you need to launch Artemis via the “Edit with Artemis” link in the GBrowse pop up of a banana gene or polypeptide or in the Gene Report (visible in figure 2).

<http://banana-genome.cirad.fr/cgi-bin/artemis.pl?species=musa;name=chr4:4590502..4648898>

There is two ways to use the annotation tool: if you have an account, you can participate to the community annotation system by modifying features stored into the Chado Database. If you do not have an account, you can ask an account (see <http://banana-genome.cirad.fr/content/annotation-account-request> on the http://banana-genome.cirad.fr/gene_annotation accessible in the TOOLS tab of the Banana Genome Hub Web site) or connect to Artemis with a read-only account: banana_read_only without password and modifications can be saved on your local computer.

Once Artemis is open, you can visualize the area around the gene of interest. By clicking on the Edit tab and then on the “Select Feature on the Editor” part, you can access the Artemis gene builder. In this builder you can access information such as comments of the curators that annotated this gene. Clicking on the commit button calls the Chado Controller annotation inspector module. The annotation Inspector validation window will appear. Clicking OK will commit the modifications whereas clicking the upper right cross will rollback to the original state. The changes are stored into the Chado database. To have more information about how to annotate a gene structure, see the Protocol for structural gene annotation and the Chado Controller publication.

<http://banana-genome.cirad.fr/documentation>

<http://www.gnpannot.org/biblio>

Like in GBrowse, you can have access the cross references for other databases, in the References tab of the gene builder.

Results

As mentioned before, we can see in the Artemis entry edit window that exons 11 and 12 of GSMUA_Achr4G16070_001 were merged.

In the annotator_comment qualifier of the Core tab of the gene builder, another correction is indicated: “one GC / AG intron-exon junction between exons 5 and 6” (exon five was extended from 14258178 to 14258199). This modification took into account the rare splicing site, GC, whereas the GAZE combiner predicted instead the classical GT.

2.2. Molecular Markers

!

Figure 4: Genetic Map (CMap).

Method

In order to get information about the genetic environment of the genes, we studied the nearest markers of the genes from the pop-up of the marker track on the GBrowse by clicking on “View marker on genetic map”.

http://banana-genome.cirad.fr/cmap_viewer?ref_map_set_acc=108;ref_map_accs=3150;highlight=111183;comparative_maps=1=map_acc=AF_PAHANG_LG4

CMap provides a physical map of the markers, and via an addition of map, links it to the genetic map (“Add a map” link).

A marker in CMap is linked to the Banana GBrowse and TropGeneDB

http://banana-genome.cirad.fr/cgi-bin/gbrowse/musa_acuminata/?name=mMaCIR1745

<http://tropgenedb.cirad.fr/tropgene/JSP/link.jsp?module=BANANA&value=mMaCIR1745&tref=marker&cref=name&tabIndex=0&onglet=MAPS&numero=10>

Results

We can see in this case that GSMUA_Achr4P16070_001 is in a region with a low recombination rate in the vicinity of the centromere. The closest actual marker mMaCIR1745 (14069511..14071481) is located at 182 kb in 5’ of GSMUA_Achr4G16070_001. The type of this marker is microsatellite, also known as Simple Sequence Repeats (SSRs). Quantitative trait locus (QTL) information generated on banana could be added when available, providing additional clues on putative function of the genes.

3. Functional Annotation

Knowing the structure of the genes of the SPS gene Family in Musa, we wanted to learn more about their function.

3.1. Gene Report

GSMUA_Achr4P06050_001, GSMUA_Achr4P06050_001 (polypeptide) Musa acuminata

Details

Name	GSMUA_Achr4P06050_001
Unique Name	GSMUA_Achr4P06050_001
Internal ID	210783
Length	1060
Type	polypeptide
Location	chr4:4615503..4623898
Organism	Musa acuminata (Banana)

Resources

- Details
- Sequence
- Relationships
- References
- Controlled vocabularies Assignments
- GO Assignments
- InterPro Report 2012-08-10
- TIGRHomologs
- SwissProtHomologs
- TrEMBLHomologs

External Link

- 1 ▫ View in Gbrowse
- 2 ▫ GreenPhyl Report
- 3 ▫ Edit with Artemis
- 4 ▫ MusaCyc Report
- 5 ▫ CoGe Report

- 1 - Return to gene visualisation on Gbrowse
- 2 - Link to Gene informations on Greenphyl
- 3 - Edition of the gene with the Artemis Software
- 4 - Link to the MusaCyc tool giving information about the pathway the gene are implied in
- 5 - Link to the CoGe tool

Figure 5: Gene report.

Method

To get detailed information about a banana gene, mRNA or polypeptide, the Gene Report (Tripal) is accessible through the quick search above the bar menu (in the banner) of the Banana Genome Hub (search for GSMUA_4P06050_001) and also via the “Gene Report” link in the GBrowse pop up of a gene or polypeptide (Figure 2).

http://banana-genome.cirad.fr/GSMUA_4P06050_001

The gene report provides resources about the protein-coding genes:

- Details : General information : Name, ID, length, type, Location, Organism
- Sequence : Sequence on fasta format
- Relationship : Other features linked in the Banana Genome Hub
- References : cross references to other external databases
- Controlled vocabularies Assignments
- GO Assignment
- InterPro Report : Links to Interpro Domains
- TIGRHomologs : Best Blast hits against MSU (ex-TIGR) Database
- SwissProtHomologs : Best Blast hits against SwissProt Database
- TrEMBLHomologs : Best Blast hits against TrEMBL Database

External links are also provided to lead to several tools giving complementary information about our gene of interest (GBrowse, Figure 2; Greenphyl, Figure 6; Artemis, Figure 3; MusaCyc, Figure 7, CoGe).

Results

Tableau 1 General functional information about the search results.

Name	Length	Location	InterPro Domains	GO Assignment
GSMUA_Achr4P06050_001	1060	chr4:4615503..4623898	IPR000368	GO:0009058
			IPR001296	GO:0005985
			IPR006380	GO:0046524
			IPR012819	
GSMUA_Achr4P16070_001	1083	chr4:14253795..14261027	IPR000368	GO:0009058
			IPR001296	GO:0005985
			IPR006380	GO:0046524
			IPR012819	
GSMUA_Achr9P22510_001	1062	chr9:27608557..27620605	IPR001296	GO:0009058
			IPR006380	GO:0005985
			IPR012819	GO:0046524
GSMUA_Achr6P17480_001	1044	chr6:11756040..11760311	IPR001296	GO:0009058
			IPR006380	GO:0005985
			IPR012819	GO:0046524

Three InterPro domains are common to all genes:

- IPR001296 : Glycosyl transferase, group 1
- IPR003680 : Sucrose-phosphate synthase
- IPR012819 : Sucrose phosphate synthase, plant

An additional InterPro domain was found only in GSMUA_Achr4P16070_001 and GSMUA_Achr9P22510_001: IPR000368 Sucrose synthase.


The GO Assignments were the same for the four genes:

- GO:0009058 : biosynthetic process
- GO:0005985 : sucrose metabolic process
- GO:0046524 : sucrose-phosphate synthase activity

The controlled vocabulary is available on the CV tab of the Artemis Gene Builder. To have more information about how to annotate a polypeptide function, see the Protocol for functional gene annotation.

<http://banana-genome.cirad.fr/documentation>

3.2. Gene Families



[Home](#)
[Quick search](#)
[Gene Family lists](#)
[Toolbox](#)
[Documentation](#)
[Statistics](#)
[My List](#)

Gene GSMUA_Achr4T06050_001

Sequence ID	GSMUA_Achr4T06050_001 add to my list
Species	Musa acuminata
Alias	No gene alias
Length	1038aa
Cross-reference(s)	<input checked="" type="checkbox"/> Display link(s) (1)
Gene Annotation	Sucrose-phosphate synthase 1
Gene Ontology	<input checked="" type="checkbox"/> Display term(s) (3)
Filter status	Not filtered

[Gene Classification](#)
[Gene Model](#)
[Gene Sequence](#)
[Domain Pattern](#)
[Ortholog Predictions](#)
[Similar Sequences \(BBMH\)](#)

ID	Name	Type
IPR001296	Glycosyl transferase, group 1	Domain
IPR006380	Sucrose-phosphate synthase	Domain
IPR012819	Sucrose phosphate synthase, plant	Family
IPR000368	Sucrose synthase	Domain

Note: sort multiple columns simultaneously by holding down the shift key and clicking other column headers.

[IPR001296](#) [IPR006380](#) [IPR012819](#) [IPR000368](#)




Figure 1: IPR domains for GSMUA_Achr4T06050_001

[Gene Classification](#)
[Gene Model](#)
[Gene Sequence](#)
[Domain Pattern](#)
[Ortholog Predictions](#)
[Similar Sequences \(BBMH\)](#)

[[FASTA](#) | [Excel](#) | [CSV](#) | [XML](#)]

Orthology

Sequence id	Species	Evolutionary Distance	Node Distance	GO
PDK_30s883591g005	Phoenix dactylifera	0.11	4	GO:0005985 GO:0009058 GO:0046524
PDK_30s713301g003	Phoenix dactylifera	0.113	4	GO:0005985 GO:0009058 GO:0046524
Os08g20660.1	Oryza sativa	0.181	6	GO:0005985 GO:0009058 GO:0046524 sucrose-phosphate synthase activity
Bradi3g20120.1	Brachypodium distachyon	0.186	5	GO:0005985 GO:0009058 GO:0046524
Sb09g028570.1	Sorghum bicolor	0.187	7	GO:0005985 GO:0009058 GO:0046524
GRMZM2G055331_P01	Zea mays	0.209	7	GO:0005985 GO:0009058 GO:0046524
Bradi1g30520.1	Brachypodium distachyon	0.338	6	GO:0005985 GO:0009058 GO:0046524
Os06g43630.1	Oryza sativa	0.351	7	GO:0005985 GO:0009058 GO:0046524
GRMZM2G471083_P01	Zea mays	0.357	7	GO:0005985 GO:0009058 GO:0046524
Sb10g025240.1	Sorghum bicolor	0.365	7	GO:0005985 GO:0009058 GO:0046524
GRMZM2G049076_P01	Zea mays	0.368	6	GO:0005985 GO:0009058 GO:0046524
Bradi3g06220.1	Brachypodium distachyon	0.37	8	GO:0005985 GO:0009058 GO:0046524
Os02g09170.1	Oryza sativa	0.387	9	GO:0005985 GO:0009058 GO:0046524
GRMZM2G462613_P01	Zea mays	0.392	10	GO:0005985 GO:0009058 GO:0046524
Bradi5g12860.1	Brachypodium distachyon	0.392	11	GO:0005985 GO:0009058 GO:0046524
Sb04g005720.1	Sorghum bicolor	0.392	11	GO:0005985 GO:0009058 GO:0046524
Cre12.g524000.t1.1	Chlamydomonas reinhardtii	1.41	10	GO:0005985 GO:0009058 GO:0046524
MDP0000265332	Malus domestica	1.49	5	
MDP0000203711	Malus domestica	1.58	5	
MDP0000790027	Malus domestica	1.96	5	
MDP0000311125	Malus domestica	2.01	6	
MDP0000255896	Malus domestica	2.06	6	

Note: sort multiple columns simultaneously by holding down the shift key and clicking other column headers.

[Gene Classification](#)
[Gene Model](#)
[Gene Sequence](#)
[Domain Pattern](#)
[Ortholog Predictions](#)
[Similar Sequences \(BBMH\)](#)

[[FASTA](#) | [Excel](#) | [CSV](#) | [XML](#)]

Best similar sequences	Species	Score	E-value	GO
At5g20280.1	Arabidopsis thaliana	1489	0	GO:0005985 GO:0009058 GO:0046524
Os08g20660.2	Oryza sativa	1627	0	GO:0005985 GO:0009058 GO:0046524 sucrose-phosphate synthase activity
Sb09g028570.1	Sorghum bicolor	1607	0	GO:0005985 GO:0009058 GO:0046524
POPTR_0018s01960.1	Populus trichocarpa	1607	0	GO:0005985 GO:0009058 GO:0046524
Bradi3g20120.1	Brachypodium distachyon	1612	0	GO:0005985 GO:0009058 GO:0046524
Medtr4g115620.1	Medicago truncatula	1478	0	GO:0005985 GO:0009058 GO:0046524

Note: sort multiple columns simultaneously by holding down the shift key and clicking other column headers.

Figure 6: Sucrose-phosphate synthase family (GreenPhylDB).

Method

In order to study further the SPS gene family, we clicked on the GreenPhyl gene report accessible from GBrowse or the Gene report through the “GreenPhyl Report” link.

http://www.greenphyll.org/cgi-bin/sequence.cgi?p=id&seq_textid=GSMUA_Achr4T06050_001

From the GreenPhyl report, we can access to the InterPro Domains and their Location, to BBMH and to orthologs prediction. For more see Figure 9 and Figure 10.

Results

The observation of the BBMH and the Orthologs confirmed the functional annotation of our genes. Indeed, the GO:0046524 sucrose Phosphatase activity is present in all of them, as well as InterPro IPR006380 Sucrose-phosphate synthase.

3.3. Metabolic Pathways

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

Home Search Tools Help Gene

***Musa acuminata* Enzyme: Sucrose-phosphate synthase**

Protein Sequence Nucleotide Sequence Nucleotide Sequence, Ad

Gene: [GSMUA_Achr4P06050_001](#) Accession Number: GSMUA_ACHR4P06050_001

Map Position: [4,615,503 -> 4,623,898] (15.36 centisomes) on Chromosome chr4 [Genome](#)

Unification Links: [Pmid:16876912](#), [Tigr:Os01g69030.2](#), [Tigr:Os01g69030.3](#), [Tigr:Os01g69030.4](#), [Tigr:Os02g09170.2](#), [Tigr:Os06g43630.1](#), [Tigr:Os08g20660.1](#), [Tigr:Os08g20660.2](#), [Trembl:C8TEV8](#), [Trembl:D6R6Y6](#), [Trembl:Q2ABX9](#), [Trembl:Q6SXU0](#), [Trembl:Q6U0U0](#), [Trembl:Q8H1Y1](#), [UniProtKB:O04932](#), [UniProtKB:O04933](#), [UniProtKB:Q22060](#), [UniProtKB:P31928](#), [UniProtKB:P49031](#), [UniProtKB:Q43845](#), [UniProtKB:Q43876](#)

Relationship Links: [InterPro:IN-FAMILY:IPR000368](#), [InterPro:IN-FAMILY:IPR001296](#), [InterPro:IN-FAMILY:IPR012819](#)

Gene-Reaction Schematic: [?](#)

GO Terms:

Biological Process: [GO:0005985 - sucrose metabolic process](#)
[GO:0009058 - biosynthetic process](#)

Molecular Function: [GO:0046524 - sucrose-phosphate synthase activity](#)

MultiFun Terms: [UNCLASSIFIED](#)

Enzymatic reaction of: sucrose-phosphate synthase (Sucrose-phosphate synthase)

D-fructose-6-phosphate + UDP-D-glucose <=> sucrose-6-phosphate + uridine-5'-diphosphate

The reaction direction shown, that is, A + B <=> C + D versus C + D <=> A + B, is in accordance with the Enzyme Commission system.

The reaction is favored in the direction shown.

In Pathways: [sucrose biosynthesis](#)

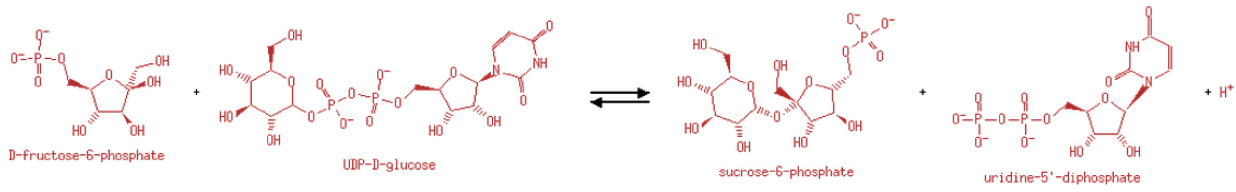
***Musa acuminata* Reaction: 2.4.1.14**

Species Comparison

Superclasses: [Reactions-Classified-By-Conversion-Type-> Simple-Reactions-> Chemical-Reactions-> EC-Reactions-> 2-- Transferases-> 2.4 -- Glycosyltransferases-> 2.4.1 -- Hexosyltransferases](#)
[Reactions-Classified-By-Substrate-> Small-Molecule-Reactions](#)

Enzymes and Genes:
[Sucrose-phosphate synthase: GSMUA_Achr9P22510_001](#)
[Sucrose-phosphate synthase: GSMUA_Achr6P17480_001](#)
[Sucrose-phosphate synthase: GSMUA_Achr4P16070_001](#)
[Sucrose-phosphate synthase: GSMUA_Achr4P06050_001](#)

In Pathway: [sucrose biosynthesis](#)



D-fructose-6-phosphate + UDP-D-glucose <=> sucrose-6-phosphate + uridine-5'-diphosphate + H⁺

The reaction direction shown, that is, A + B <=> C + D versus C + D <=> A + B, is in accordance with the Enzyme Commission system.

Most BioCyc compounds have been protonated to a reference pH value of 7.3, and some reactions have been computationally balanced for hydrogen by adding free protons. Please see the [BioCyc Guide](#) for more information.

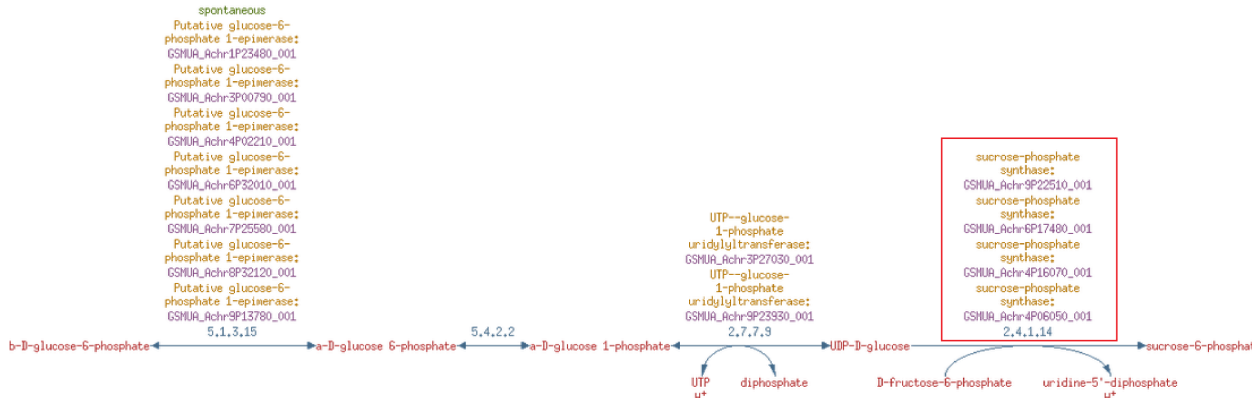
Enzyme Commission Primary Name for this Reaction: Sucrose-phosphate synthase

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

Home Search Tools Help Pathway

***Musa acuminata* Pathway: sucrose biosynthesis**

Show Predicted Enzymes More Detail Less Detail Species Comparison



spontaneous
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr1P23480_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr3P00790_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr4P02210_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr6P32010_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr7P25580_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr8P32120_001
 Putative glucose-6-phosphate 1-epimerase: GSMUA_Achr9P13780_001

5.1,3,15 5.4,2,2 2.7,7,9

UTP--glucose-1-phosphate uridylyltransferase: GSMUA_Achr3P27030_001
 UTP--glucose-1-phosphate uridylyltransferase: GSMUA_Achr9P23930_001

sucrose-phosphate synthase: GSMUA_Achr9P22510_001
 sucrose-phosphate synthase: GSMUA_Achr6P17480_001
 sucrose-phosphate synthase: GSMUA_Achr4P16070_001
 sucrose-phosphate synthase: GSMUA_Achr4P06050_001

2.4,1,14

b-D-glucose-6-phosphate a-D-glucose 6-phosphate a-D-glucose 1-phosphate UDP-D-glucose sucrose-6-phosphate

UTP H⁺ diphosphate D-fructose-6-phosphate uridine-5'-diphosphate H⁺

Figure 7: Enzyme Sucrose-phosphate synthase, reaction 2.4.1.14 and pathway sucrose biosynthesis (Pathway Tools / MusaCyc).

Method

MusaCyc displays information about all enzymes implied in the reaction and also in the pathway. It provides biological information about the reaction and the pathway. In order to get information about the metabolic pathways of the SPS gene family in plants, we clicked on the “MusaCyc Report” link available on GBrowse and the gene report.

http://banana-genome.cirad.fr/musacyc_report?GSMUA_Achr4P06050_001

As in other systems, you can have access to cross references to other databases, in the “Unification Links” and the “Relationship Links” paragraphs. If you click on the Enzyme Commission (EC) number (here 2.1.4.14), you can also have cross references to the BRENDA and ENZYME databases.

<http://enzyme.expasy.org/EC/2.4.1.14>

By clicking on the Enzymatic reaction, you can have access to the catalyzing step of sucrose biosynthesis from UDP-glucose and fructose- 6-phosphate.

<http://gohelle.cirad.fr:1555/MUSA/NEW-IMAGE?type=REACTION&object=SUCROSE-PHOSPHATE-SYNTHASE-RXN>

By clicking on the Pathway, you can have access to the sucrose biosynthesis pathway.

<http://gohelle.cirad.fr:1555/MUSA/NEW-IMAGE?type=PATHWAY&object=SUCSYN-PWY>

Results

The EC number 2.1.4.14 was predicted for GSMUA_Achr4P06050_001. It corresponds to the enzyme named the Sucrose-phosphate synthase catalyzing the last reaction of the sucrose biosynthesis pathway.

The four SPS polypeptides (GSMUA_Achr4P16070_001, GSMUA_Achr4P06050_001, GSMUA_Achr6P17480_001 and GSMUA_Achr9P22510_001) have an enzymatic activity characterized by the EC number 2.1.4.14.

3.1. Transcriptome Expression

Figure 8: EST report (ESTtik).

Method

Evidence of expression is sometimes available with ESTs from ESTtik system. You can go to ESTtik from the “ESTtik reference” link present in the GBrowse pop up of some EST of the track MusaGenomics consortium ESTs (group E. Similarity with cDNA or EST).

[http://esttik.cirad.fr/cgi-](http://esttik.cirad.fr/cgi-bin/public_quick_search.cgi?name_seq=MUC4RO1011_B12_b_093&type=singleton&project=MUC4_ROOTINVITRO&clean=cleaned&navigate=2&species=MusaESTtik)

[bin/public_quick_search.cgi?name_seq=MUC4RO1011_B12_b_093&type=singleton&project=MUC4_ROOTINVITRO&clean=cleaned&navigate=2&species=MusaESTtik](http://esttik.cirad.fr/cgi-bin/public_quick_search.cgi?name_seq=MUC4RO1011_B12_b_093&type=singleton&project=MUC4_ROOTINVITRO&clean=cleaned&navigate=2&species=MusaESTtik)

ESTtik report provides information about ESTs such as the sequence, the blast annotation and the database from which is issued the EST.

Results

In this case, the EST MUC4RO1011_B12_b_093 matches with GSMUA_AChr9T22510_001. As described in the detailed report, the EST has been reported on roots, that confirm the expression of this gene in *Musa* roots. Moreover the best blastx hit with nr databank is a *Musa* polypeptide (gb|AAC23914.1 in ESTtik Report) encoded by a CDS (GenBank:U59489.1) that has a *Musa* reference by Nascimento, J.R.O. et al. 1997 (PMID:9431676). They said:

“A 952-base pair polymerase-chain-reaction product of sucrose-phosphate synthase (SPS) (EC 2.3.1.14) from banana (*Musa acuminata* cv. Nanicão) fruit was cloned and used to study expression of the enzyme during development and ripening. The deduced amino acid sequence shows that banana SPS has a high similarity with the leaf, tap-root and bean enzymes from other species. Enzyme activity, and mRNA and protein levels point to an increase in SPS expression during ripening. The accumulation of sucrose was correlated to starch degradation and happened 4 d after SPS mRNA and activity had reached their maxima. These results indicate that access to substrate and transcriptional activation with increase in SPS expression might be important regulatory events of sweetening during banana fruit ripening. »

These proposals could be confirmed by analysis of recent results on *Musa* RNA-seq.

<http://www.ncbi.nlm.nih.gov/protein/3237273>

<http://www.ncbi.nlm.nih.gov/nuccore/3237272>


<http://www.ncbi.nlm.nih.gov/pubmed/9431676>

4. Phylogenetic and Paralogous Region Analyses

To explore the evolutionary history of the SPS gene family. We looked at the phylogenetic et paralogous relationship at different scales.

4.1. Phylogenomics System

Sequence	Species	Annotation	Clustering Level 1	Clustering Level 2	Clustering Level 3	Clustering Level 4
<input checked="" type="checkbox"/> GSMUA_Achr4T06050_001	Musa acuminata	Sucrose-phosphate synthase 1	GP000333	GP016032	GP040672	GP070110



[Home](#)
[Search](#)
[Gene Family lists](#)
[Toolbox](#)
[Documentation](#)

[Family Structure](#)
[Family Composition](#)
[Protein Domains](#)
[Protein List](#)
[Phylogenomic Analysis](#)
[Homolog Pred](#)

Family GP016032

Family ID	GP016032 add to my list
Family name	Sucrose-phosphate synthase subfamily suggest a name
Synonym(s)	SPS
Cross-reference(s)	PMID: 16876912
Family Ontology	<div> <input type="checkbox"/> Hide </div> <div> <p>molecular_function GO:0046524 (70%) - sucrose-phosphate synthase activity</p> <p>biological_process GO:0005985 (70%) - sucrose metabolic process</p> <p>GO:0009058 (80%) - biosynthetic process</p> </div>
Curation status	
Found in	Viridiplantae
Phylogenetic analyzes	Not available

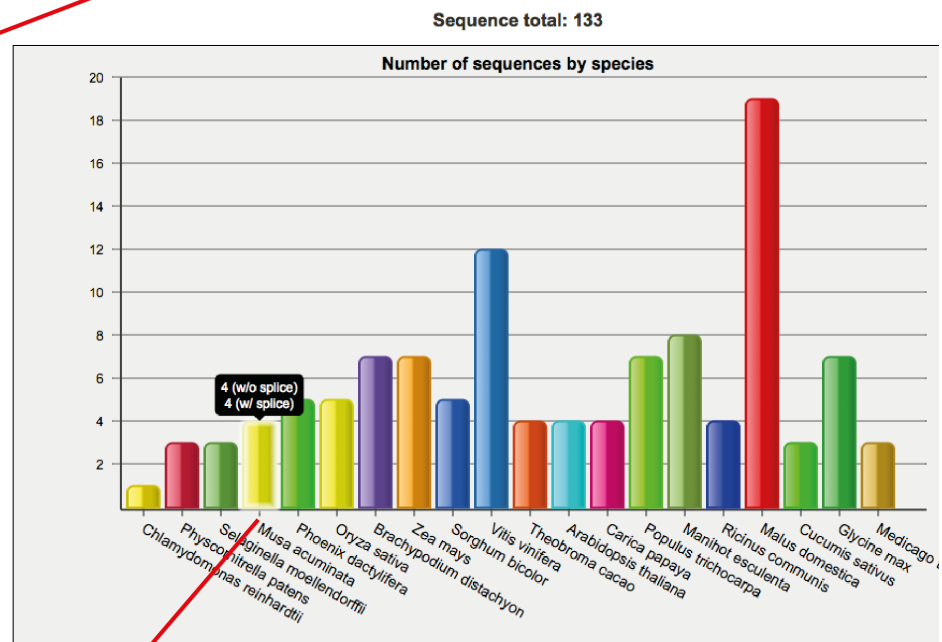


Figure 1: Number of sequence by species in current family. Lower bars indicate count of sequences without splice forms, brighter bars

[Excel](#)
[CSV](#)
[XML](#)
[FASTA](#)

Sequence	Species	Gene name	UniProt	InterPro	Annotation	Gene Ontology
<input checked="" type="checkbox"/> GSMUA_Achr4T16070_001	Musa acuminata			IPR001296 IPR006380 IPR012819 IPR000368	Sucrose-phosphate synthase	GO:0005985 GO:0009058 GO:0046524
<input checked="" type="checkbox"/> GSMUA_Achr6T17480_001	Musa acuminata			IPR001296 IPR006380 IPR012819	Sucrose-phosphate synthase 2	GO:0005985 GO:0009058 GO:0046524
<input checked="" type="checkbox"/> GSMUA_Achr9T22510_001	Musa acuminata			IPR001296 IPR006380 IPR012819	Sucrose-phosphate synthase 1	GO:0005985 GO:0009058 GO:0046524
<input checked="" type="checkbox"/> GSMUA_Achr4T06050_001	Musa acuminata			IPR001296 IPR006380 IPR012819 IPR000368	Sucrose-phosphate synthase 1	GO:0005985 GO:0009058 GO:0046524

Note: sort multiple columns simultaneously by holding down the shift key and clicking other column headers.

Figure 9: Phylogenomics system (GreenPhylDB)

Method

We can access to the gene families by clicking the links on gene families available from the GreenPhyl Report previously described (GSMUA_Achr4T06050_001; Figure 6),

http://www.greenphyl.org/cgi-bin/sequence.cgi?p=id&seq_textid=GSMUA_Achr4T06050_001

or by searching the family name (sucrose-phosphate synthase),

http://www.greenphyl.org/cgi-bin/family.cgi?p=id&family_id=16032

or by searching the identifier of the InterPro entry of family type (IPR012819).

http://www.greenphyl.org/cgi-bin/family.cgi?p=id&family_id=333

GreenPhyl provides gene clusters at 4 levels of stringency, gene families, Interpro domains, the orthologs prediction and the BBMHs, accessible by clicking on the respective tabs. Through the clusters links, information is provided about the gene family. You can get the Gene ontologies, the distribution of the families between the species, the InterPro domains, the list of proteins, the prediction of homologs and a phylogenetic analysis.

Results

The four genes previously found (GSMUA_Achr4T16070_001, GSMUA_Achr4T06050_001, GSMUA_Achr6T17480_001 and GSMUA_Achr9T22510_001) belong to a cluster curated as Sucrose Phosphate Synthase subfamily (cluster GP016032, second level of clustering), which is a subfamily of the superfamily named Sucrose Synthase (cluster GP000333, first level of clustering). In this particular case, the phylogenetic analysis was not available for the sucrose phosphate synthase family, but available at the upper level.

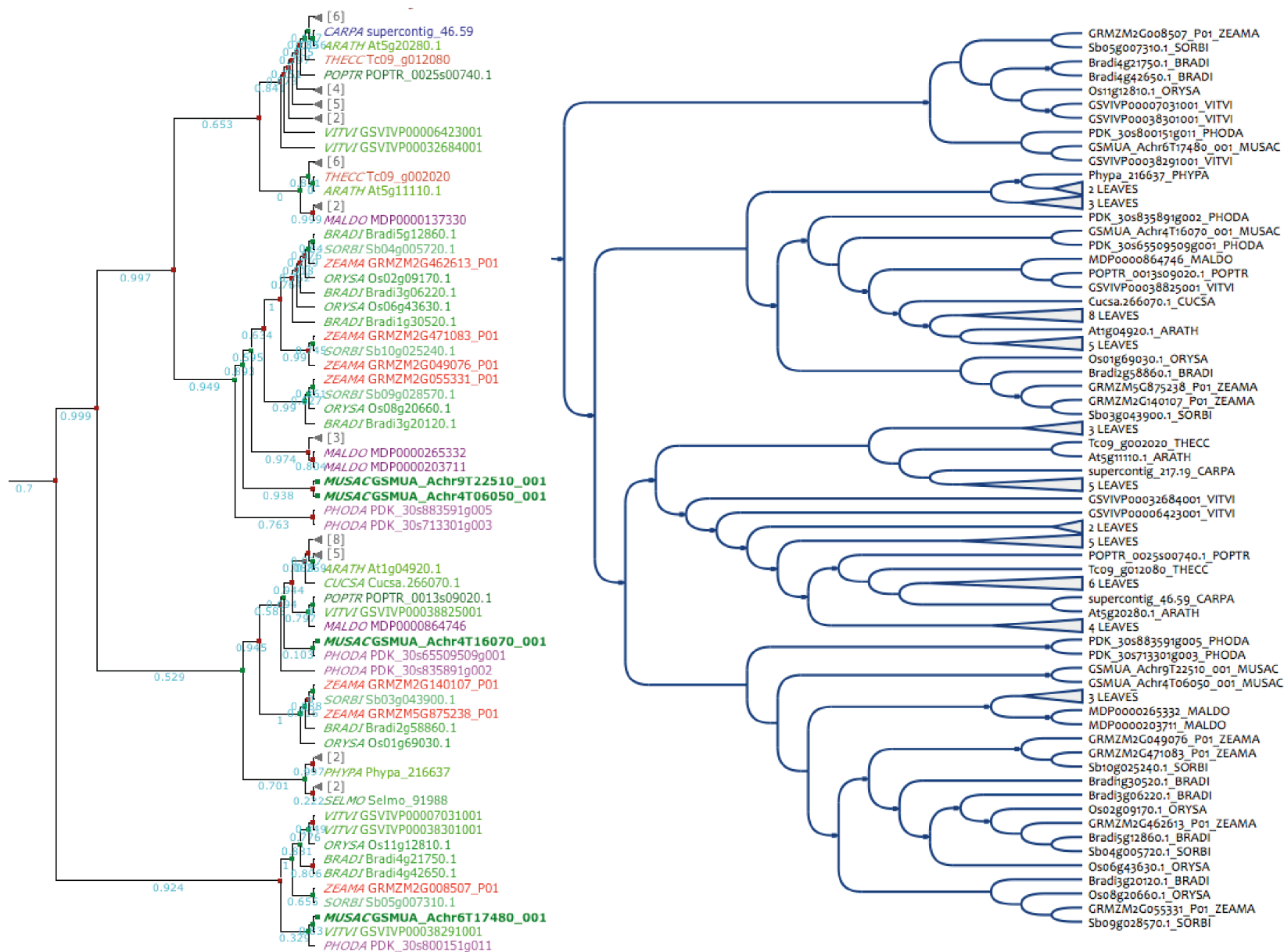


Figure 10: Phylogenetic trees (GreenphylDB)

Method

The phylogenetic tree of this family is accessible from the “Phylogenomic Analysis” tab and then the button “View gene tree with APTX”.

An alternate visualization is proposed through the button “View gene tree with IntreeGreat”. InTreeGreat is developed in the context of the SouthGreen platform.

The *Musa* SPS polypeptides were flagged with green squares whereas Arabidopsis polypeptides were flagged with avocat squares.

Results

We noticed in this phylogeny that two of our genes of interest are really close together: it appeared that GSMUA_Achr9G22510_001 and GSMUA_Achr4G06050_001 were the result of a recent duplication.

4.2. Phylogeny Refinement

Family Structure

Family Composition

Protein Domains

Protein List

Phylogenomic Analysis

Homolog Predictions

Build your own display: 1

Species

☐ Check/uncheck all

☒ Oryza sativa
 ☒ Arabidopsis thaliana
 ☐ Medicago truncatula

☒ Sorghum bicolor
 ☐ Populus trichocarpa
 ☐ Vitis vinifera

☐ Physcomitrella patens
 ☐ Selaginella moellendorffii
 ☐ Glycine max

☐ Chlamydomonas reinhardtii
 ☐ Brachypodium distachyon
 ☐ Carica papaya

☐ Ricinus communis
 ☐ Zea mays
 ☒ Musa acuminata

☐ Theobroma cacao
 ☐ Manihot esculenta
 ☐ Malus domestica

☐ Cucumis sativus
 ☐ Phoenix dactylifera

Information to display

☒ Gene name
 ☒ UniProt
 ☒ InterPro
 ☒ Annotation
 ☒ Gene Ontology
 ☒ EC
 ☒ PubMed
 ☒ Filter Status

Filter state: any

Display table

Sequence	Species	Gene name	UniProt	InterPro	Annotation	Gene Ont
<input checked="" type="checkbox"/> At1g04920.1	Arabidopsis thaliana			IPR001296 IPR009380 IPR012812 IPR000368	I Symbolic ATP5SF1 ATP5SF1 (sucrose phosphate synthase 3F) sucrose- phosphate synthase/ transferase, transferring glycosyl groups chr1:1391674- 1395756 protein	GO:000598 GO:000905 GO:000422

Galaxy | Excel | CSV | XML | FASTA

The following job has been successfully added to the queue:

234: Banana Genome

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

You are now being redirected back to Galaxy

Search By

Putative Function Search

Locus Search

InterPro Domain Search

GO Term Search

Primer Designer

Primer Blaster

Location Search

Example : IPR000009 or try typing "GRA" to get a list of InterPro domain

Entry an Inter Pro domain

IPR012819, Sucrose pl

Output

FASTA file

☐ CDS sequences. Does not include intron sequences or UTRs. (ie Coding sequence only)
 ☒ Translated protein sequences.

Action

☒ Send output to Galaxy
 Run Reset

Tools

Get Data

Send Data

Tools

Convert Formats

Evolution

EST/BL

Filter and Sort

Gene/Protein prediction

SAI

NGS: Quality Control

NGS: Mapping

NGS: SAM/BAM Manipulations

NGS: Coverage

NGS: SNP Detection

NGS: Structural Variation

NGS: Variant Calling

Fetch Sequences

Protein Structures

Sequence comparisons

Genomics

UNTESTED TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Alignments

Operate on Genomic Intervals

Statistics

Get Data

VCFToFastaAlignments

Creates a collection of FASTA alignments from a VCF output obtained from GATK

Upload File from your computer

UCSC Main table browser

UCSC Test table browser

UCSC Archive table browser

BX table browser

Get Microbial Data

Galaxy

Analyze Data

Workflow

Shared Data

Visualization

Help

User

South Green

bioinformatics platform

Welcome to GALAXY

... at your disposal as part of the services provided by SouthGreen

In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "southgreen" in the lookup filter.

Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr

The GALAXY project is supported in part by IISG (INRA) and the many Institutes of the Life Sciences

History

Greenphyl for HUBMUSA

33.9 MB

234: Banana Genome

233: Banana Genome

232: Banana Genome

230: Rapgreen statistics

229: Reconciled Newick

228: Output species

227: Output phyloxml

226: Newick

225: Phyml statistics

224: Newick Tree

221: Gblocks on data

220: MAFFT on data

215: Newick

209: Aligned

24

Figure 11: Upload data to the workflow manager (Galaxy)

Method

As the SPS genes have been manually annotated, we wanted to study if the phylogeny would be different. We performed a new phylogenetic analysis via an in-house Galaxy Workflow corresponding to the one implemented in GreenPhylDB.

<http://gohelle.cirad.fr/galaxy/>

Data uploading to Galaxy can be made from the Advanced Search tool of the Banana Genome Hub or from the Protein List tab in GreenPhylDB, after choosing species. You can also upload our own data through the uploading tool of Galaxy. Once the data uploaded, they will appear in green on the right of the screen. They will be available with a number that can be used by a Galaxy tool. Even if you are not logged in, you can use the Galaxy tools accessible in the left panel of the screen.

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 34.1 MB

Workflow renamed to 'Workflow for phylogenomic analysis based on proteic sequences'

Your workflows

Name	# of Steps
Workflow for phylogenomic analysis based on nucleic sequences	9
Workflow for phylogenomic analysis based on proteic sequences	7

Edit
 Run
 Share or Publish
 Download or Export
 Clone
 Rename
 View
 Delete

Create new workflow Upload or import workflow

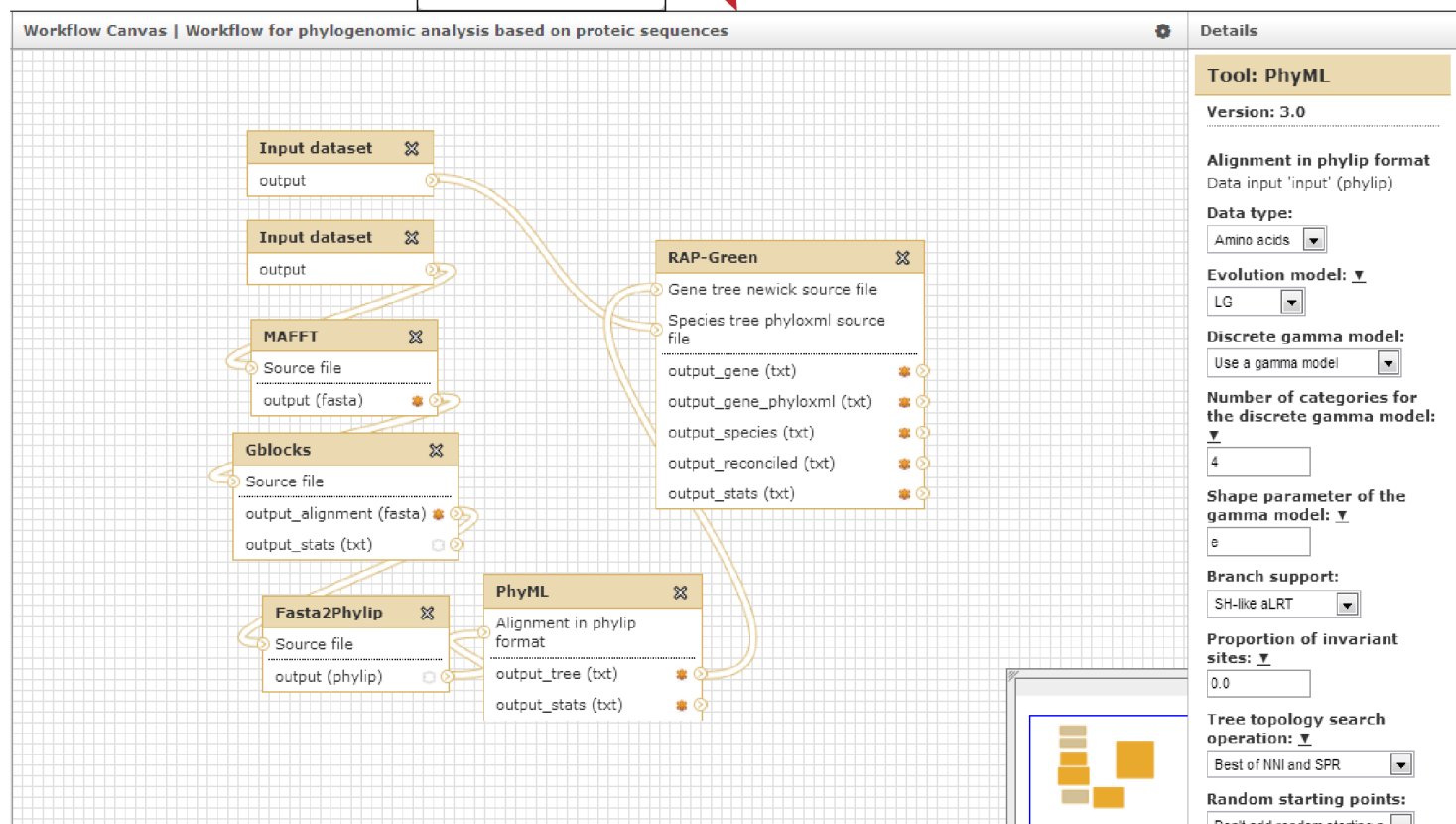


Figure 12: Editing a workflow (Galaxy)

Method

For users who have an account, the workflows are available on the Workflows tab. For those who do not have one, the workflows will not be available, but the tools are still accessible. You can apply for a account in the computer center section of the SouthGreen Web site.

http://gohelle.cirad.fr/cluster/compte_galaxy.php

The workflows can be modified in the edit section of the workflow menu accessible by clicking on the arrow next to the workflow's name. The different steps are visible in the central part of the screen. The parameters can be modified for each step in the Details panel.

Results

The workflow “GreenPhyl Phylogenomic analysis workflow (nucleic)” displayed in this example is accessible in the Shared Data / Published Workflows menu of the Galaxy web site.

http://gohelle.cirad.fr/galaxy/workflow/list_published

The protein version is also available (“GreenPhyl Phylogenomic analysis workflow”).

Galaxy

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 34.1 MB

Workflow renamed to 'Workflow for phylogenomic analysis based on proteic sequences'

Run

Your workflows

Name	# of Steps
Workflow for phylogenomic analysis based on nucleic sequences	9
Workflow for phylogenomic analysis based on proteic sequences	7

Running workflow "Workflow for phylogenomic analysis based on proteic sequences"

Expand All Collapse

Step 1: Input dataset

Input phyloxml species tree file

1: viridiplantae.phyloxml

type to filter

Step 2: Input dataset

Fasta sequences

234: Banana Genome

type to filter

Step 3: MAFFT (version 1.0.0)

Step 4: Gblocks (version 1.0.0)

Step 5: Fasta2Phylip (version 1.0.0)

Step 6: PhyML (version 3.0)

Step 7: RAP-Green (version 1.0.0)

☐ Send results to a new history

Run workflow

Successfully ran workflow "Workflow for phylogenomic analysis based on proteic sequences". The following datasets have been added to the queue:

1: viridiplantae.phyloxml

234: Banana Genome

235: MAFFT on data 234

236: Gblocks on data 235

237: Gblocks on data 236

238: Fasta2Phylip on data 236

239: Newick Tree

240: Phylml statistics output

241: Newick

242: Output phyloxml

243: Output species

244: Reconciled Newick

245: Rapgreen statistics

History

GreenPhyL_for_HUBMUSA

34.0 MB

245: Rapgreen statistics

244: Reconciled Newick

243: Output species

242: Output phyloxml

241: Newick

240: Phylml statistics output

239: Newick Tree

238: Fasta2Phylip on data 236

236: Gblocks on data 235

235: MAFFT on data 234

234: Banana Genome

233: Banana Genome

235: MAFFT on data 234

4 sequences

format: fasta, database: 2

>GSMR_Achr4F06090_001 Sucrose-phosph

MAQNDHNSYLEATLDAGPSI---DAAGAGLLIRER

FPPTATVVEEVINGPDTDLVKWVRRAAMGSPQGR

EAQGISWRLERERARADATADMSDLSGEGKSID

AKASQYKDK---LYIVLISHLIRGENMELARDSC

VLLITRQILAPVDMSYGEPTMLTPRESEPTMETC

Figure 13: Running a workflow (Galaxy)

Method

The workflow is launched from the run section. Choose the entry data and then click on the “run workflow” button. The steps will be shown on the history, in grey while they are waiting and in yellow while process. The workflow finishes running when all the steps are green. If an error occurs, the steps will be red. You can visualize and download result files of each step by clicking on it.

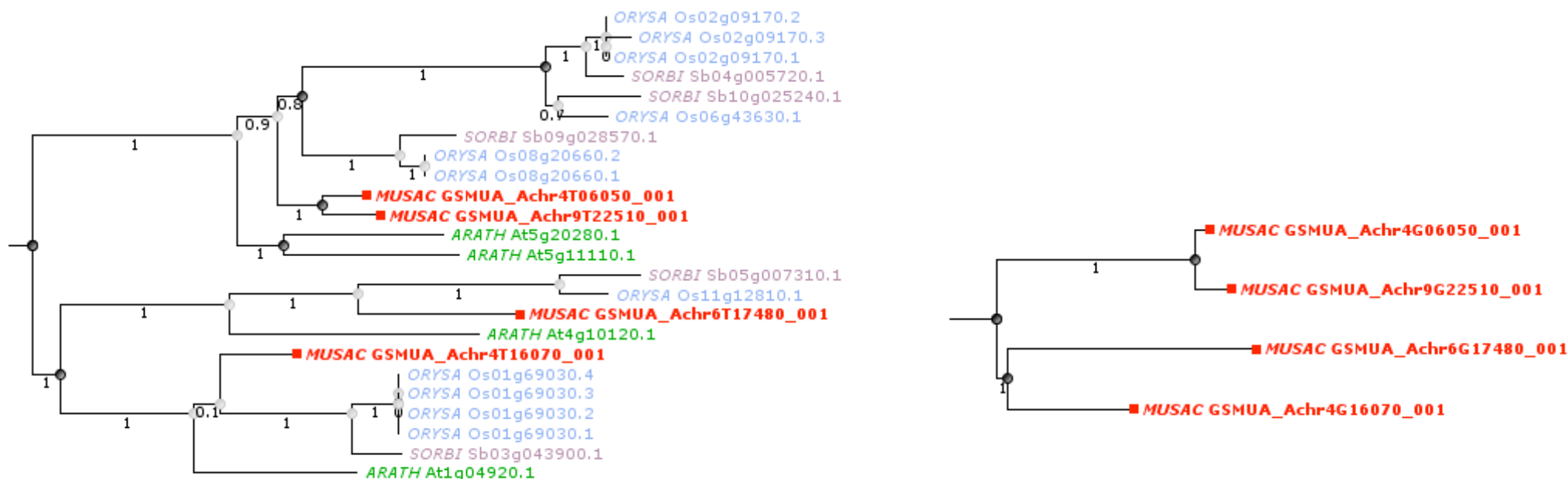


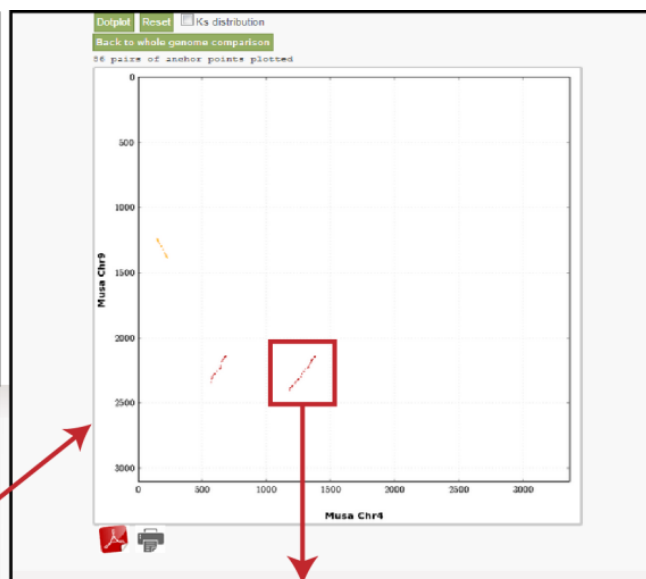
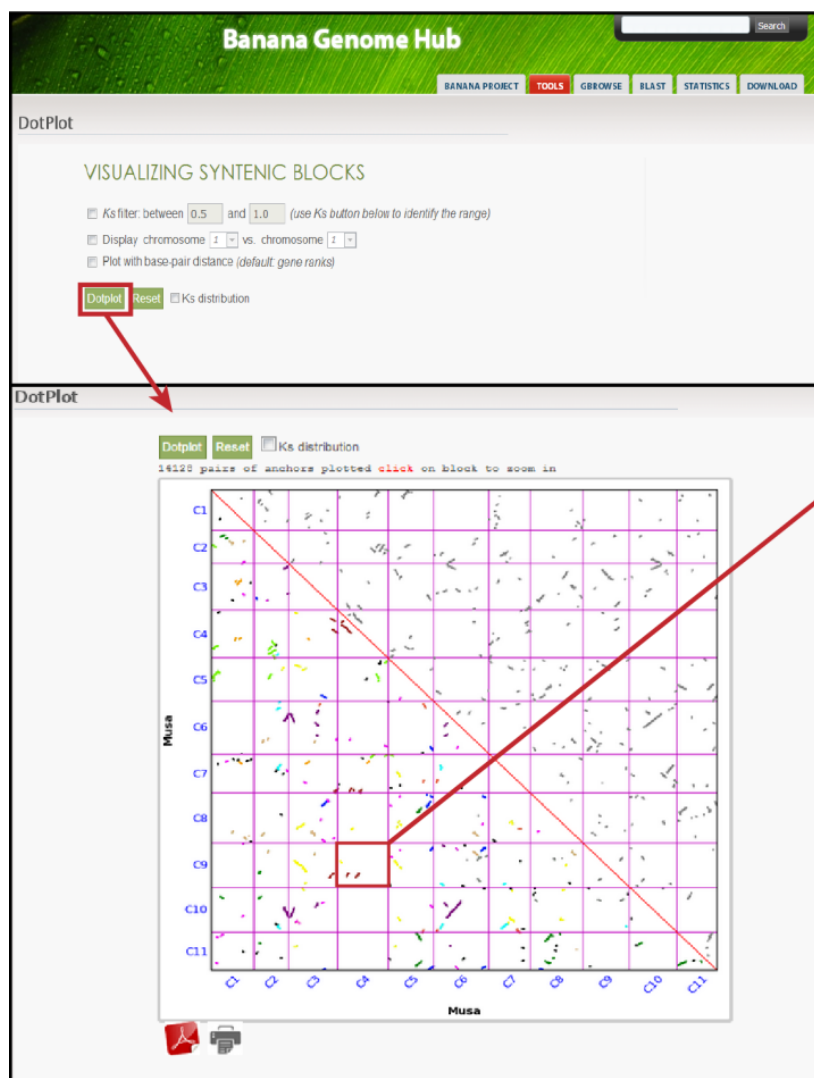
Figure 14: Phylogenetic trees (Galaxy)

Results

For the protein analysis, species considered are Oryza sativa (ORYSA), Sorghum bicolor (SORBI), Arabidopsis thaliana (ARATH) and Musa acuminata (MUSAC). We ran the phylogeny on fewer species and the phylogeny did not change significantly. The genes GSMUA_Achr9G22510_001 and GSMUA_Achr4G06050_001 were still very close together.

For the nucleotide analysis, only *Musa* was considered. The results are consistent; the duplication between the genes GSMUA_Achr9G22510_001 and GSMUA_Achr4G06050_001 is very more recent than between the other genes of the family in *Musa*.

4.3. Synteny of Paralogous Musa Regions



(1) Block (Score 0.0, E-value 0.0) with 282 anchors.

Order within block	Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks
1	GENUA_Achr4P05820_001	Putative uncharacterized protein	GENUA_Achr9P23380_001	Hypothetical protein	0.28	0.80
2	GENUA_Achr4P05720_001	Putative Squamosa promoter-binding-like protein 12	GENUA_Achr9P23140_001	Putative Squamosa promoter-binding-like protein 12	0.20	0.48
3	GENUA_Achr4P05730_001	Putative U-box domain-containing protein 51	GENUA_Achr9P23090_001	serine threonine kinase, putative, expressed	0.22	1.45
4	GENUA_Achr4P05760_001	Whole genome shotgun sequence of line P140024, scaffold_62, assembly12k (fragment)	GENUA_Achr9P23010_001	diphthamide biosynthesis protein 3, putative, expressed	2.14	2.70
5	GENUA_Achr4P05780_001	60S Ribosomal protein L7-2	GENUA_Achr9P22960_001	60S Ribosomal protein L7-6	0.14	0.70
6	GENUA_Achr4P05830_001	4-coumarate--CoA ligase 1	GENUA_Achr9P23840_001	Probable 4-coumarate--CoA ligase 3	0.06	1.59
7	GENUA_Achr4P05890_001	transcription factor like protein, putative, expressed	GENUA_Achr9P22780_001	transcription factor like protein, putative, expressed	0.13	0.76
8	GENUA_Achr4P05930_001	expressed protein	GENUA_Achr9P22760_001	expressed protein	0.17	0.50
9	GENUA_Achr4P06010_001	Hypothetical protein	GENUA_Achr9P22750_001	Eukaryotic initiation factor iso-4' subunit p87-34	0.58	3.09
10	GENUA_Achr4P06100_001	Floral homeotic protein APETALA 2	GENUA_Achr9P22540_001	Hypothetical protein	0.19	0.76
11	GENUA_Achr4P06340_001	zinc finger, C2H2 type domain containing protein, expressed	GENUA_Achr9P22360_001	Hypothetical protein	1.71	2.44
	Whole genome shotgun		Whole genome shotgun			

Figure 15: Dotplot of paralogous *Musa* regions coloured according to *Musa* ancestral blocks (PGDD)

Method

We then tried to determine if the high similarity of these genes was due to block duplication in *Musa*. For that, we used the Plant Genome Duplication Database (PGDD) system that provides in our case, paralogous regions between *Musa* chromosomes, coloured according to *Musa* ancestral blocks. The hypothesis is that these paralogous regions come from whole genome duplications (WGD).

PGDD, is available from the DotPlot section of the “Tools” tab (“Syntenic Dotplot” link).

<http://banana-genome.cirad.fr/dotplot>

To examine the synteny, we clicked on the Dotplot button without giving any parameters. You can zoom afterwards on a square of the dot plot to visualize the Dotplot between two chromosomes by clicking on a square in the global plot or by selecting the “Display chromosome A vs chromosome B” option. The scale then represents the gene rank. By clicking on a syntenic region, you get the list of co-localized paralogs with the values of Ks and Ka.

Results

We studied the microsynteny between the chromosomes 4 and 9 in order to conclude if the genes GSMUA_Achr9G22510_001 and GSMUA_Achr4G06050_001 that appeared very similar on the phylogeny (Figure 10 and Figure 14) were issued from the same duplicated ancestral block. We noticed that the genomic regions are issued from WGD event of the ancestral block 4 (**Figure 1**), but our two genes are not listed as syntenic in this table. Further studies will be required to confirm whether or not the duplication between GSMUA_Achr9G22510_001 and GSMUA_Achr4G06050_001 is linked to a WGD event.