### GREENPHYLDB A phylogenomic database for plant comparative genomics

http://greenphyl.cirad.fr

Matthieu CONTE, Sylvain GAILLARD, Nadège LANAU, Mathieu ROUARD and Christophe PERIN\*.

Corresponding author: perin@cirad.fr CIRAD: French Agricultural Research Center for International Development .



#### Overview

The increasing amount of sequence data provided by full or partial genome sequencing projects urgently needs a way to transfer information from model species to new sequenced ones. Orthologous and paralogous genes identification is now a major objective for gene function prediction as orthologous sequences are more likely to share the same function than paralogous sequences.

#### Results

We developed **GreenPhyIDB**, a comprehensive platform for comparative genomics analyses of **Arabidopsis thaliana** and **Oryza sativa** full genomes. First **GreenPhyIDB** offers the most complete plant family list with 6417 manually curated families. Most of these families have been analyzed with our optimized phylogenomic pipeline (4341 families). Thus GreenPhyIDB is the only plant database that provides an indicator of phylogenomic accuracy prediction.

## GreenPhyIDB



# > GOST: GreenPhyl Orthologs Search Tool

Step 1: Paste your protein sequence

**GOST** is a powerful tool developed to predict phylogenomic relationships between any plant gene and *O. sativa / A. thaliana* gene(s). This tool rapidly integrates a new sequence into a precomputed phylogenic tree developed on *O. sativa* and *A. thaliana* to infer orthologs relationships.

Step 2: Integration of your query into a GreenPhyIDB family

			GOST				Your sequence was inferred to be part of this GreenPhyl family:			
			(GreenPhyl Orthologs Search Tool)				Christering level	Earritz Name	Earniby int	
			GOST is a powerful tool developed to predict phylogenomic relationships between any plant gene and O.sativa / A.thaliana gene new sequence into a model phylogeny developed on O.sativa and A.thaliana to infer orthologs relationships.				MCL_I1.2	GRAS transcription factor family	20939	
Phylogenomic prediction on RHT1						).	MGL_12 MCL_13		25010	
wheat gene from GRAS transcription		Fland Vances	Stand. Your company will be attributed to a GreenDivid family				MCL_I5	DELLA subfamily	37810	
factor family	Step1. Your se	Step1. Your sequence will be interneted to a Greenwijn taniny				The phytogenomics analysis for the cluster "2099" named GKA's transcription factor family is available here				
Tactor family.			quence will be integrated into the p	nyiogenomic tree of the selected fa	iny .		Select the species of your query in t	the list hereunder;		
This gene have been published to be ortholog to <i>Arabidopsis thaliana</i> Gibberellin Insensitive (GAI) gene involved in gibberellin transduction. PMID: 10421366			Pask over PHOTEN sequence in task hond protection 10 (PHOTEN Sequ				Cere advance (ZCREU) Materia Tree     Academic processing VACPM Manderkaar for     Academic processing VAC			
			Find family Clear				Our plant tree of life Link to NCBI Taxonomy			
			Tool developed and maintained by MG. Conte, M. Rouard and G. Droc				Search homologs			
Step 4: Phylogenomic integration of your sequence into the selected family Step 5: Generated tree including your query										
					-	🍰 ATV				
Phylogenomic integration							File Edit Searc	ch View Options Help		
	Step 1/5 Input analyzes								seq name ext nodes	
Step 2:5 Align query to family alignment								Os05g49930.1_ORYZA	seq name int nodes species ext nodes	
	to query	query					Cs01a45880.1 ORYZA	EC ext nodes		
	nomic tree							EC int nodes branch length values		
		of the second states					A12g01570.1_ARATH [87	duplic vs spec		
	aryses of the gene	ated tree					Attg14920.1 ARATH [87]	display orthology display subtr-neighb		
JOB DONE			View associated tran						color accord to log L	
			view generated i	ee			→   -+	At1966350.1_ARATH [ 77 ]	editable	
	Orthologs found							A15x17400.1 APATH ( 90.1	click on node to: C display/edit information	
Orthology @ Subtree-neighbor @ SuperOrthology @ Distance									C collapse/uncollapse C go to swiss-prot	
	Id	Ortholog 5	SubtreeNeighbor %	SuperOrthogy %	Distance			At3g03450.1_ARATH [ 80 ]	C root/reroot	
GreenPhyl	Os03a49990.1	87	100	87	0.03184				C swap children	
Dhula non amia nas distisma	At1g14920.1	87	97	0	0.27414			©s03g49990.1_ORYZA [87]	zoom in X	
Phylogenomic predictions	At2g01570.1	87	97	0	0.31311				zoom in Y	
	At3g03450.1	80	90	0	0.33925			Your query TRIAE [Q]	C	
	At5g17490.1	80	90	0	0.36147				show whole	
	At1g66350.1	77	87	0	0.34259				<ul> <li>uncolapse all</li> </ul>	
	Os11g11600.1	5	6	5	1.73449		<		> collapse to deepest	
	L									