Plant ortholog database

Yang Li and Dawei Huang

Dec 31, 2014

Arabidopsis_thaliana.TAIR10.24.pep.all.fa
Brassica_rapa.IVFCAASv1.24.pep.all.fa
Glycine_max.V1.0.24.pep.all.fa
Oryza_barthii.ABRL00000000.24.pep.all.fa
Oryza_brachyantha.Oryza_brachyantha.v1.4b.24.pep.all.fa
Oryza_glaberrima.AGI1.1.24.pep.all.fa
Oryza_glumaepatula.ALNU02000000.24.pep.all.fa
Oryza_indica.ASM465v1.24.pep.all.fa
Oryza_meridionalis.ALNW00000000.24.pep.all.fa
Oryza_nivara.AWHD00000000.24.pep.all.fa
Oryza_punctata.AVCL00000000.24.pep.all.fa
Oryza_rufipogon.PRJEB4137.24.pep.all.fa
Oryza_sativa.IRGSP-1.0.24.pep.all.fa
Populus_trichocarpa.JGI2.0.24.pep.all.fa
Solanum_lycopersicum.SL2.40.24.pep.all.fa
Sorghum_bicolor.Sorbi1.24.pep.all.fa
Zea_mays.AGPv3.24.pep.all.fa

The purpose of IC4R is to build a rice reference genome with standard and comprehensive annotations, based on a variety of omics data and literature.

Rice Gene Expression Database (RiceGED)

Collaboration with Dr. Song-nian Hu (under development)

Rice Non-coding RNA Database (RiceNRD)

In preparation

Rice Variation Database (RiceVD)

Collaboration with Dr. Xing-Wang Deng (under development)

IC4R

http://www.ic4r.org

(Sequence database)

Rice Methylation Database (RiceMD)

Collaboration with Drs. Yingfeng Luo & Rujiao Li (In preparation)

Rice Literature Miner

Homolog database

RiceWiki

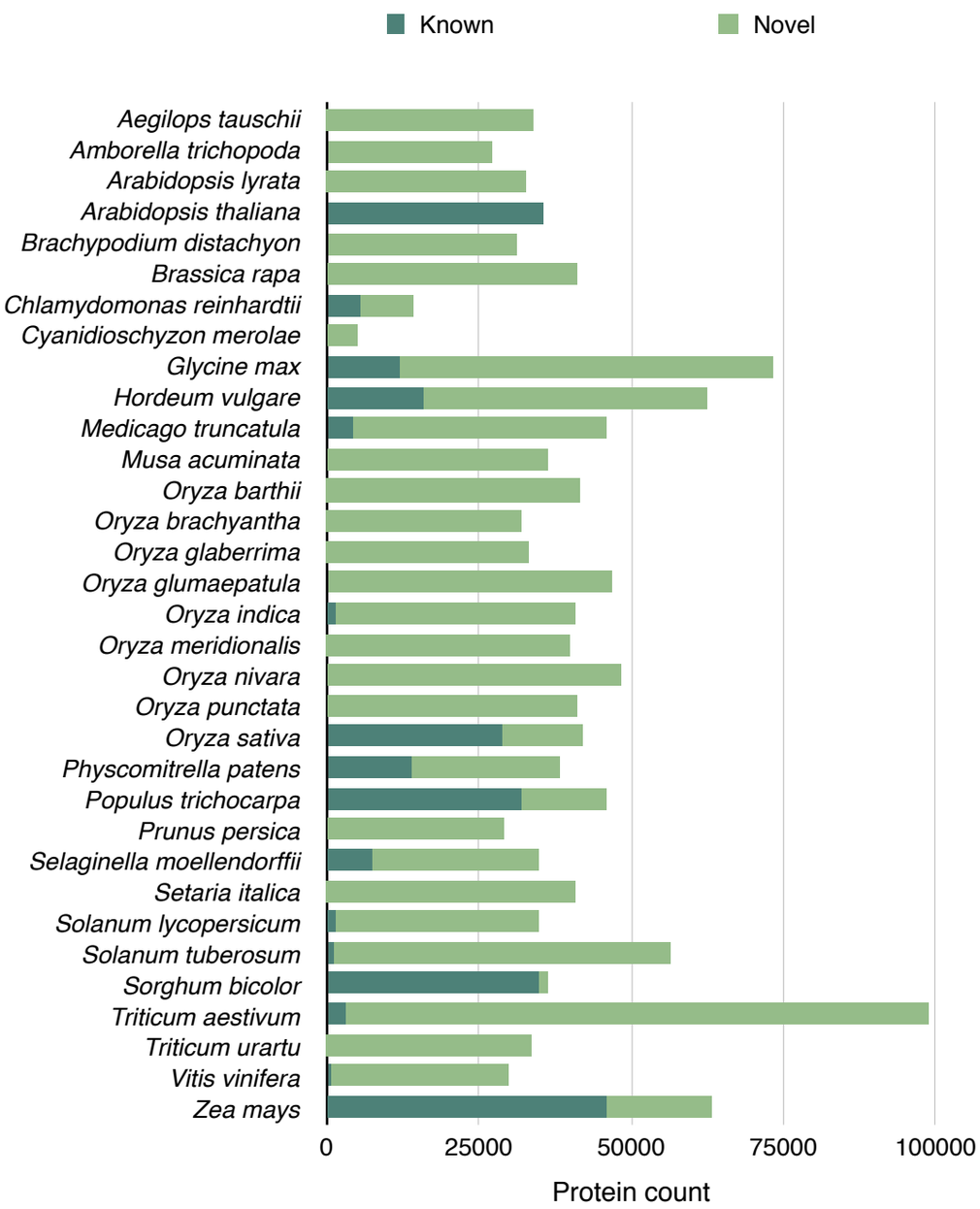
http://ricewiki.big.ac.cn

Spcies long name	Species Name	release 22	单/双子叶/其他	Chinese Name
aegilops_tauschii	*Aegilops tauschii*		单子叶	粗山羊草(Aegilops tauschii) 隶属于禾本科(Poaceae) 小麦族(Triticeae) 山羊草属(Aegilops L .2 n = 14,DD) ,是小麦D 基因组的供体
amborella_trichopoda	*Amborella trichopoda*		无油樟目	无油樟
arabidopsis_lyrata	*Arabidopsis lyrata*		双子叶	深山南芥
arabidopsis_thaliana	*Arabidopsis thaliana*		双子叶	拟南芥
brachypodium_distachyon	*Brachypodium distachyon*		单子叶	二穗短柄草
brassica_oleracea	*Brassica oleracea*	NA	双子叶	甘蓝
brassica_rapa	*Brassica rapa*		双子叶	芸薹，油菜
caenorhabditis_elegans	*Caenorhabditis elegans*		动物	新秀丽隐杆线虫
chlamydomonas_reinhardtii	*Chlamydomonas reinhardtii*		-	莱茵衣藻
ciona_savignyi	*Ciona savignyi*		-	太平洋透明海鞘
cyanidioschyzon_merolae	*Cyanidioschyzon merolae*		-	红藻
drosophila_melanogaster	*Drosophila melanogaster*		-	黑腹果蝇
glycine_max	*Glycine max*		双子叶	大豆
homo_sapiens	*Homo sapiens*		动物	人
hordeum_vulgare	*Hordeum vulgare*		单子叶	大麦
leersia_perrieri	*Leersia perrieri*	NA	-	Familia: Poaceae Subfamilia: Ehrhartoideae Tribus: Oryzeae Subtribus: Oryzinae Genus: Leersia Species: Leersia perrieri
medicago_truncatula	*Medicago truncatula*		双子叶	蒺藜苜蓿:是国际上广泛用于研究仅属于豆科植物或者与之有关的某些生物过程的豆科模式植 物，这些生物过程无法采用模式植物拟南芥进行研究。
musa_acuminata	*Musa acuminata*		单子叶	小果野蕉
oryza_barthii	***Oryza barthii***		单子叶	**短舌野生稻 Oryza barthii, also called wild rice, or African wild rice, is a grass in the rice genus Oryza.**
oryza_brachyantha	***Oryza brachyantha***		单子叶	**短花药野生稻**
oryza_glaberrima	***Oryza glaberrima***		单子叶	**非洲栽培稻**
oryza_glumaepatula	***Oryza glumaepatula***		单子叶	**展颖野生稻**
oryza_indica	***Oryza indica***		单子叶	**亚洲栽培稻 *indica* 种**
oryza_meridionalis	***Oryza meridionalis***		单子叶	**南方野生稻**
oryza_nivara	***Oryza nivara***		单子叶	**尼瓦拉野生稻 Oryza nivara is a wild progenitor of the cultivated rice Oryza sativa.**
oryza_punctata	***Oryza punctata***		单子叶	**斑点野生稻**
oryza_rufipogon	***Oryza rufipogon***	**NA**	单子叶	**普通野生稻 Oryza rufipogon, known as brownbeard rice, wild rice and red rice, is a member of the genus Oryza.**
oryza_sativa	***Oryza sativa***		单子叶	**亚洲栽培稻 *japonica* 种**
ostreococcus_lucimarinus	*Ostreococcus lucimarinus*		-	藻类
physcomitrella_patens	*Physcomitrella patens*		-	小立碗藓
populus_trichocarpa	*Populus trichocarpa*		双子叶	欧洲大叶杨
prunus_persica	*Prunus persica*		双子叶	桃
saccharomyces_cerevisiae	*Saccharomyces cerevisiae*		-	酿酒酵母
selaginella_moellendorffii	*Selaginella moellendorffii*		-	江南卷柏
setaria_italica	*Setaria italica*		单子叶	小米
solanum_lycopersicum	*Solanum lycopersicum*		双子叶	番茄
solanum_tuberosum	*Solanum tuberosum*		双子叶	马铃薯
sorghum_bicolor	*Sorghum bicolor*		单子叶	高粱
theobroma_cacao	*Theobroma cacao*	NA	双子叶	可可树
triticum_aestivum	*Triticum aestivum*		单子叶	普通小麦
triticum_urartu	*Triticum urartu*		单子叶	一粒小麦（小麦属中最原始的二倍体栽培种）
vitis_vinifera	*Vitis vinifera*		双子叶	酿酒葡萄
zea_mays	*Zea mays*		单子叶	玉米

All protein count

Species	Known	Novel
Aegilops tauschii	47	33802
Amborella trichopoda	0	27313
Arabidopsis lyrata	72	32595
Arabidopsis thaliana	35386	0
Brachypodium distachyon	155	30874
Brassica rapa	0	41025
Chlamydomonas reinhardtii	5559	8930
Cyanidioschyzon merolae	107	4891
Glycine max	11722	61597
Hordeum vulgare	15769	46467
Medicago truncatula	4387	41629
Musa acuminata	0	36519
Oryza barthii	5	41667
Oryza brachyantha	1	32036
Oryza glaberrima	8	33156
Oryza glumaepatula	0	46893
Oryza indica	1480	39265
Oryza meridionalis	11	39921
Oryza nivara	95	48266
Oryza punctata	0	41060
Oryza sativa	28831	13301
Physcomitrella patens	13873	24481
Populus trichocarpa	31991	13787
Prunus persica	0	28927
Selaginella moellendorffii	7342	27483
Setaria italica	23	40576
Solanum lycopersicum	1506	33169
Solanum tuberosum	1071	55139
Sorghum bicolor	34820	1518
Triticum aestivum	2986	95793
Triticum urartu	5	33419
Vitis vinifera	669	29258
Zea mays	46146	17095

Important Plants



Important Plants

Select species protein sta

Species	Known*	Novel*	All
Arabidopsis thaliana	35386	0	35386
Brassica rapa	0	41025	41025
Glycine max	11722	61597	73319
Oryza barthii	5	41667	41672
Oryza brachyantha	1	32036	32037
Oryza glaberrima	8	33156	33164
Oryza glumaepatula	0	46893	46893
Oryza meridionalis	11	39921	39932
Oryza nivara	95	48266	48361
Oryza punctata	0	41060	41060
Oryza sativa	28831	13301	42132
Populus trichocarpa	31991	13787	45778
Solanum lycopersicum	1506	33169	34675
Sorghum bicolor	34820	1518	36338
Zea mays	46146	17095	63241

* Defined by v22

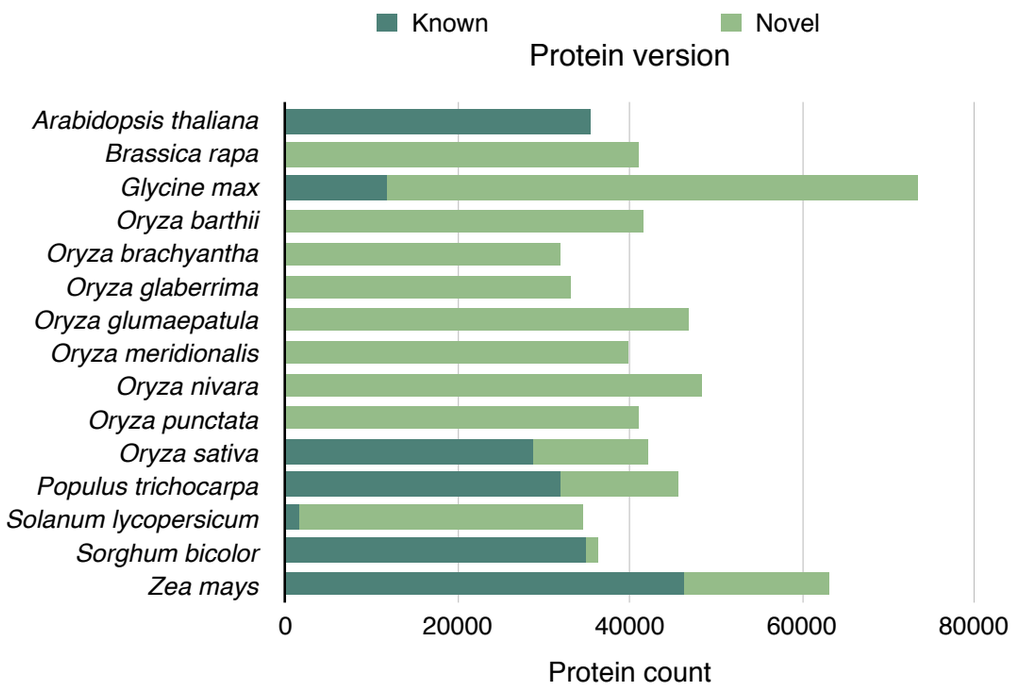
Select species homolog count

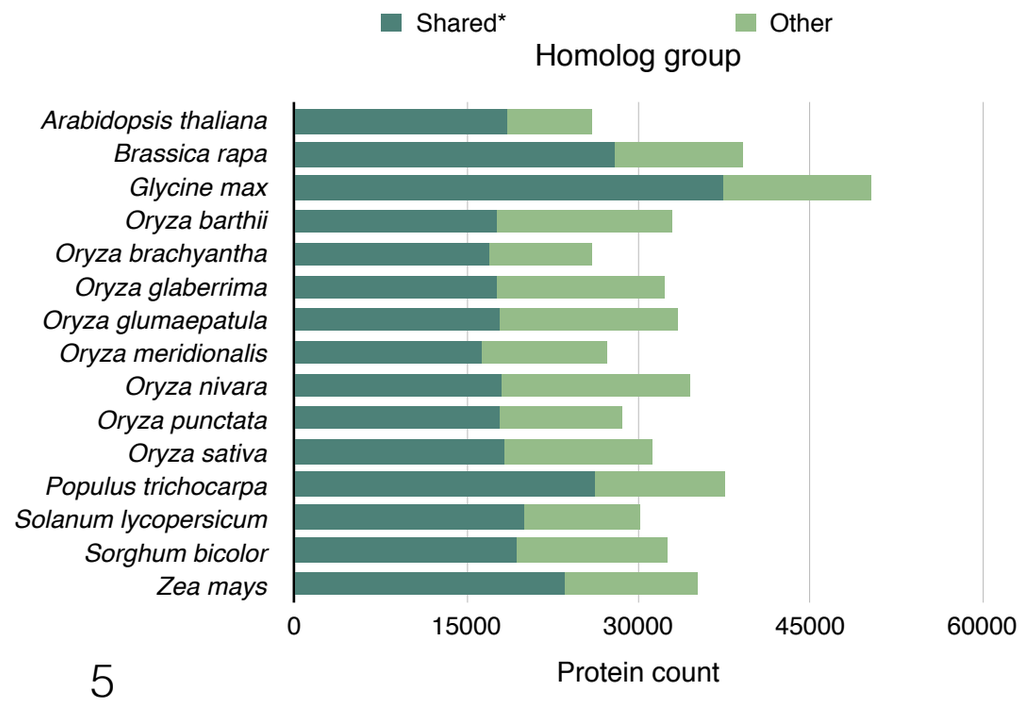
Species	Shared**	Other	Protein count
Arabidopsis thaliana	18453	7565	26018
Brassica rapa	28039	11131	39170
Glycine max	37380	12872	50252
Oryza barthii	17621	15377	32998
Oryza brachyantha	17020	8897	25917
Oryza glaberrima	17731	14568	32299
Oryza glumaepatula	17949	15493	33442
Oryza meridionalis	16248	11140	27388
Oryza nivara	17990	16451	34441
Oryza punctata	17893	10666	28559
Oryza sativa	18255	13034	31289
Populus trichocarpa	26237	11391	37628
Solanum lycopersicum	19973	10249	30222
Sorghum bicolor	19490	13127	32617
Zea mays	23468	11716	35184

** 5382 homolog groups with proteins in all 15 species.

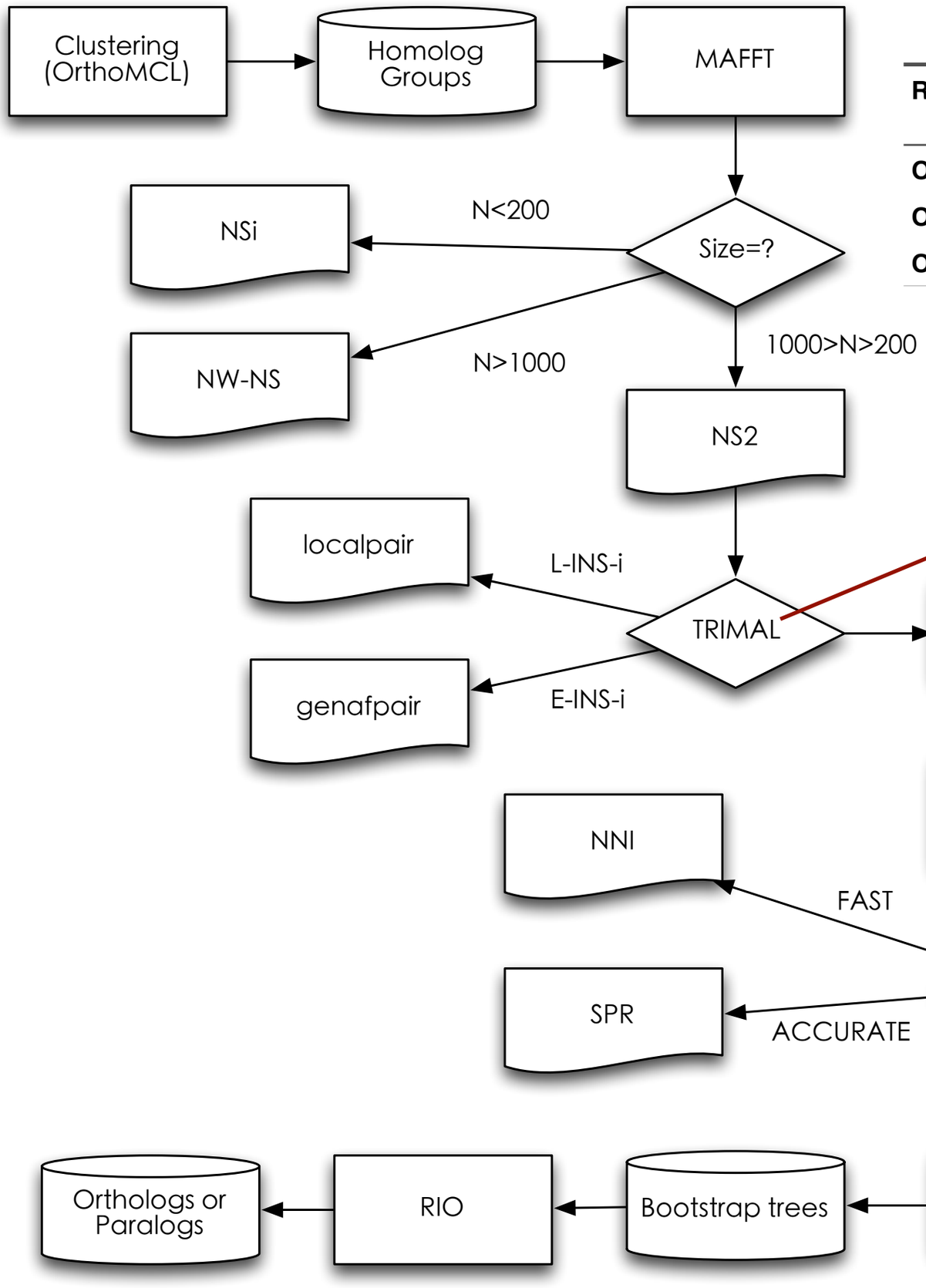
Chromosomes/Scaffolds Count (Selected Species)

Species	Count
Arabidopsis_thaliana	7
Brassica_rapa	284
Glycine_max	148
Oryza_barthii	12
Oryza_brachyantha	210
Oryza_glaberrima	904
Oryza_glumaepatula	12
Oryza_meridionalis	12
Oryza_nivara	12
Oryza_punctata	12
Oryza_sativa	12
Populus_trichocarpa	976
Solanum_lycopersicum	13
Sorghum_bicolor	217
Zea_mays	132





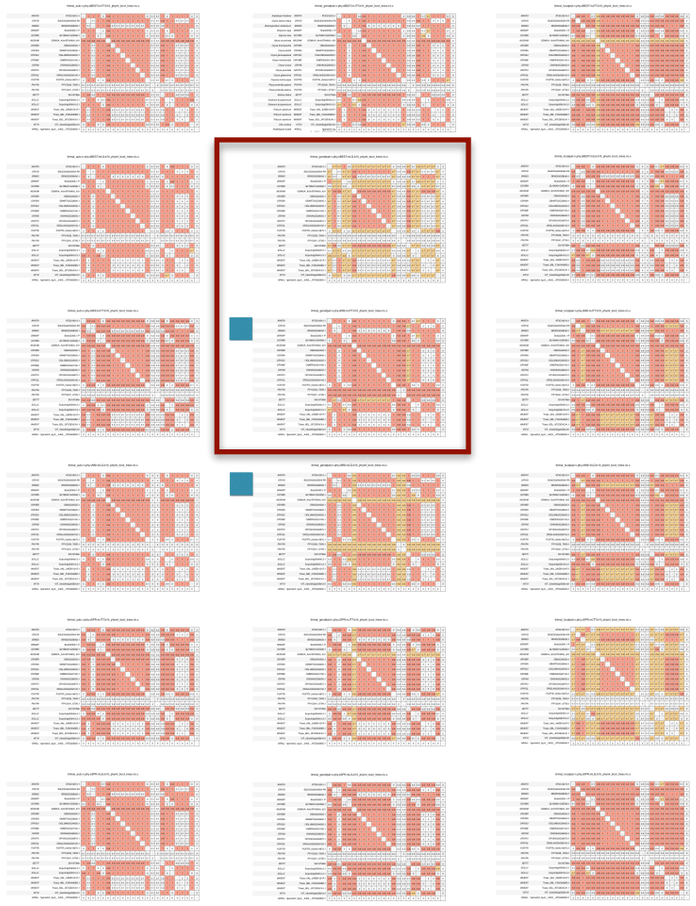
Parameters in Ortholog inference



Test Dataset

Rice Gene ID	CDS count	Species Count	Max Member CNT	MM Species
OS02G0762800	153	33	13	WHEAT
OS05G0113900	236	29	67	WHEAT
OS09G0133200	266	34	22	ORYSI

Parameters in Ortholog inference



itrimal_genafpair.n.phy.sBEST.mLG.b10_phyml_boot_trees.rio.s

Arabidopsi thaliana	-	0.7	0.7	1	0.2	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0	0.4	0.4	0.7	0.1	0.6	0.7	0.7	0.7	0
Oryza sativa Indica	0.7	-	0.1	0.7	0.7	0.9	1	1	1	1	1	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Brachypodium distachyon	0.7	0.1	-	0.7	0.7	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.4	0.4	0.2	0.7	0.7	1	1	1	0.7
Brassica rapa	1	0.7	0.7	-	0.2	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0	0.4	0.4	0.7	0.1	0.6	0.7	0.7	0.7	0
Glycine max	0.2	0.7	0.7	0.2	-	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.1	0.4	0.4	0.7	0.5	0	0.7	0.7	0.7	0.1
Musa acuminata	0.7	0.9	0.9	0.7	0.7	-	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.7	0.4	0.4	0.9	0.7	0.7	0.9	0.9	0.9	0.7
Oryza brachyantha	0.7	1	0.1	0.7	0.7	0.9	-	1	1	1	1	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza barthii	0.7	1	0.1	0.7	0.7	0.9	1	-	1	1	1	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza glumaepatula	0.7	1	0.1	0.7	0.7	0.9	1	1	-	1	1	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza meridionalis	0.7	1	0.1	0.7	0.7	0.9	1	1	1	-	1	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza nivara	0.7	1	0.1	0.7	0.7	0.9	1	1	1	1	-	1	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza punctata	0.7	1	0.1	0.7	0.7	0.9	1	1	1	1	1	-	1	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Oryza glaberrima	0.7	1	0.1	0.7	0.7	0.9	1	1	1	1	1	1	-	0.7	0.4	0.4	0.9	0.7	0.7	0.1	0.1	0.1	0.7
Populus trichocarpa	0	0.7	0.7	0	0.1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	-	0.4	0.4	0.7	0.5	0	0.7	0.7	0.7	0.8
Physcomitrella patens	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	-	0	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Physcomitrella patens	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0	-	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Setaria italica	0.7	0.9	0.2	0.7	0.7	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.7	0.4	0.4	-	0.7	0.7	0.2	0.2	0.2	0.7
Solanum lycopersicum	0.1	0.7	0.7	0.1	0.5	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.5	0.4	0.4	0.7	-	0	0.7	0.7	0.7	0.4
Solanum lycopersicum	0.6	0.7	0.7	0.6	0	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0	0.4	0.4	0.7	0	-	0.7	0.7	0.7	0
Triticum aestivum	0.7	0.1	1	0.7	0.7	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.4	0.4	0.2	0.7	0.7	-	0	0	0.7
Triticum aestivum	0.7	0.1	1	0.7	0.7	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.4	0.4	0.2	0.7	0.7	0	-	0	0.7
Triticum aestivum	0.7	0.1	1	0.7	0.7	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.4	0.4	0.2	0.7	0.7	0	0	-	0.7
Vitis vinifera	0	0.7	0.7	0	0.1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.4	0.4	0.7	0.4	0	0.7	0.7	0.7	-

- Sub model: JTT
  - ProtTest
- Tree Searching: SPR
- Trial: genefpair

itrimal_genafpair.n.phy.sNNI.mJTT.b10_phyml_boot_trees.rio.s

AT3G19210.1	-	1	1	1	0	0.8	1	1	1	1	1	1	1	0	0.8	0.8	1	0.1	0.7	1	1	1	0
BGIOSGA005564-PA	1	-	0	1	1	0.8	1	1	1	1	1	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
BRADI3G58092.1	1	0	-	1	1	0.8	0	0	0	0	0	0	0	1	0.8	0.8	0.3	1	1	1	1	1	1
Bra022395.1-P	1	1	1	-	0	0.8	1	1	1	1	1	1	1	0	0.8	0.8	1	0.1	0.7	1	1	1	0
GLYMA01G45596.1	0	1	1	0	-	0.8	1	1	1	1	1	1	1	0.2	0.8	0.8	1	0.4	0	1	1	1	0.2
GSMUA_Achr5P24600_001	0.8	0.8	0.8	0.8	0.8	-	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
OB02G40520.1	1	1	0	1	1	0.8	-	1	1	1	1	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
OBART02G33600.1	1	1	0	1	1	0.8	1	-	1	1	1	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
OGLUM02G34050.1	1	1	0	1	1	0.8	1	1	-	1	1	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
OMERI02G31160.1	1	1	0	1	1	0.8	1	1	1	-	1	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
ONIVA02G36050.1	1	1	0	1	1	0.8	1	1	1	1	-	1	1	1	0.8	0.8	0.7	1	1	0	0	0	1
OPUNC02G30870.1	1	1	0	1	1	0.8	1	1	1	1	1	-	1	1	0.8	0.8	0.7	1	1	0	0	0	1
ORGLA02G0290100.1	1	1	0	1	1	0.8	1	1	1	1	1	1	-	1	0.8	0.8	0.7	1	1	0	0	0	1
POPTR_0004s14870.1	0	1	1	0	0.2	0.8	1	1	1	1	1	1	1	-	0.8	0.8	1	0.4	0	1	1	1	0.2
PP1S236_78V6.1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	-	0	0.8	0.8	0.8	0.8	0.8	0.8	0.8
PP1S341_67V6.1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0	-	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Si019738m	1	0.7	0.3	1	1	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	1	0.8	0.8	-	1	1	0.3	0.3	0.3	1
Solyc04g056400.2.1	0.1	1	1	0.1	0.4	0.8	1	1	1	1	1	1	1	0.4	0.8	0.8	1	-	0	1	1	1	0.2
Solyc04g056410.2.1	0.7	1	1	0.7	0	0.8	1	1	1	1	1	1	1	0	0.8	0.8	1	0	-	1	1	1	0
Traes_6AL_2AEB1031F.1	1	0	1	1	1	0.8	0	0	0	0	0	0	0	1	0.8	0.8	0.3	1	1	-	0	0	1
Traes_6BL_F3A5A99E6.1	1	0	1	1	1	0.8	0	0	0	0	0	0	0	1	0.8	0.8	0.3	1	1	0	-	0	1
Traes_6DL_3FCDE4CA1.1	1	0	1	1	1	0.8	0	0	0	0	0	0	0	1	0.8	0.8	0.3	1	1	0	0	-	1
VIT_03s0063g00580.t01	0	1	1	0	0.2	0.8	1	1	1	1	1	1	1	0.2	0.8	0.8	1	0.2	0	1	1	1	-

- More test: release-24
  - 1840 homolog group with unique gene in each species

Parameters in Ortholog inference

- MAFFT: NSi (n<200), NS2 (1000>n>200)
- TRIMAL: genefpair
- PhyML: JTT
- Tree searching: SPR
- RIO bootstrap cutoff
- Test data: gramene.org (based on ensembl release-24)
- Update data to ensembl release-24

Website Design

- Module
    - Bar chart: homolog gene/ortholog in species
    - Phylogenetic tree
    - Table of ortholog
    - Alignment of protein, CDSs
    - Ka/Ks, phylogenetic bootstrap
    - Sequence download
- Linkout/in
    - Sequence database of each species
    - PubMed id, uniprot id of each sequence for outsite link
    - Link from other database/moudle

Thanks for your attention