

利用生物信息技术解析水稻颗粒结合 淀粉合成酶1

Analysis of rice Granule-bound starch
synthase 1 by bioinformatics

组 号：3G01

组 员：安瑞虎 陈昌招 郭锐

报告人：安瑞虎

组 长：安瑞虎

汇报内容

- 研究背景
- 蛋白结构预测与分析
- 同源序列及突变情况分析
- 总结

研究背景

- 颗粒结合型淀粉合成酶Granule-bound starch synthase 1 (GBSS1)，由是禾谷类作物Waxy基因编码的，其作用是特异性的延长直链淀粉，而且Waxy是典型W_x突变体，由于其发生突变，GBSS1不能正常表达或酶活性降低，致使缺少直链淀粉，这就是糯性性状形成的根本原因。直链淀粉含量不同会导致稻米的硬度的改变，直链淀粉含量越高，稻米硬度越大，因此，由WAXY编码的GBSSI对稻米食味品质影响较大。
- 登录号：Q0DEV5
- 属于glycosyltransferase 1 family，糖基转移酶1家族
- 其功能是胚乳中直链淀粉合成必需酶，参与淀粉生物合成和聚糖生物合成

二、蛋白结构预测与分析

2.1、氨基酸组成分析

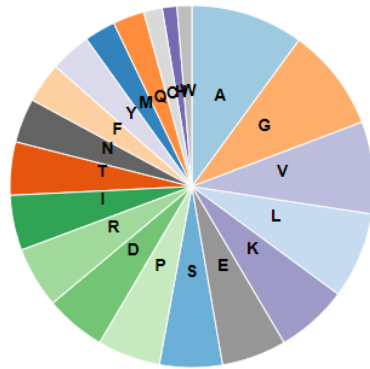
- 利用predict protein对GBSS1进行预测，分析其氨基酸组成
- 经图示可以看出,GBSS1总氨基酸数为609，丙氨酸含量最高为10.02%，甘氨酸次之为9.20%，缬氨酸为8.21%，亮氨酸7.72%。

Summary

Sequence Length

609

Amino Acid
composition

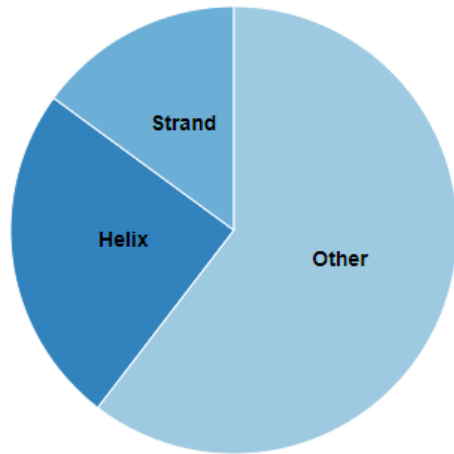


二、蛋白结构预测与分析

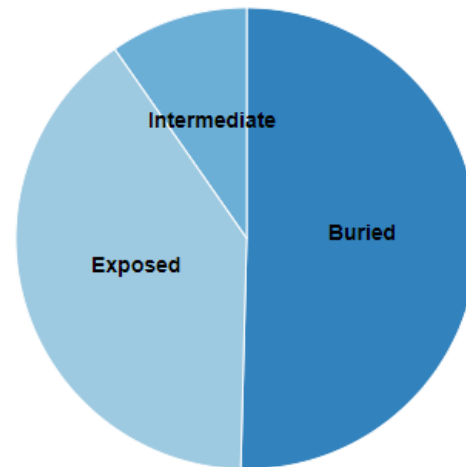
2.2 二级结构与可溶性分析

- 如图所示二级结构组成为24.63%的螺旋，14.94%的折叠，60.43%的无规则卷曲。通过蛋白质残基可溶性分析可以看出在内部的氨基酸残基占50.41%，在蛋白质表面的残基占39.90%，处于中间状态的为9.69%，通过<https://www.novopro.cn/>分析，预测其溶解度为0.283，证明该蛋白疏水性较强。

Secondary Structure Composition



Solvent Accessibility



预测的溶解度 (归一化后) : 0.283

二、蛋白结构预测与分析

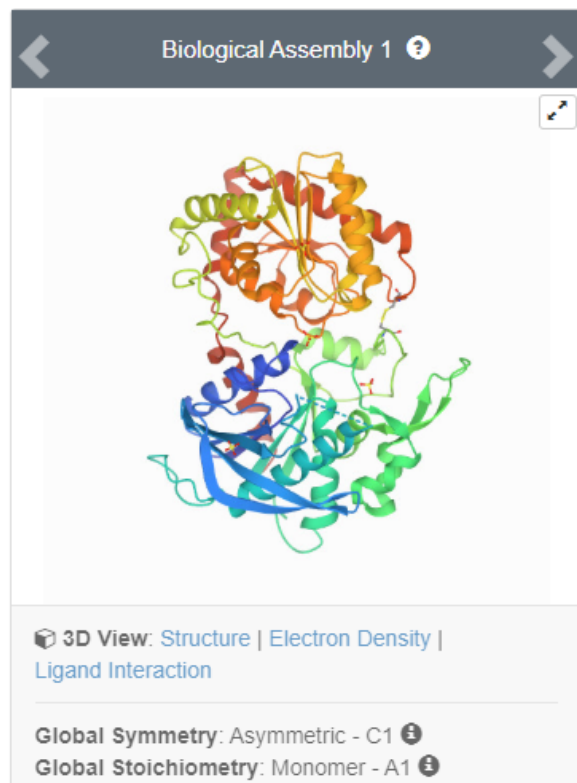
2.3亚细胞定位

在Uniprot中可以看到， GBSS1只存在于叶绿体中， 叶绿体是高等植物进行光合作合成淀粉等有机物的场所， 与其功能相符合



二、蛋白结构预测与分析

2.4三级结构



3VUE

Crystal Structure of Rice Granule bound Starch Synthase I Catalytic Domain

DOI: [10.2210/pdb3VUE/pdb](https://doi.org/10.2210/pdb3VUE/pdb)

Classification: **TRANSFERASE**

Organism(s): *Oryza sativa Japonica Group*

Expression System: *Escherichia coli*

Mutation(s): No

Deposited: 2012-06-28 Released: 2012-09-05

Deposition Author(s): Momma, M., Fujimoto, Z.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.70 Å

R-Value Free: 0.267

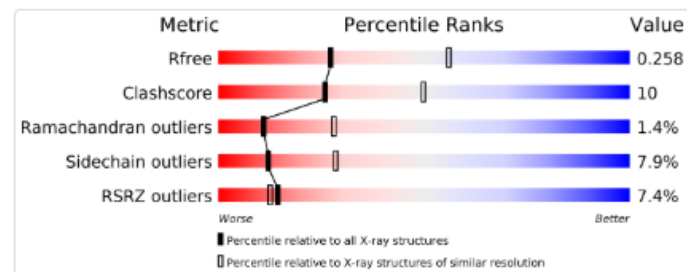
R-Value Work: 0.218

R-Value Observed: 0.221

wwPDB Validation

3D Report

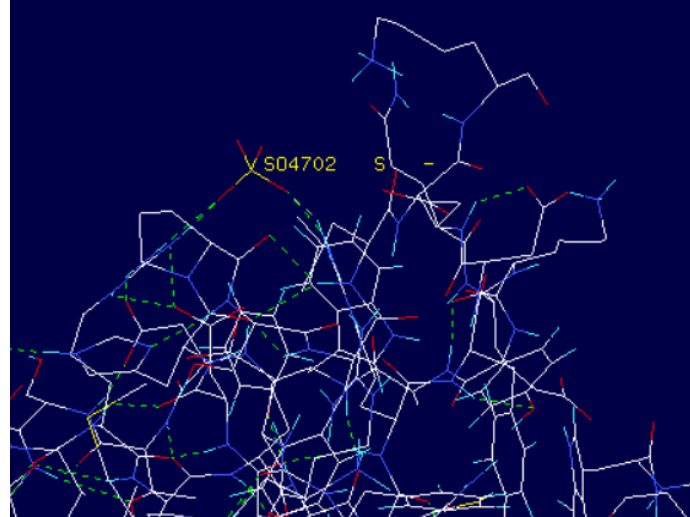
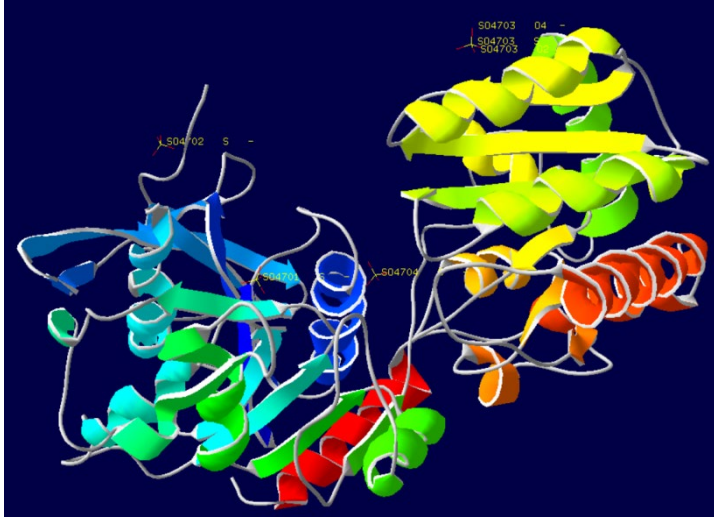
Full Report



二、蛋白结构预测与分析

2.5三级结构分析

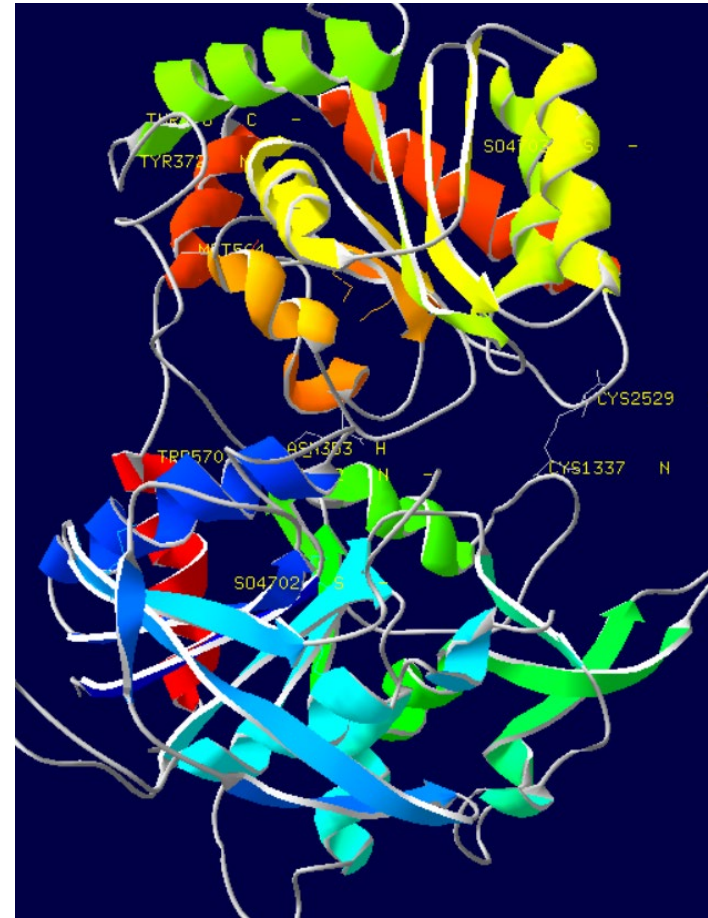
- GBSS1三级结构中 α 螺旋、 β 折叠和 β 转角都有，含有4个 SO_4 ，这四个 SO_4 与蛋白质以氢键相连



二、蛋白结构预测与分析

2.6三级结构分析

CYS 337和CYS 529形成一个二硫键，水稻GBSS1中的第一个半胱氨酸Cys 337在禾本科植物中是保守的，但在非禾本科植物中被缬氨酸取代。因此，尽管两个结构域在植物GBSSIs中似乎是保守的，但二硫键只存在于禾本科GBSSIs中。由该二硫键引起的开放和封闭结构之间的结构域移动，可能有助于淀粉的有效生物合成。

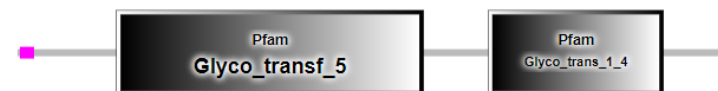


结构域预测

Pfam



SMART



Source	Domain	Start	End
low_complexity	n/a	2	14
disorder	n/a	29	36
disorder	n/a	38	65
Pfam	Glyco_transf_5	84	345
low_complexity	n/a	99	108
Pfam	Glycos_transf_1	387	548

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
low_complexity	2	14	N/A
Pfam:Glyco_transf_5	84	345	3.7e-74
Pfam:Glycos_transf_1	387	528	7.3e-18
Pfam:Glyco_trans_1_4	400	549	1e-15

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Outlier homologues and homologues of known structure: ?

Name	Sequence	Start ▲	End	E-value
PDB:3VUFJA	3vuf	83	609	0
SCOP:d1f6da_	d1f6da_	83	584	1e-46

这两个结构域所在位点与之前预测的跨膜结构大致符合；

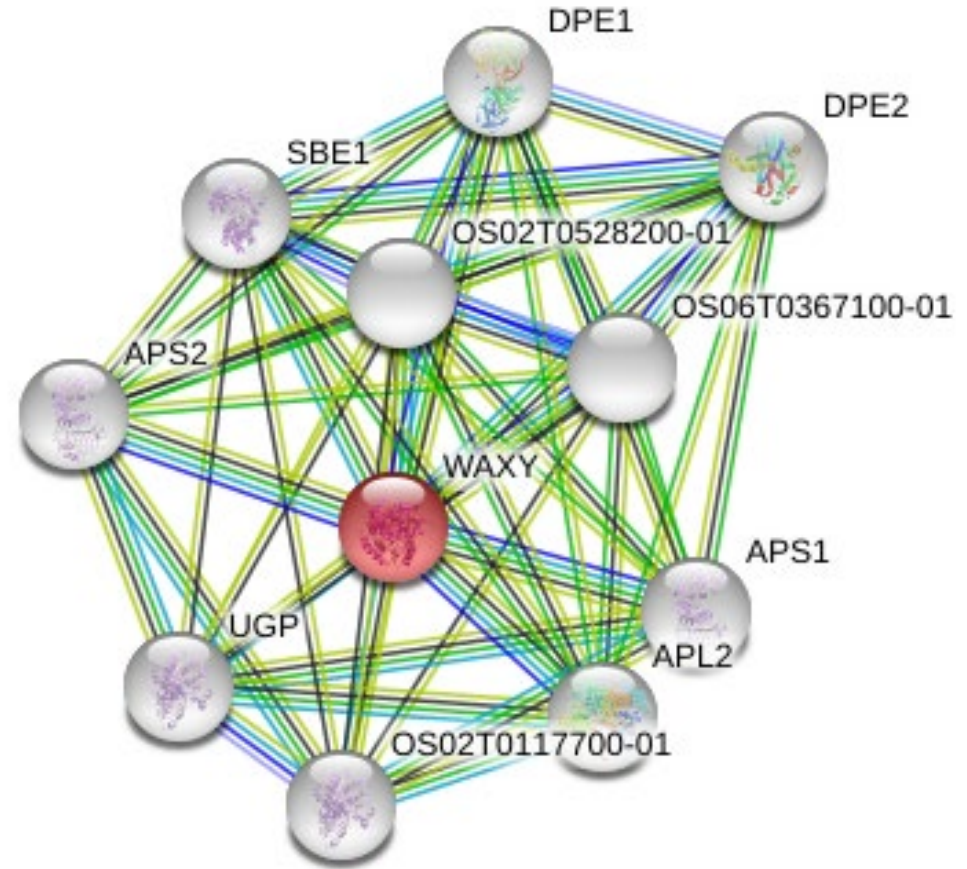
Glyco_transf属于GT-B家族，即糖基转移酶家族；

Glyco_transf_5是淀粉合酶催化结构域；

Glyco_transf_1是糖基转移酶I；

GBSS1蛋白互作情况

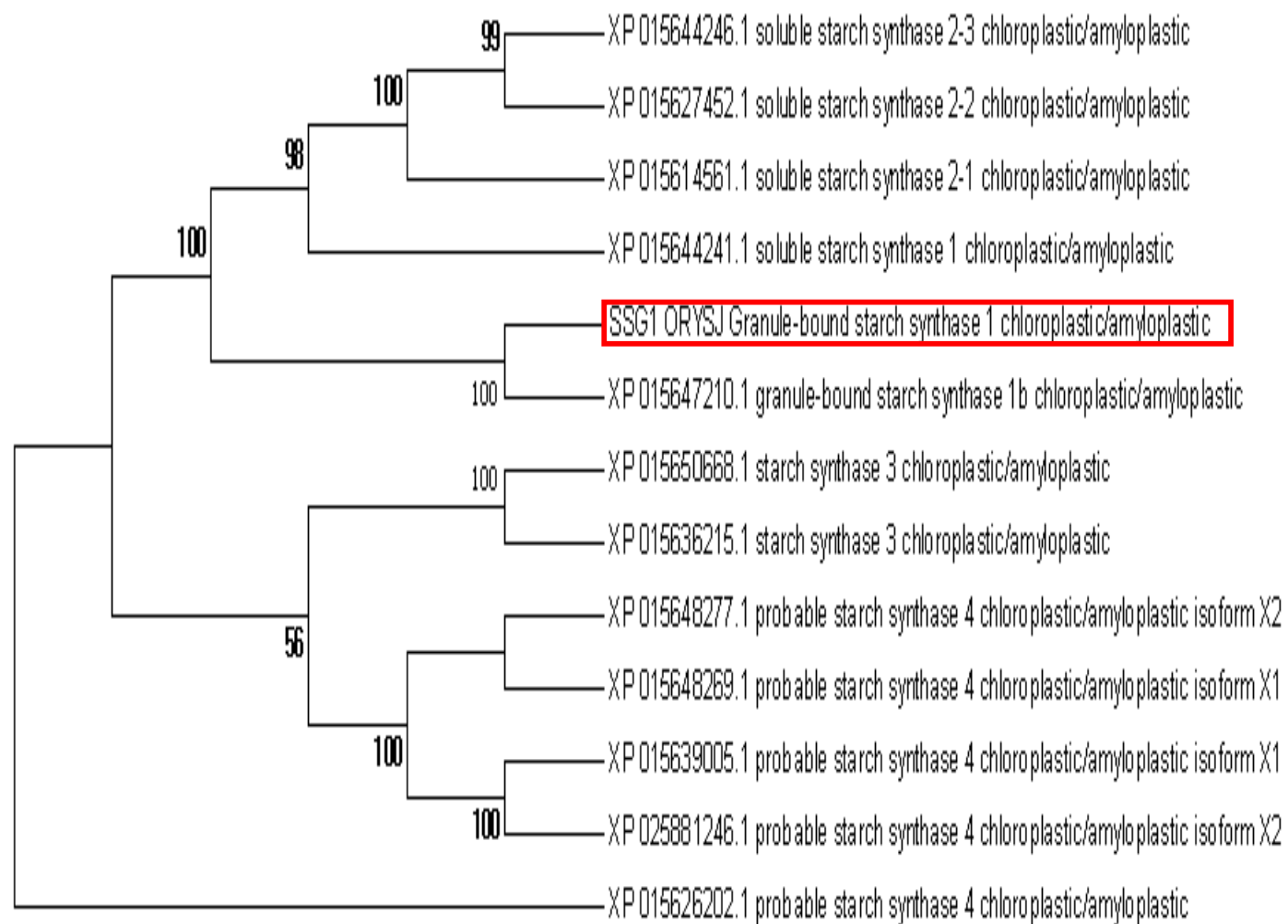
与GBSS1互作的蛋白大多都是与葡萄糖合成有关的酶



系统发育树分析

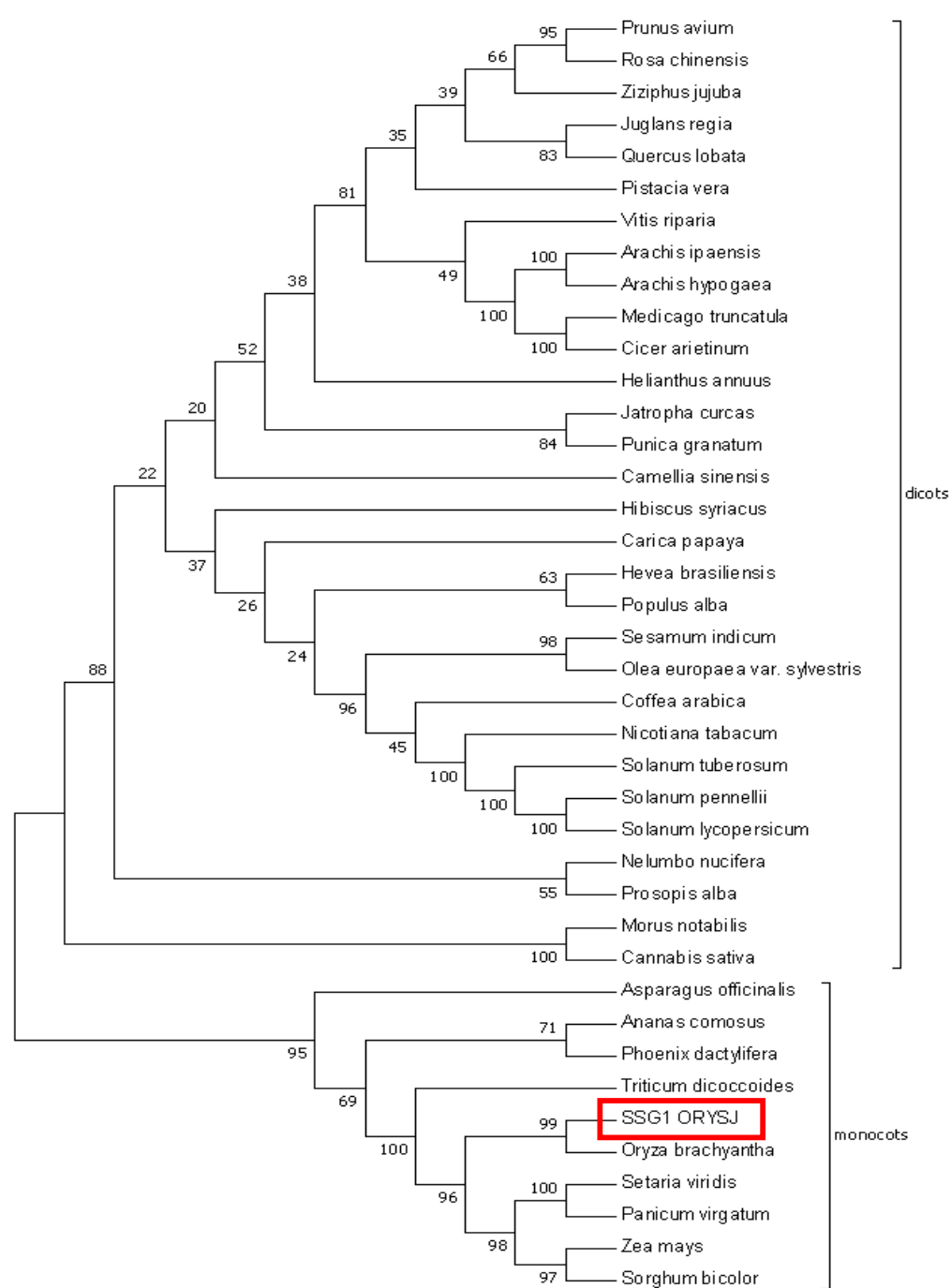
发现旁系同源蛋白中

SSG1 ORYSJ与XP015647210.1
(颗粒结合型淀粉合成酶1 b) 相
似性最高, 同源性最近。



系统发育树分析

通过分析GBSS1的直系同源蛋白发现单子叶与双子叶植物之间存在差异，这说明颗粒结合型淀粉合成酶在高等植物进化过程中相对保守



保守性分析

将GBSS1与其他几种单子叶植物的同源蛋白进行多序列对比发现同源蛋白的结构域保守性很高。可以看出突变位点多为不完全保守位点。

SSG1_ORYSJ	MSALTTSQLATSSAGFGIADRSAPSSMLRHGFGQLKFRSPAG-GDATSLVTT SARATPK	59
Oryza	MSALTTSQLATSSAGFGIADRSAPSSMLRHGFGQLKFRSPAG-ADG-SLSVTT SARATPK	58
Setaria	MAALATSQLATT RAGFGLGDAS-SSMP RFGVQGLSRGRASS-PAATLSVTT SARAAPR	57
Panicum	MAALATSQLATT HAGFGLGGD-TSMP RRGVQGLRGP R-F SAGALSVRT SARAAPR	53
Zea	MAALATSQLVAT RAGLVVPDAS-T-FRRGAAQGLRGGRTAS-AADT L SMRTSARAAPR	55
Sorghum	MSTLATSQLVATHAGLVVPDAS-M-FRRGGVQGLRAAARASAAAGDALSMRTSACPAPR	57
SSG1_ORYSJ	QQRS-VQRGSRFPFVVVYATGAGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	116
Oryza	QQRS-VQRGSRFPFVVVYATGAGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	115
Setaria	QQHRRAR-Q-ARFPLVVCATGAGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	114
Panicum	QQSRRPQR-GGGRLPLVVC-A-AGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	110
Zea	LHQQQQQARRGARFPFLVVCAS-AGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	114
Sorghum	QQFA-ARRGRRGRRFPFLVVCAT-AGMNVVFGAEMAPWSKTGGLGDLVGGLPFAMAANG	115
SSG1_ORYSJ	HRVMVISPRYDQYKDAWDTSVVAEIKVADRVERVFFHCYKRGVDRVFDHPSFLEKVVWG	176
Oryza	HRVMVISPRYDQYKDAWDTSVVAEIKVADRVERVFFHCYKRGVDRVFDHPSFLEKVVWG	175
Setaria	HRVMVISPRYDQYKDAWDTSVVEIKVADRVERVFFHCYKRGVDRVFDHPSFLEKVVWG	174
Panicum	HRVMVISPRYDQYKDAWDTSVVEIKVADRVERVFFHCYKRGVDRVFDHPSFLEKVVWG	170
Zea	HRVMVVSRYDQYKDAWDTSVVEIKMGDRYETVFFHCYKRGVDRVFDHPLFLERVVWG	174
Sorghum	HRVMVVSRYDQYKDAWDTSVVEIKMGDGYETVFFHCYKRGVDRVFDHPLFLERVVWG	175
SSG1_ORYSJ	KTGEKIYGFDTGVDYKDNQMRFSLLCQAALAPRLLNLNNTFYFKGYGEDVVVFCNDWH	236
Oryza	KTGEKIYGFDTGIDYKDNQLRFSLLCQAALAPRLLNLNNTFYFSGTYGEDVVVFCNDWH	235
Setaria	KTGEKIYGF DAGVDYKDNQLRFSLLCQAALAPRLLSLNNTFYFSGPYGEDVVVFCNDWH	234
Panicum	KTGEKIYGF DAGVDYKDNQRFRFSLLCQAALAPRLLSLNNTFYFSGPYGEDVVVFCNDWH	230
Zea	KTEEKIYGF DAGTDYKDNQLRFSLLCQAALAPRLLSLNNTFYFSGPYGEDVVVFCNDWH	234
Sorghum	KTEEKIYGF DAGTDYKDNQLRFSLLCQAALAPRLLSLNNTFYFSGPYGEDVVVFCNDWH	235
SSG1_ORYSJ	TGPLASYLNKNYQPNGIYRNKAVAFCIHNSVQGRFAFEDFELNLSERFRSSDFIDY	296
Oryza	TGPLPSYLNKNYQPNGIYRNKAVAFCIHNSVQGRFAFEDFELNLSERFRSSDFIDY	295
Setaria	TGPLSSYLKSNVQSNGIYRNKATAFCIHNSVQGRFAFSDYFELNLSERFRSSDFIDY	294
Panicum	TGPLSSYLKSNVQSNGIYRNKATAFCIHNSVQGRFAFSDYFELNLSERFRSSDFIDY	290
Zea	TGPLSCYLKSNVQSHGIYRDATAFCIHNSVQGRFAFSDYFELNLSERFRSSDFIDY	294
Sorghum	TGPLSCYLKSNVQSNGIYKDATAFCIHNSVQGRFAFSDYFELNLSERFRSSDFIDY	295
SSG1_ORYSJ	DTFVEGRKLNWMKAGILEADRVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	356
Oryza	DKFVEGRKLNWMKAGILESDRVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	355
Setaria	EKFVEGRKLNWMKAGILEADRVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	354
Panicum	EKFVEGRKLNWMKAGILEADKVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	350
Zea	EKFVEGRKLNWMKAGILEADRVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	354
Sorghum	EKFVEGRKLNWMKAGILEADRVLTVSPYYAEELISGIARGCELDNDMRLTGITGI VNGMD	355
SSG1_ORYSJ	VSEWDPSKDKYITAKYDATTAEAKALNHEALQAEAGLPVDRKIPLIAFIGRLEEQKGF	416
Oryza	VSEWDPSKDKYITVKYDATTAEAKALNHEALQAEVGLPVDRIKIPLIAFIGRLEEQKGF	415
Setaria	VSEWDPSKDKYIATKYDVSTATAAKALNHEALQAAAGLPVDRKIPLVAFVGRLEEQKGF	414
Panicum	VSEWDPSKDKYIATKYDVSTATAAKALNHEALQAAAGLPVDRKIPLVAFVGRLEEQKGF	410
Zea	VSEWDPSKDKYIAVKYDVSTAVEAKALNHEALQAEVGLPVDRIKIPLVAFVGRLEEQKGF	414
Sorghum	VSEWDPSKDKYIAVKYDVSTAVEAKALNHEALQAEVGLPVDRIKIPLVAFVGRLEEQKGF	415
SSG1_ORYSJ	VMAAAIPELME-EDVQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	474
Oryza	VMAAAIPELME-ENVQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	473
Setaria	VMAAAIPELME-EDVQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	472
Panicum	VMAAAIPELME-EDVQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	468
Zea	VMAAAIPELME-EDVQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	474
Sorghum	VMAAAIPELME-EDIQIVLGTGKHKFERMLMSAEEKYPDKVRAVVKFNAALAHHIMAG	473
SSG1_ORYSJ	ADVLAVTSRFEPFCGLIQLQGMRYGTFCACASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	534
Oryza	ADTLAVTSRFEPFCGLIQLQGMRYGTFCACASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	533
Setaria	ADLLAVTSRFEPFCGLIQLQGMRYGTFCVCASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	532
Panicum	ADLLAVTSRFEPFCGLIQLQGMRYGTFCVCASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	528
Zea	ADVLAVTSRFEPFCGLIQLQGMRYGTFCACASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	534
Sorghum	ADLLAVTSRFEPFCGLIQLQGMRYGTFCACASTGGLVDTTIEGKTGFHMGRLSVDCRVVEP	533
SSG1_ORYSJ	SDVQKVAATLKRRAIKVVGTFAIEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	594
Oryza	SDVQKVAATLKRRAIKVVGTFAIDEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	593
Setaria	ADVQKVAATLKRRAIKVVGTFAIEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	592
Panicum	ADVQKVAATLKRRAIKVVGTFAIEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	588
Zea	ADVQKVAATLKRRAIKVVGTFAIEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	594
Sorghum	ADVQKVAATLKRRAIKVVGTFAIEMVNRNCMIQDLSWKGFARWENVLLSLGVAGSFGI	593
SSG1_ORYSJ	GDEIAPLAKENVAAP 609	
Oryza	GEEIAPLAKENVAAP 608	
Setaria	GEEIAPLAKENVAAP 607	
Panicum	GEEIAPLAKENVAAP 603	
Zea	GEEIAPLAKENVAAP 609	
Sorghum	GEEIAPLAKENVAAP 608	

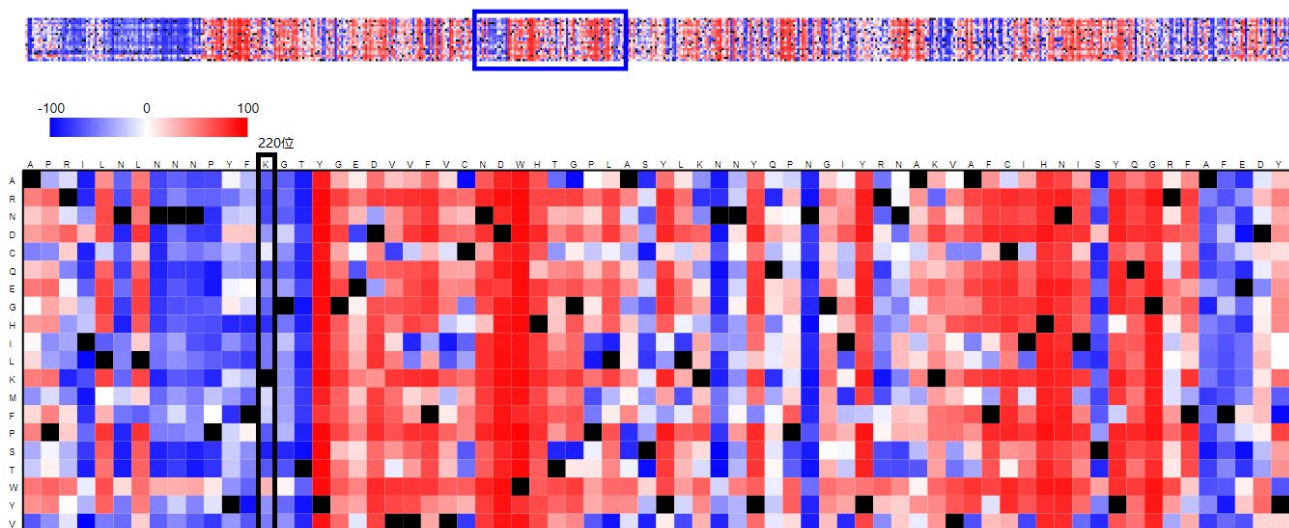
突变位点分析

利用predict protein进行分析

蓝色表示突变后影响效率较低

红色表示突变后影响效率较高

如图所示，第220位的赖氨酸突变后影响效率较低，与保守型分析相印证。



总结

- 通过本次实验，我们初步掌握了对一个基因或蛋白的解析、挖掘的基本方法。
- 生物信息技术为分子实验的研究提供了极大地便利。
- 由于小组内同学课题均未开始，通过这次分析研究，为我们今后的实验研究提供了一种新的思路。

参考文献

- 1、 Mitsuru MOMMAy, Zui FUJIMOTO, Interdomain Disulfide Bridge in the Rice Granule Bound Starch Synthase I Catalytic Domain as Elucidated by X-Ray Structure Analysis, Biosci. Biotechnol. Biochem., 76 (8), 1591–1595, 2012.
- 2、 李祥栋 ,石明,魏心元.禾谷类作物胚乳淀粉合成及 Waxy基因研究进展, 黔西南州农业科学研究所,中国农业学报,2015,31(12):181—186

谢谢