葡萄 *SNAT* 基因的功能预测和原核表达分析

Prediction of grape *VvSNAT* gene function

and analysis of its expression in prokaryotic system

G02组成员： 吴艳迪 牛　娟

张学贤 肖水平

汇　报　人：吴艳迪

指 导 老 师：罗静初

Zhengzhou Fruit Research Institute,CAAS

1. **Topics of group member**

2. **Background**

3. **Gene cloning**

4. **Sequence analysis and function prediction**

5. **Combination with molecular experiment**

G02B 牛娟

题目：八月瓜种子硬度相关基因的挖掘及生物信息学分析

拟开展工作：挖掘组学数据中细胞壁合成与降解相关相关的基因/蛋白，并对相关基因进行生物信息学分析和功能验证。

目的及意义：研究八月瓜种子硬度形成的分子机制，降低种皮硬度或使种皮变软，提高果实的可食率，为八月瓜的分子育种提供理论基础。

涉及到的软件：除本学期学习的软件外，后期可能还会涉及到热图构建，蛋白质功能，群体结构分析等相关软件等。

1. Topics of group members

G02C 张学贤

题目：棉花育性恢复相关基因挖掘与功能验证

棉花虽已实现三系配套，但由于恢复系来源狭窄，恢复力不强等因素，在生产上仍没有真正实现大规模三系杂交种生产。因此，有必要针对棉花恢复基因及其作用机制开展相关研究。

恢复系　　　　不育系

利用恢复基因等基因系材料，通过转录组测序研究其基因差异表达情况，筛选与育性恢复相关的差异表达基因，挖掘与棉花育性发育相关的候选基因并进行验证，探寻其对影响育性恢复的分子机制，为优良恢复系选育提供理论依据。

G02D 肖水平

题目：棉花纤维发育相关基因的克隆、生物信息学分析及功能验证

拟开展工作： 对棉花纤维发育相关基因进行克隆；利用生物信息学方法对克隆到的基因序列等进行分析 ；基因功能验证。

目的及意义： 棉花纤维是重要的天然纺织原料。随着人们对高品质纤维需求的增加和纺织工艺的改进，挖掘控制纤维性状的基因，深入研究棉花纤维发育的分子机制，对于实现棉花纤维品质改良具有重要意义。

G02A 吴艳迪

题目： 葡萄褪黑素合成基因 *SNAT* 原核表达分析

实验内容： 对葡萄中褪黑素合成基因*SNAT*进行初步序列分析及功能预测，设计引物克隆其cds序列，连接至表达载体，转入大肠杆菌BL(DE3)中诱导产生可溶性蛋白，加入相应前体物质从酶学功能方面初步鉴定克隆得到的葡萄*SNAT*基因的功能。

目的及意义： 褪黑素在植物生长发育过程中发挥重要作用，葡萄作为我国重要的栽培果树，鉴定葡萄中褪黑素合成基因的功能有助于为采用分子手段调控葡萄中内源褪黑素含量奠定基础

褪黑素（N-乙酰-5-甲氧基色胺）是一种吲哚胺类物质，它普遍存在于细菌、真菌、藻类、植物、动物、人类等绝大多数生物有机体中。
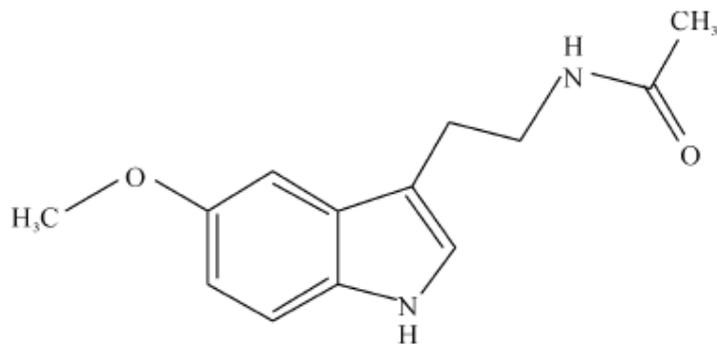

图1. 褪黑素的化学结构

2.1 Function of melatonin in animals

在动物体内，褪黑素具有**调节昼夜节律**和光周期反应、**改善睡眠障碍、清除体内的自由基**、减缓衰老、抗肿瘤、提高机体免疫力等多种功能。
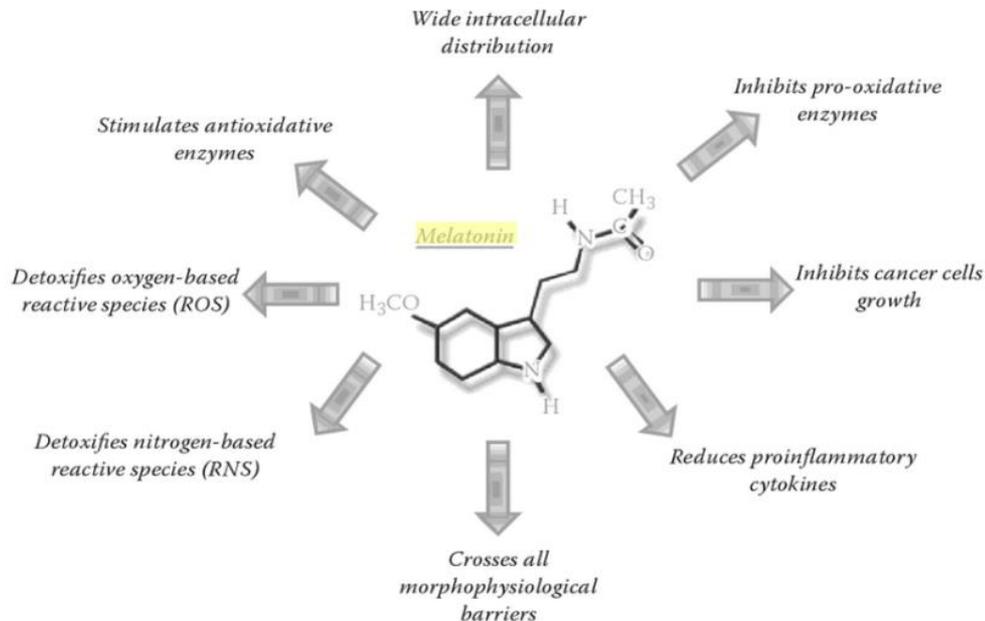
## 2.2 Function of melatonin in plants

生长发育起到调节作用
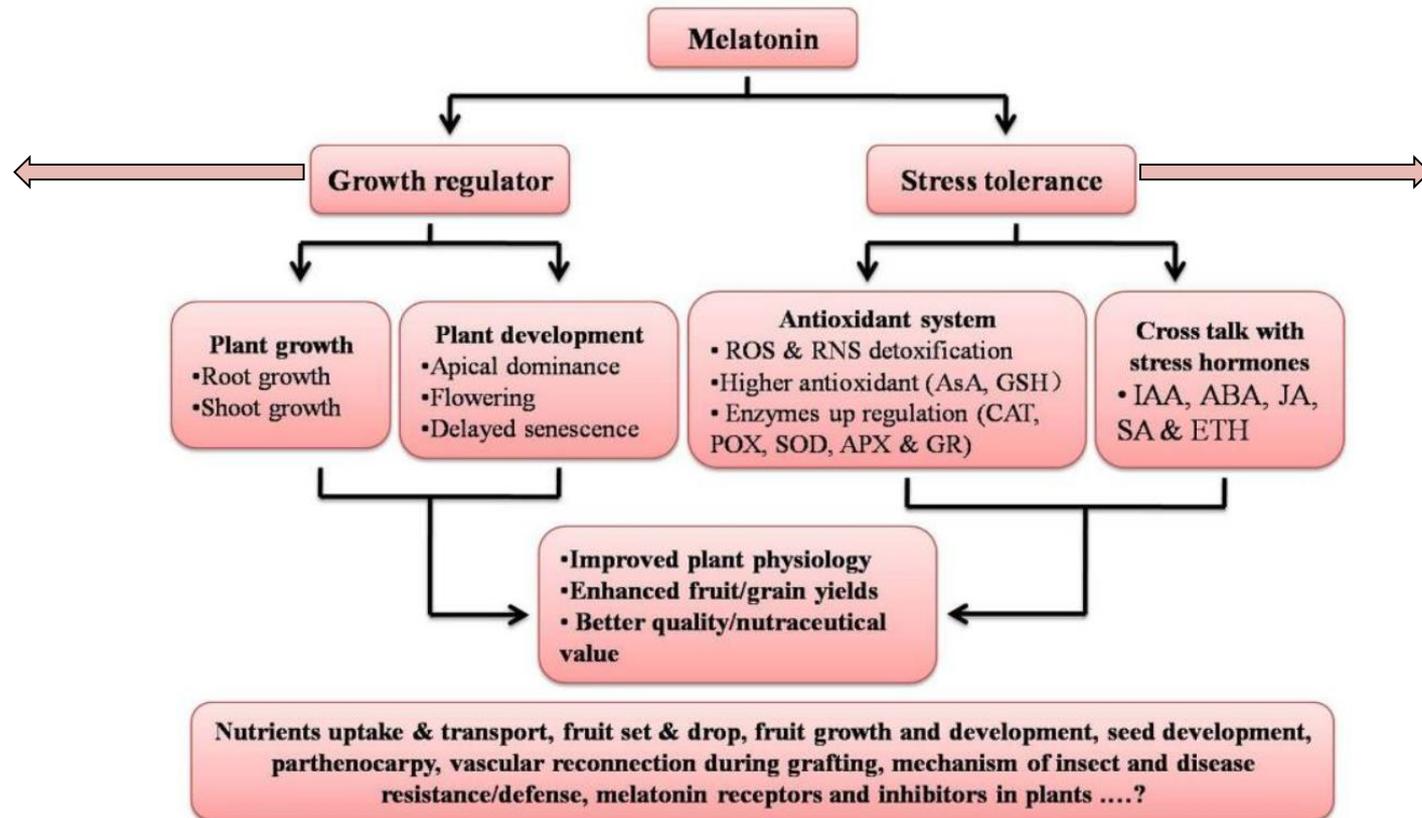
缓解逆境胁迫对植物的损伤

2.3 Biosynthesis of melatonin
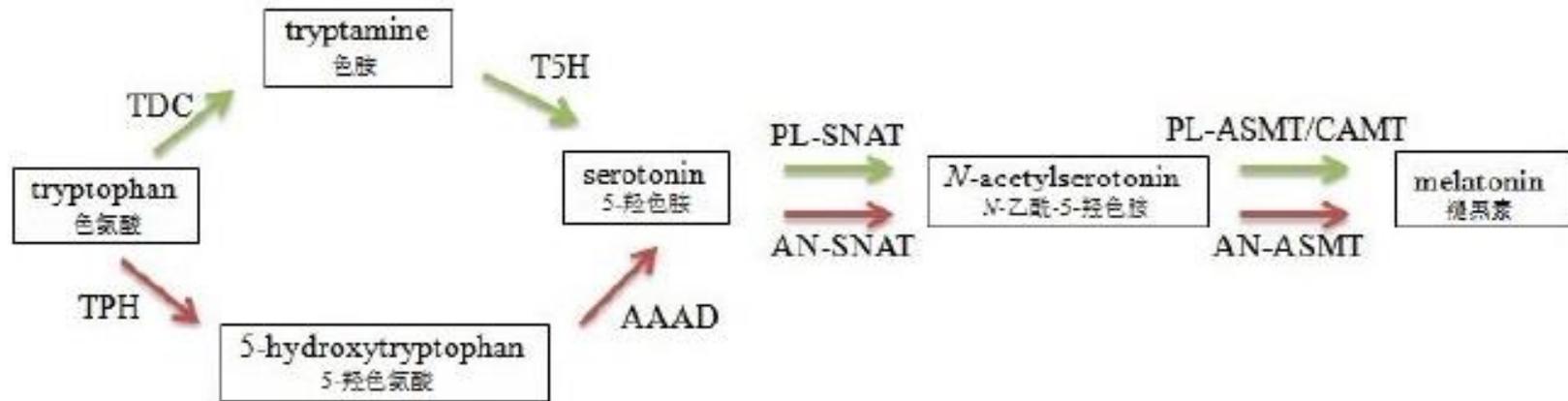
The green arrows identify the preferred pathway in plants while the red arrows identify the major pathway in animals. SNAT, serotonin N-acetyltransferase

褪黑素合成路径相关基因只在模式植物中被鉴定，其他高等植物（比如经济价值高的栽培果树——葡萄）中褪黑素合成基因的鉴定很有必要

3. Gene cloning

3.1 Screening of grape *SNAT* candidate genes

Run Blast: (blastp)

OsSNAT Genebank: BAG86973.1;

（**Lei et al., 2013)**

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq.
Specific hits
Superfamilies

Coenzyme A binding pocket

Acetyltransf_7

NAT_SF superfamily
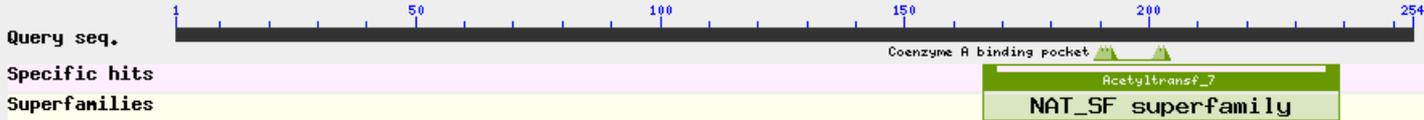
Distribution of the top 18 Blast Hits on 18 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

| <40 | 40-50 | 50-80 | 80-200 | >=200 |
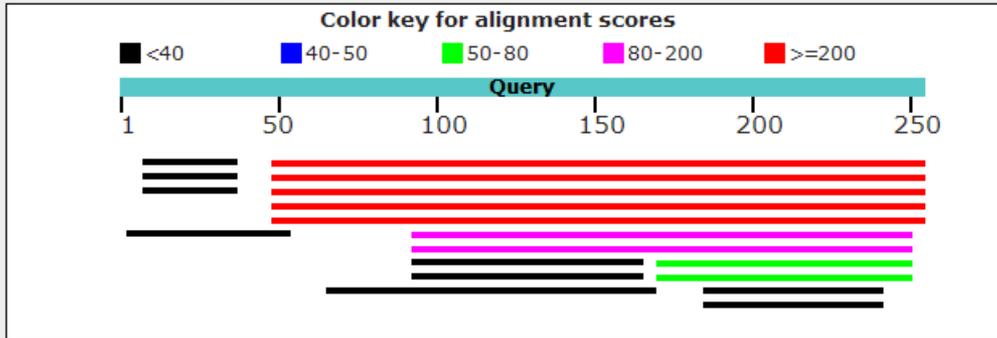
Query

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
unnamed protein product [Vitis vinifera]	299	299	81%	5e-103	72%	CBI31163.3		✔
PREDICTED: acetyltransferase NSI isoform X4 [Vitis vinifera]	299	299	81%	1e-102	72%	XP_002266361.2	✔	✔
PREDICTED: acetyltransferase NSI isoform X3 [Vitis vinifera]	291	291	81%	1e-99	69%	XP_010653125.1	✔	✔
PREDICTED: acetyltransferase NSI isoform X2 [Vitis vinifera]	283	283	81%	6e-96	60%	XP_019077452.1	✔	✔
PREDICTED: acetyltransferase NSI isoform X1 [Vitis vinifera]	276	276	81%	7e-93	58%	XP_019077451.1	✔	✔
hypothetical protein VITISV_037700 [Vitis vinifera]	113	113	62%	2e-30	40%	CAN70315.1		✔
PREDICTED: serotonin N-acetyltransferase 2, chloroplastic [Vitis vinifera]	113	113	62%	2e-30	40%	XP_002271276.1		✔
PREDICTED: uncharacterized N-acetyltransferase ycf52 [Vitis vinifera]	57.8	57.8	31%	2e-09	37%	XP_002270840.1		✔
unnamed protein product [Vitis vinifera]	57.4	57.4	31%	3e-09	37%	CBI25520.3		✔

PREDICTED: acetyltransferase NSI isoform X4 [Vitis vinifera]

http://www.zzgss.cn

Note：Principle for primer design

Primer Premier 6.0 Primer Premie...

Oligo 7

Vector NTI

Primer-BLAST
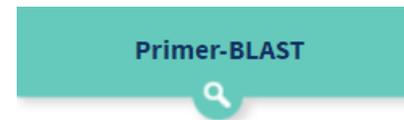
Design primers specific to your PCR template

（1）引物长度控制在15-30bp以内（18-25bp）

（2）GC含量一般设为40%-60%

（3）Tm一般设为55 ℃-65 ℃，退火温度一般要比解链温度低5 ℃

（4） 3'端不应超过4个连续的G或C

（5）引物自身不应存在互补序列，

（6）如扩增cds序列，引物3'端不要终止于密码子的第3位，因密码子的第3位易发生简并，会影响扩增特异性与效率。

3.2 Primer design（PCR ） (XM_002266325.4)

DNAMAN 8

4. sequence analysis and function prediction

4.1 Sequence align（DNAMAN，Needle）

匹配率是**99.33 %**

但是其编码的氨基酸
序列相同，而且也并
未出现移码现象

4.2.1 Search from UniProt datase

	Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
☑	PREDICTED: acetyltransferase NSI isoform X4 [Vitis vinifera]	519	519	100%	0.0	100%	XP_002266361.2	☑	
☐	PREDICTED: acetyltransferase NSI isoform X3 [Vitis vinifera]	510	510	100%	0.0	96%	XP_010653125.1	☑	
☐	PREDICTED: acetyltransferase NSI isoform X2 [Vitis vinifera]	504	504	100%	8e-180	86%	XP_019077452.1	☑	
☐	PREDICTED: acetyltransferase NSI isoform X1 [Vitis vinifera]	495	495	100%	3e-176	83%	XP_019077451.1	☑	
☐	unnamed protein product [Vitis vinifera]	433	433	83%	6e-153	100%	CBI31163.3	☑	
☐	PREDICTED: acetyltransferase NSI isoform X1 [Juglans regia]	394	394	100%	6e-137	77%	XP_018814362.1	☑	
☐	PREDICTED: acetyltransferase NSI isoform X1 [Theobroma cacao]	384	384	100%	7e-133	78%	XP_007033470.2	☑	
☐	PREDICTED: acetyltransferase NSI isoform X2 [Juglans regia]	382	382	88%	1e-132	83%	XP_018814363.1	☑	
☑	PREDICTED: serotonin N-acetyltransferase 1, chloroplastic isoform X3 [Nelumbo nucifera]	382	382	100%	2e-132	75%	XP_010244575.1	☑	
☐	Nuclear shuttle interacting isoform 2 [Theobroma cacao]	382	382	100%	4e-132	77%	EOY04396.1	☑	
☐	acetyltransferase NSI [Hevea brasiliensis]	382	382	95%	9e-132	81%	XP_021684353.1	☑	
☐	acetyltransferase NSI isoform X2 [Lactuca sativa]	379	379	96%	4e-131	76%	XP_023770682.1	☑	
☐	acetyltransferase NSI [Momordica charantia]	378	378	99%	1e-130	75%	XP_022159397.1	☑	
☐	acetyltransferase NSI [Cynara cardunculus var. scolymus]	376	376	94%	5e-130	80%	XP_024975398.1	☑	
☑	serotonin N-acetyltransferase 1, chloroplastic [Cajanus cajan]	376	376	96%	1e-129	76%	XP_020203403.1	☑	
☐	acetyltransferase NSI [Citrus clementina]	376	376	100%	2e-129	76%	XP_006429638.1	☑	
☐	hypothetical protein CUMW_001790 [Citrus unshiu]	375	375	100%	7e-129	76%	GAY32321.1	☑	
☑	serotonin N-acetyltransferase 1, chloroplastic isoform X2 [Ziziphus jujuba]	373	373	100%	7e-129	77%	XP_015897769.1	☑	

4.3 Phylogeny analysis c

通过Blastp搜索NR数据库中同源序列，选取identity ＞ 70%以上且注释为serotonin N-acetyltransferase的序列25条；搜索swiss-prot数据库中搜索到已审阅的identity ＞ 60%的拟南芥、水稻serotonin N-acetyltransferase同源序列3条；另从Uniprot数据库中找出5种已审阅的动物serotonin N-acetyltransferase序列加入系统进化树构建，共33个物种进行MEGA建树。

4.4 Alignment of SN

根据Mega建树图，选出与葡萄 serotonin N-acetyltransferase序列亲缘最近的其它3个物种序列，另加2个模式生物拟南芥、水稻序列进行**Clustal Omega法多序列比对**。

星号标记的为高度保守序列位点。
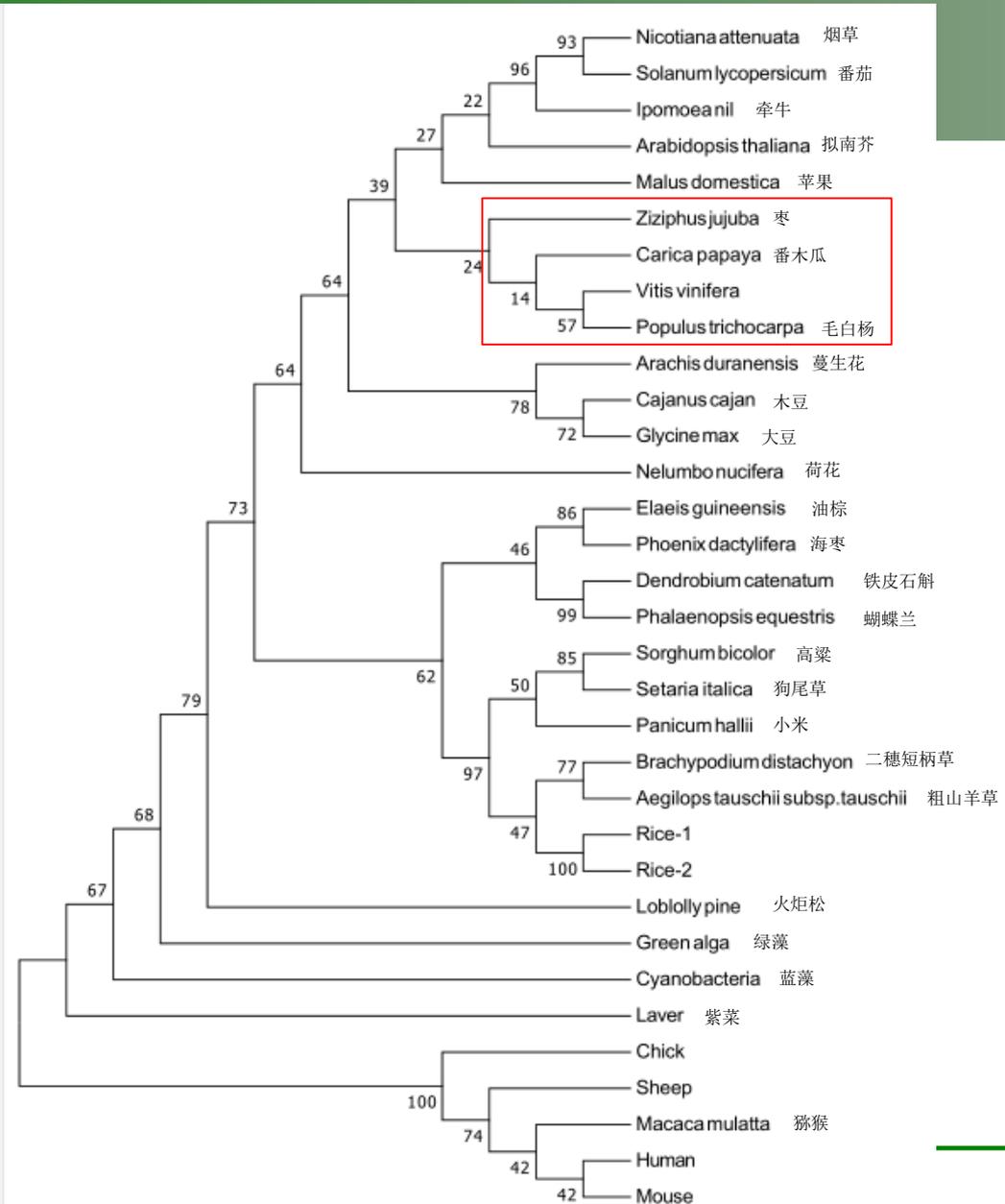
4.5.1 Analysis of domain

Pfam
Acetyltransf_7

低复杂度区域

乙酰转移酶结构域

Low complexity region

This is a region of low compositional complexity, as detected by the SEG program. The region starts at position **79** and ends at position **92**.

Low complexity region (14 aa):

| Submit to BLAST | Copy to clipboard |

EEEEPLPEEFVLVE

Acetyltransf_7 domain

This is a Pfam domain. Please see the Acetyltransf_7 entry in Pfam for full annotation.

Pfam
Acetyltransf_7

| Position: | 156 to 232 |
| E-value: | 1.4e-13 (HMMER3) |

Accession:

Pfam domain sequence (77 aa):

| Submit to BLAST | Copy to clipboard |

ERNEQKKLIGMARATSDHAFNATIWDVLVDPSYQGQGLGKALVEKTIRALLQRDIGNITL
FADSQVVEFYRNLGFEP

4.5.2 Analysis of ProtParam

```
Amino acid composition:    CSV format
Ala (A)   16      6.5%
Arg (R)   14      5.7%
Asn (N)   12      4.9%
Asp (D)   12      4.9%
Cys (C)    5      2.0%
Gln (Q)   12      4.9%
Glu (E)   17      6.9%
Gly (G)   18      7.3%
His (H)    4      1.6%
Ile (I)   12      4.9%
Leu (L)   26     10.6%
Lys (K)   13      5.3%
Met (M)    4      1.6%
Phe (F)   12      4.9%
Pro (P)   13      5.3%
Ser (S)   20      8.1%
Thr (T)   12      4.9%
Trp (W)    4      1.6%
Tyr (Y)    6      2.4%
Val (V)   14      5.7%
Pyl (O)    0      0.0%
Sec (U)    0      0.0%

 (B)       0      0.0%
 (Z)       0      0.0%
 (X)       0      0.0%
```

Number of amino acids: 246

Molecular weight: 27502.25

Theoretical pI: 6.12

Total number of negatively charged residues (Asp + Glu): 29
Total number of positively charged residues (Arg + Lys): 27

Instability index:

The instability index (II) is computed to be 43.99
This classifies the protein as unstable.

4.6.1 Hydrophilic and hydrophobic analysis （ProtScale）

数值越高表明该区域多肽链
的疏水性越强，该蛋白存在数
段亲水性较强的区域，蛋白
总体亲水性较强。
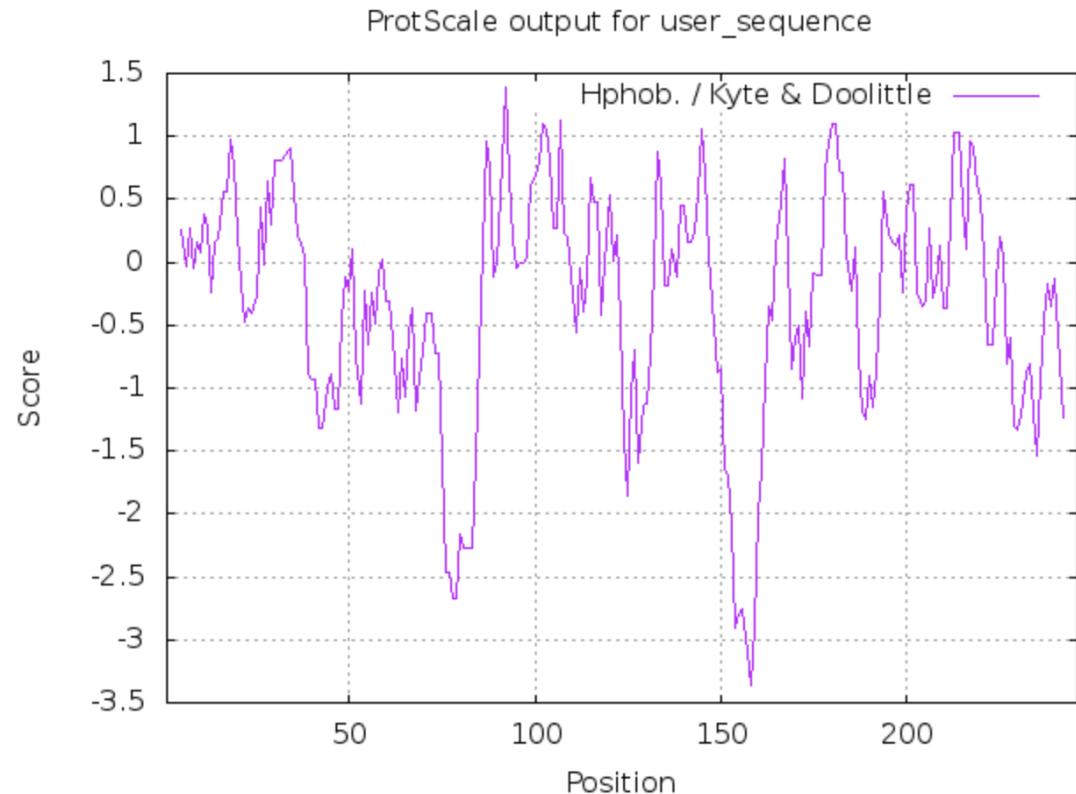
4.6.2 Signal peptide prediction

该蛋白无信号肽序列

TMHMM posterior probabilities for XP_002266361.2
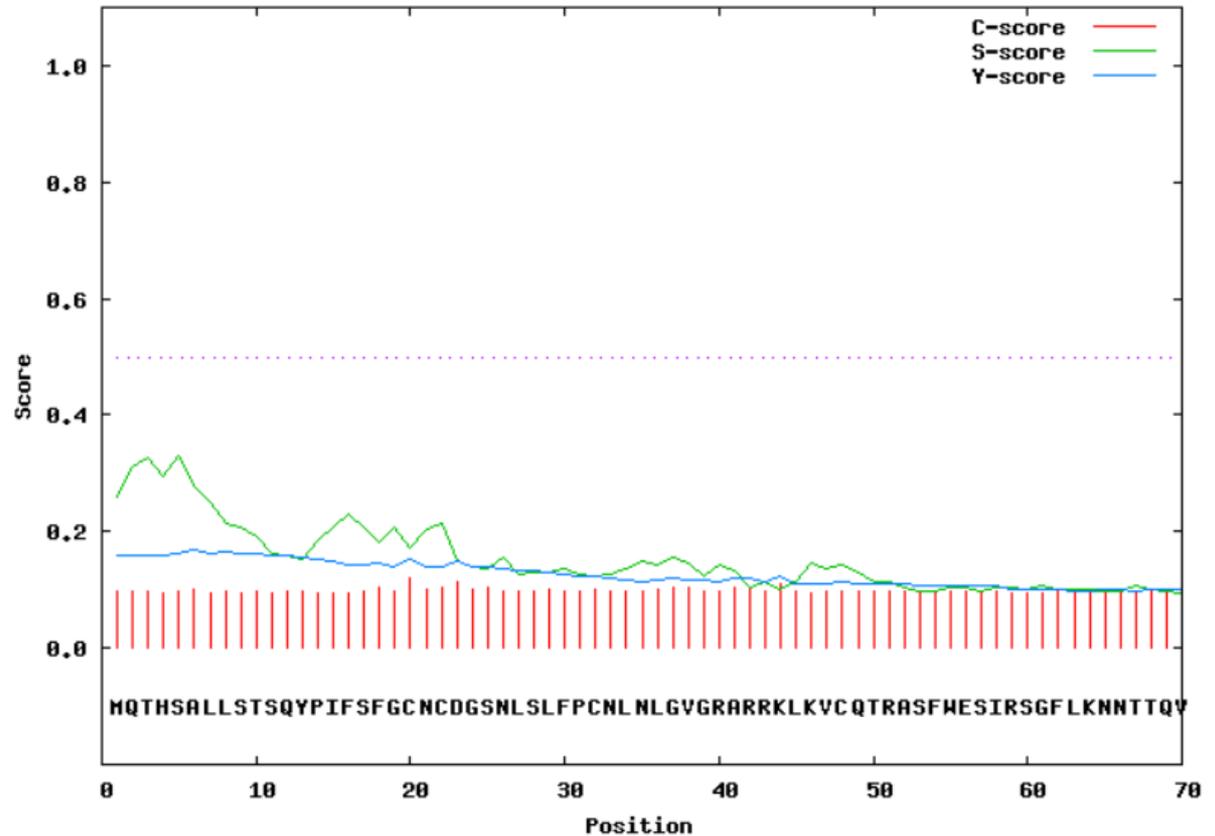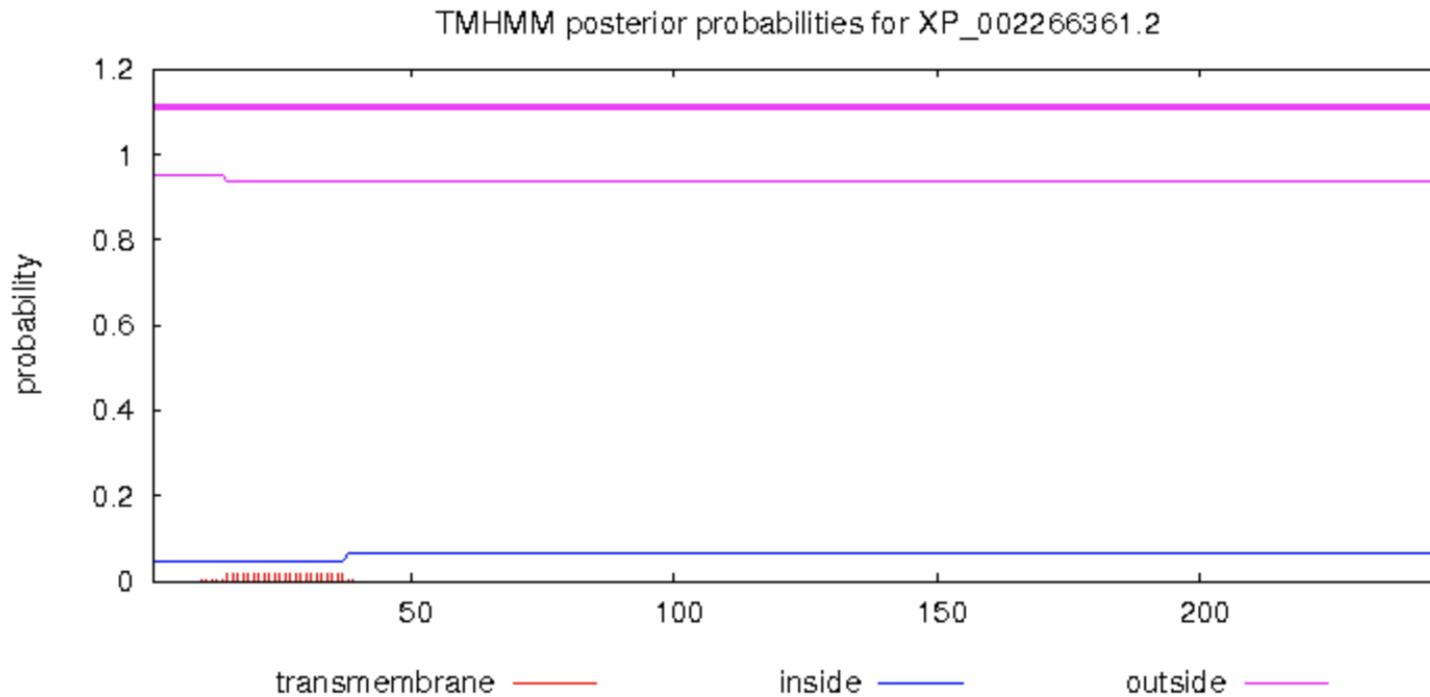
该蛋白无跨膜结构

4.6.4 Transit peptide prediction


```
CENTERFOR
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS
```

```
### chlorop v1.1
Number of query s

Name
_____
Sequence
_____
---Detailed outpu

Residue    --NN-sc
           Raw   D
_____
```

```
1      ATGCAAACCCACAGCGCCCTCCTCTCCACTTCTCAGTACCCCATTTTCTCTTTTGGTTGCAATTGTGACGGATCAAATTTATCATTGTTC

       M  Q  T  H  S  A  L  L  S  T  S  Q  Y  P  I  F  S  F  G  C  N  C  D  G  S  N  L  S  L  F

91     CCTTGCAATTTAAATCTTGGGGTTGGGAGAGCAAGAAGAAAGCTTAAGGTTTGTCAAACGAGAGCTAGCTTTTGGGAATCCATCAGATCC

       P  C  N  L  N  L  G  V  G  R  A  R  R  K  L  K  V  C  Q  T  R  A  S  F  W  E  S  I  R  S

181    GGGTTTTTGAAGAATAATACAACACAAGTTATTGAACCACCCTCCACAAACCAAGAAGAGGAAGAACCATTGCCCGAGGAGTTTGTTCTT

       G  F  L  K  N  N  T  T  Q  V  I  E  P  P  S  T  N  Q  E  E  E  E  P  L  P  E  E  F  V  L

271    GTTGAAAAGACTCTAGCTGACGGAGCAATTGAACAGATAATATTTTCTTCAGGTGGAGATGTTGATGTCTATGATCTCCAGGCCTTATGT

       V  E  K  T  L  A  D  G  A  I  E  Q  I  I  F  S  S  G  G  D  V  D  V  Y  D  L  Q  A  L  C

361    GATAAGGTGGGCTGGCCCCGAAGGCCTCTATCAAAACTATCTGCAGCTTTAAAAAATAGCTACATGGTAGCCACATTGCATTCCTTAAGG

       D  K  V  G  W  P  R  R  P  L  S  K  L  S  A  A  L  K  N  S  Y  M  V  A  T  L  H  S  L  R

451    AAATCACCTGGAGAAGAGAGGAATGAGCAAAAGAAGCTTATTGGCATGGCCCGTGCTACATCAGATCATGCCTTCAATGCCACAATTTGG

       K  S  P  G  E  E  R  N  E  Q  K  K  L  I  G  M  A  R  A  T  S  D  H  A  F  N  A  T  I  W

541    GATGTCCTTGTTGATCCTTCATATCAGGGCCAGGGACTTGGAAAGGCCCTCGTTGAAAAGACTATAAGAGCTCTTCTGCAAAGGGACATT

       D  V  L  V  D  P  S  Y  Q  G  Q  G  L  G  K  A  L  V  E  K  T  I  R  A  L  L  Q  R  D  I

631    GGCAATATAACACTCTTTGCAGATAGCCAAGTGGTGGAGTTCTATCGAAATTTAGGTTTTGAACCTGACCCAGAGGGCATCAAAGGTATG

       G  N  I  T  L  F  A  D  S  Q  V  V  E  F  Y  R  N  L  G  F  E  P  D  P  E  G  I  K  G  M

721    TTCTGGCACCCAAGGTATTAG

       F  W  H  P  R  Y  *
```

4.6.4 Subcellular localization（Target P）

Localization of OsSNAT. (A) Bright-field image of Agrobacterium- infiltrated tobacco. (B) The red fluorescence of OsSNAT-mCherry and (C) the yellow fluorescence of chlorophyll (Chl) were monitored 12 hr after XVE induction, and (D) the two fluorescence images were merged. N. benthamiana leaves infiltrated with Agrobacterium with XVE-inducible OsSNAT-mCherry were grown for 2 days in a growth room before XVE induction and visualization by confocal microscopy. Bars = 25 lm.

4.3 3D structure prediction
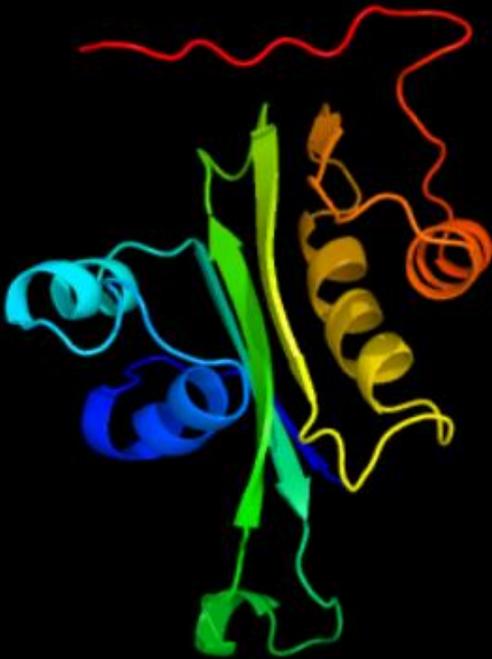
Model (left) based on template d2atra1

Top template information

Fold: Acyl-CoA N-acyltransferases (Nat)
Superfamily: Acyl-CoA N-acyltransferases (Nat)
Family:N-acetyl transferase, NAT

Confidence and coverage

Confidence: **99.9%**   Coverage: **52%**

129 residues (52% of your sequence) have been modelled with 99.9% confidence by the single highest scoring template.

(Phary2)

4.7.1 The protein interaction network of Human SNAT
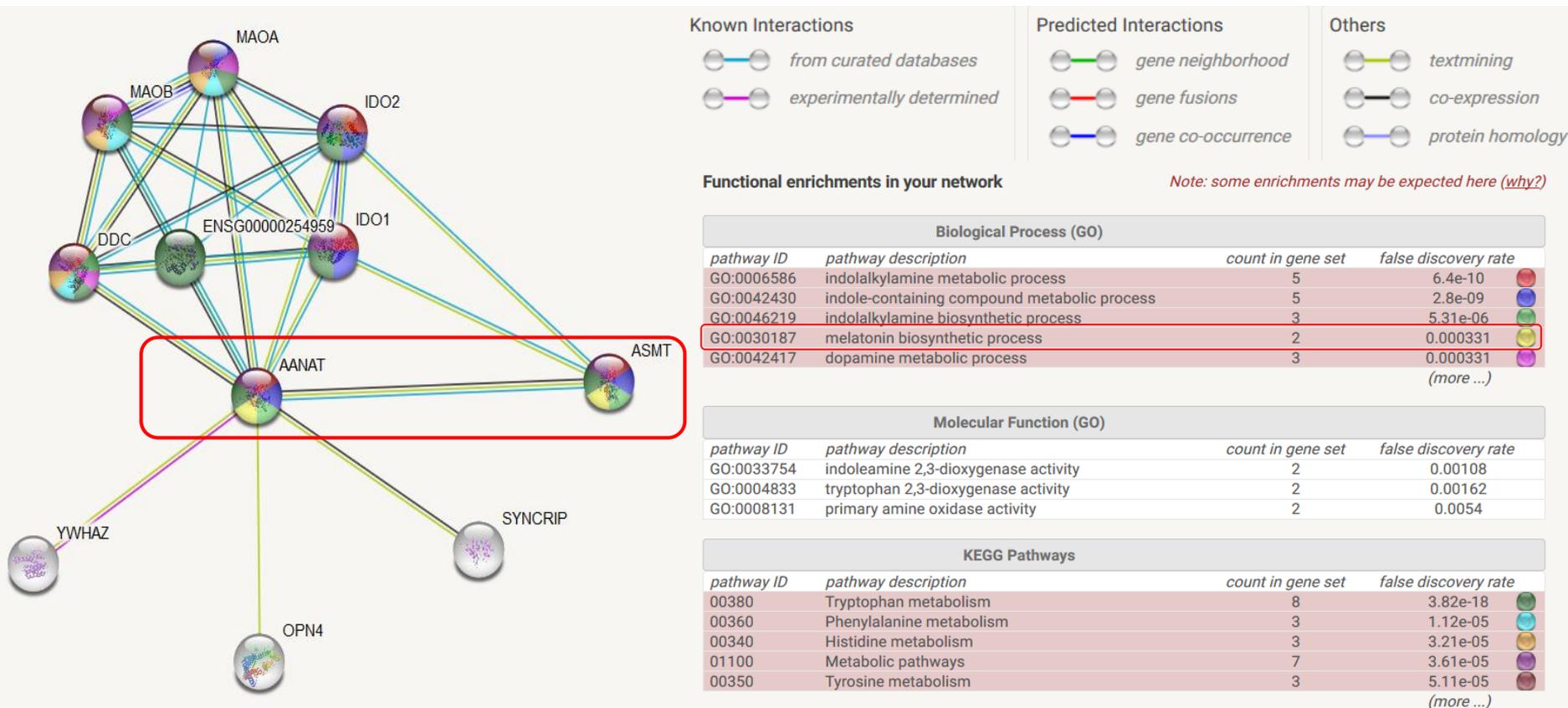
4.7.2 Metabolic pathway of rice

TRYPTOPHAN METABOLISM
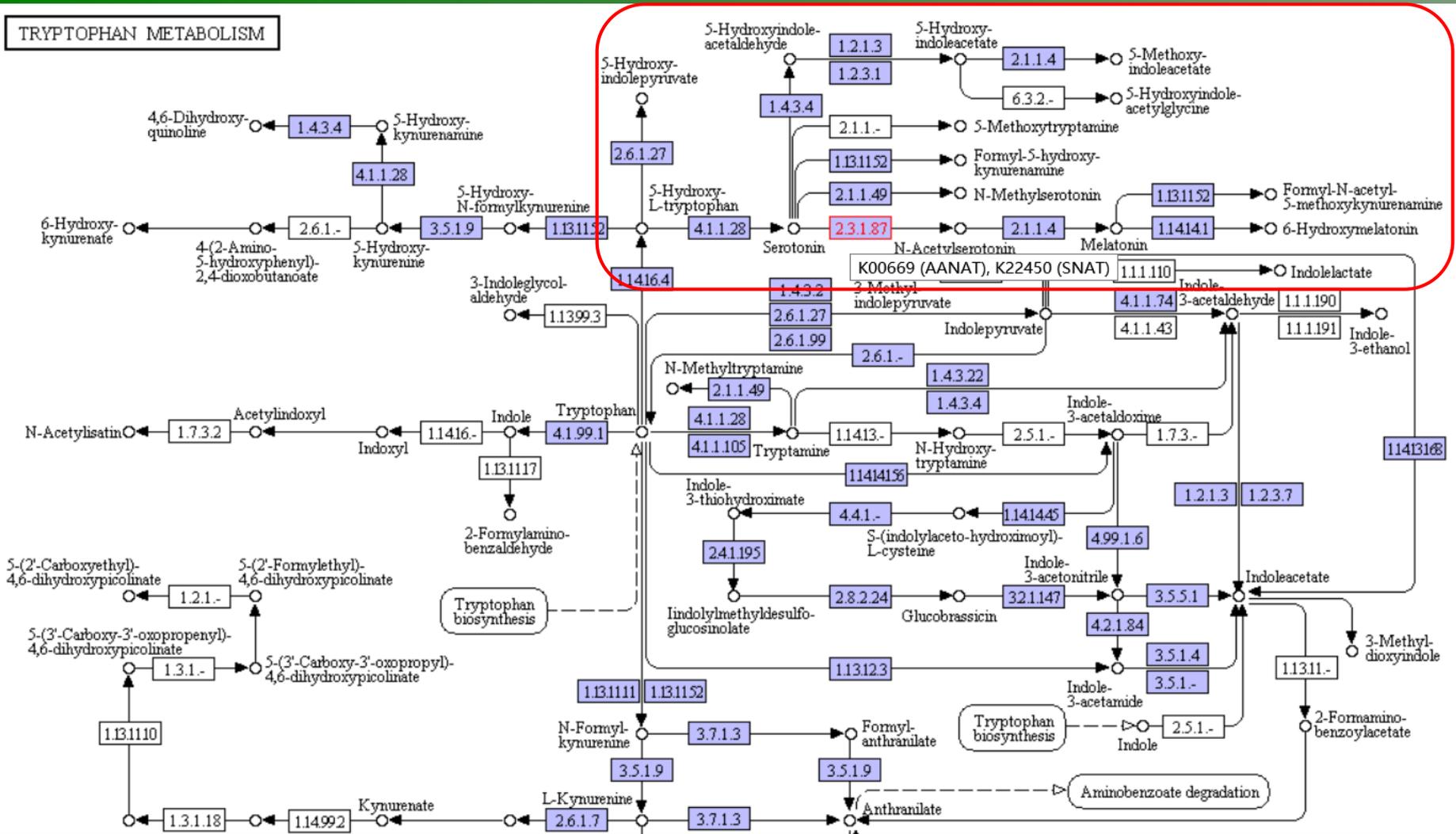
Problem

葡萄中**SNAT**（**5-羟色胺-N-乙酰基转移酶**）是不是可能也具有与人，拟南芥中**SNAT**相似的功能，可以将**5-羟色胺**转化为**N-乙酰-5-羟色胺**？

推理

验证

?

!

目的基因 → 重组表达载体 ← 表达载体Pgex-4T-1

连接

重组表达载体

转化

大肠杆菌BL21(DE3)

诱导

可溶性蛋白

纯化+western blot

目的蛋白

酶学功能验证
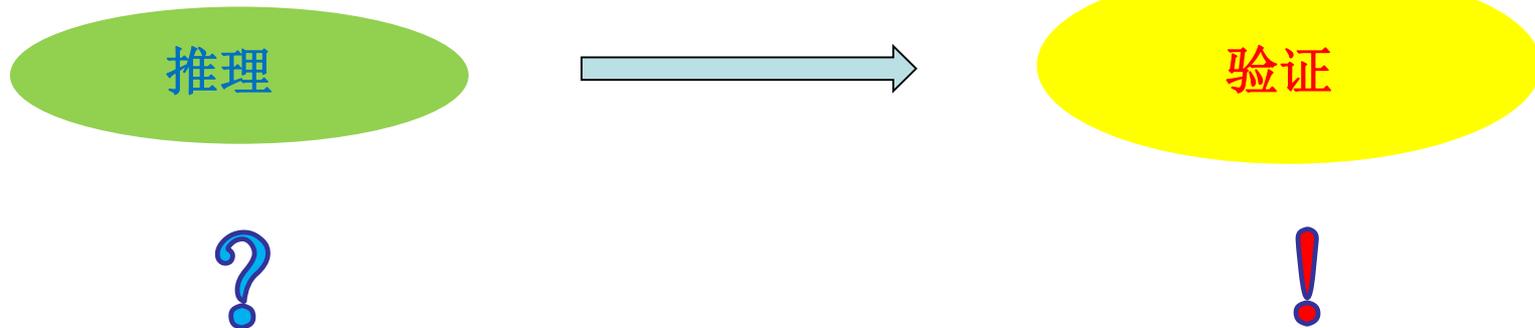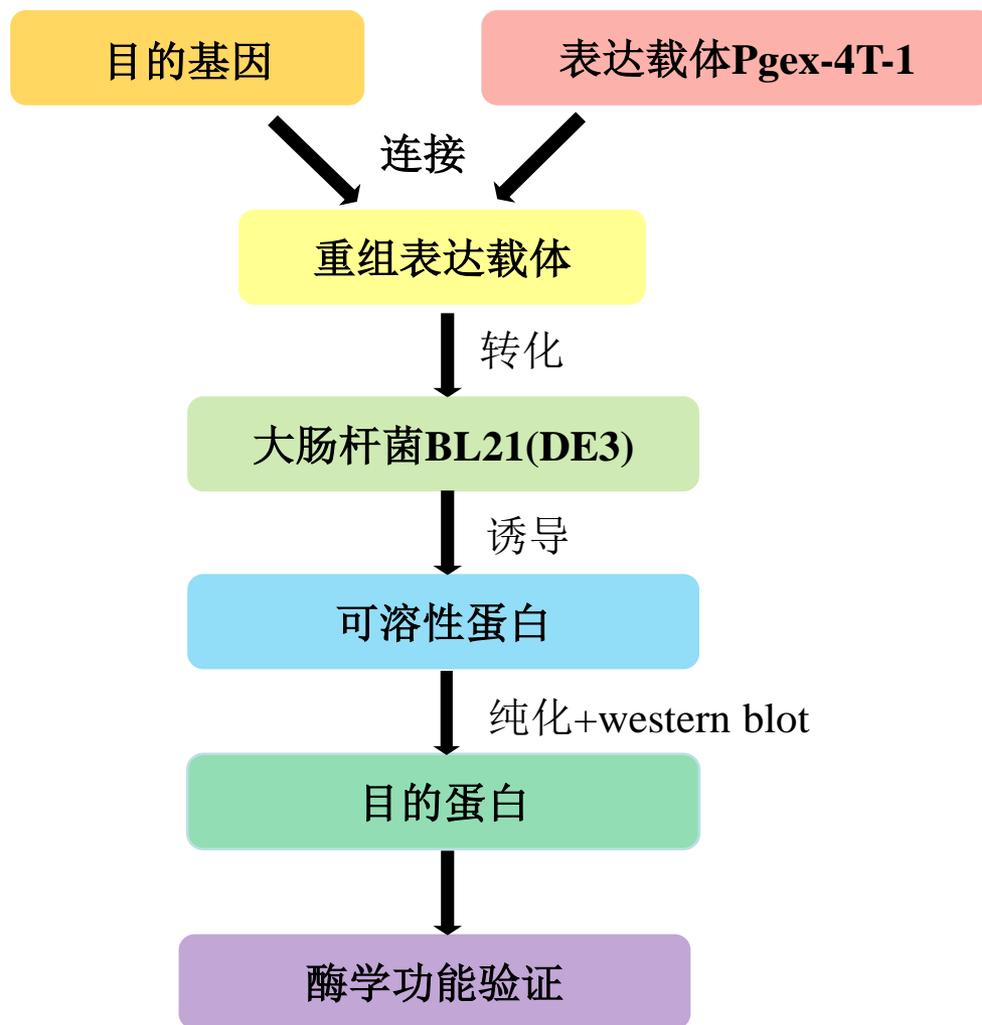
5.2 Rare codon analysis of amino acid

E. coli Codon Usage Analyzer 2.1

SNAT

ATG CAA ACC CAC AGC GCC CTC CTC TCC ACT TCT CAG TAC CCC ATT TTC TCT TTT GGT TGC
AAT TGT GAC GGA TCA AAT TTA TCA TTG TTC CCT TGC AAT TTA AAT CTT
GGG GTT GGG AGA GCA AGA AGA AAG CTT AAG GTT TGT CAA ACG AGA GCT AGC TTT TGG
GAA TCC ATC AGA TCC GGG TTT TTG AAG AAT AAT ACA ACA CAA GTT ATT GAA CCA CCC TCC
ACA AAC CAA GAA GAG GAA GAA CCA TTG CCC GAG GAG TTT GTT CTT GTT GAA AAG
ACT CUA GCT GAC GGA GCA ATT GAA CAG AUA AUA TTT TCT TCA GGTGGA GAT GTT GAT GTC
TAT GAT CTC CAG GCC TTA TGT GAT AAG GTG GGC TGG CCC CGA AGG CCT CUA TCA
AAA CUA TCT GCA GCT TTA AAA AAT AGC TAC ATG GTA GCC ACA TTG CAT TCC TTA AGG AAA
TCA CCT GGA GAA GAG AGG AAT GAG CAA AAG AAG CTT ATT GGC ATG GCC CGT GCT ACA
TCA GAT CAT GCC TTC AAT GCC ACA ATT TGG GAT GTC CTT GTT GAT CCT TCA TAT CAG GGC
CAG GGA CTT GGA AAG GCC CTC GTT GAA AAG ACT AUA AGA GCT CTT CTG CAA AGG GAC
ATT GGC AAT AUA ACA CTC TTT GCA GAT AGC CAA GTG GTG GAG TTC TAT CGA AAT TTA GGT
TTT GAA CCT GAC CCA GAG GGC ATC AAA GGT ATG TTC TGG CAC CCA AGG TAT TAG

稀有密码子均匀分布，而且没有连续三个的稀有密码子排列，对翻译蛋白的可溶性影响不大

http://www.zzgss.cn

5.3 Plasmid profile of pGEX-4T-1

含有**GST**促溶标签，有助于重组蛋白的可溶性表达

请各位老师、同学批评指导

谢谢！