实用生物信息技术期末总结交流报告会

百合花青苷调控相关转录因子*LhWRKY44* 的序列、结构与功能分析

Sequence, Structure, and Function Analysis of Anthocyanin Regulatory Transcrption Factor LhWRKY44 in Asiatic hybrid lilies

2020年12月12日

报告人:毕蒙蒙

中国农业科学院研究生院

2020级博士班

邮箱：1262665213@qq.com

微信号：15944458060

小组成员

编号	姓名	研究所	导师	研究方向
G1A	毕蒙蒙	蔬菜花卉	明军	观赏园艺-百合种质资源与遗传育种
G1B	常立春	蔬菜花卉	王晓武	蔬菜学-芥菜叶片性状研究
G1C	孙鹏雷	植保所	李香菊	杂草学
G1D	汤金荣	植保所	梁革梅	昆虫毒理学

CONTENT

PART1   Background

PART2   Sequence Analysis

PART3   Structure Analysis

PART4   Function Prediction

PART5   Primer Design

Background

What factors affect the value of Lily?

Flower color

Plant type and flower shape

Flower fragrance

Anthesis

Resistance

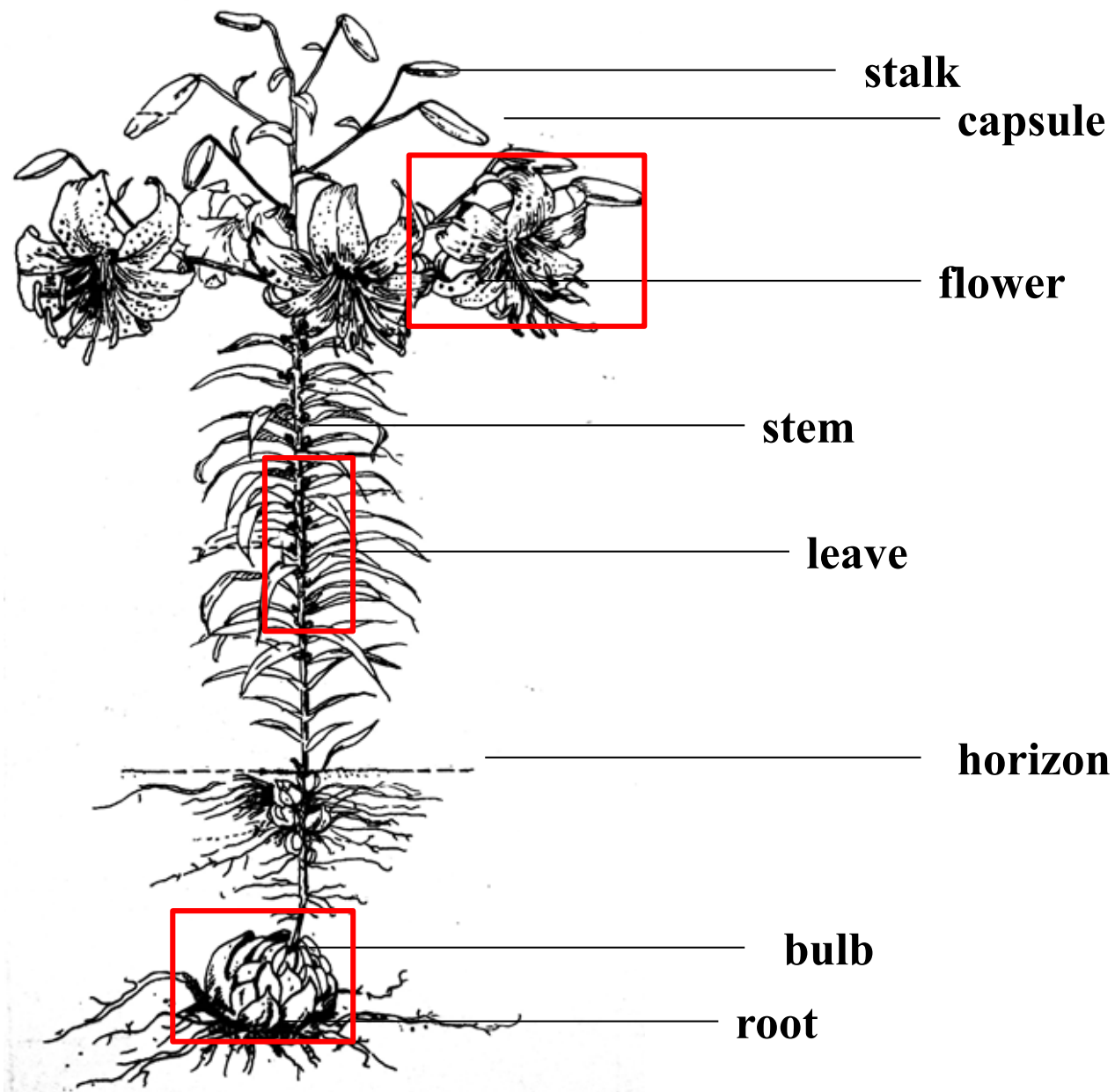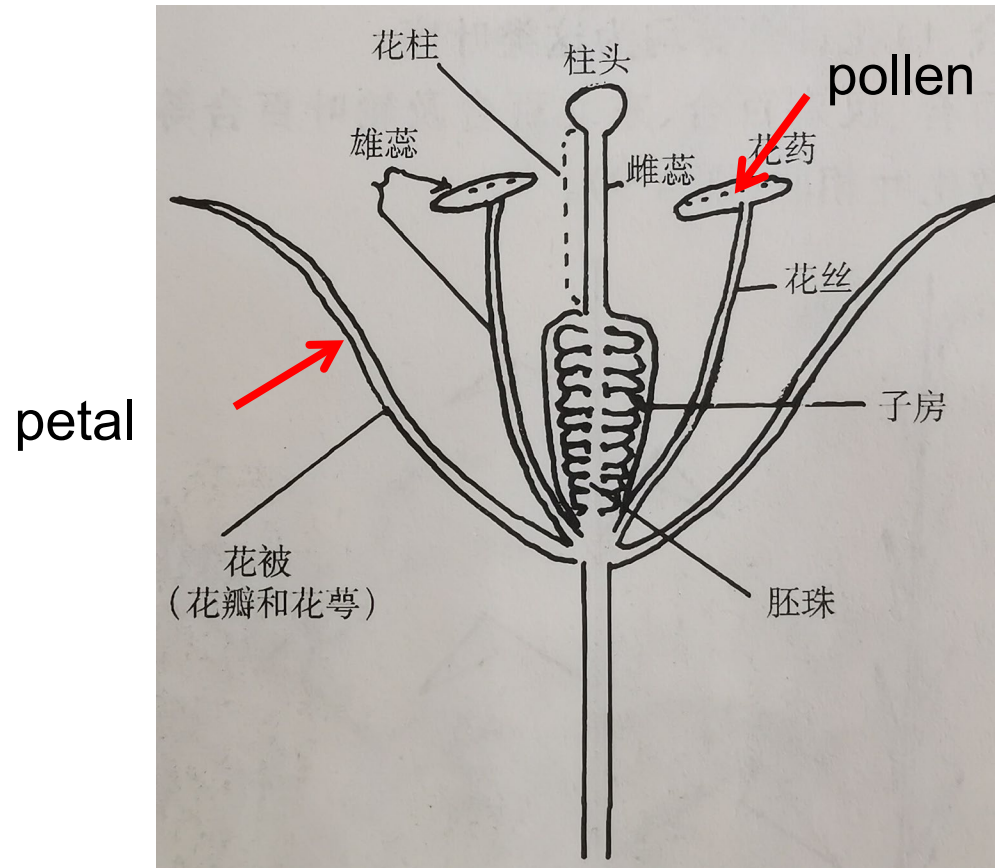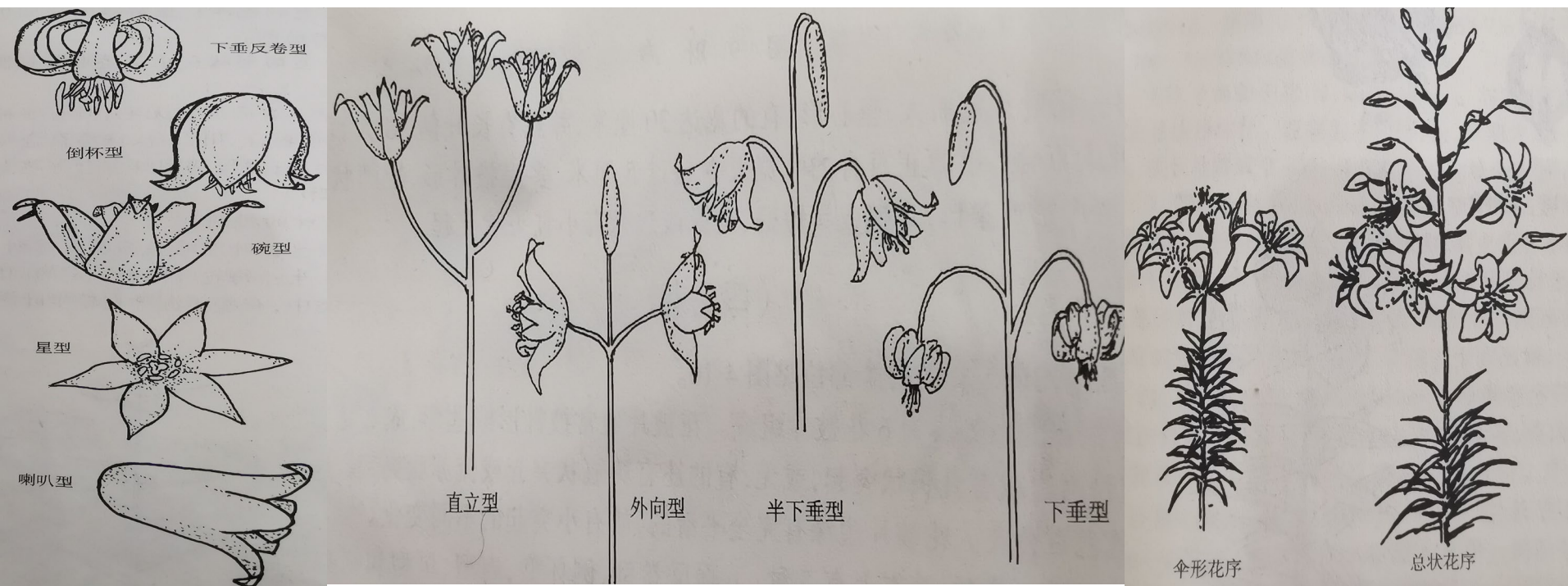明黄的文心兰、火红的玫瑰、鲜艳的红掌、芬芳的百合
百合具有纯洁、伟大的爱和深深祝福的意义

stalk

capsule

flower

stem

leave

horizon

bulb

root

Fig. 1 The main parts of Lily

摘自《百合，球根花卉之王》

花柱 柱头 pollen

雄蕊 雌蕊 花药

petal 花丝

子房

花被
（花瓣和花萼）

胚珠

Flower structure

下垂反卷型

倒杯型

碗型

星型

喇叭型

直立型　　　外向型　　　半下垂型　　　下垂型

伞形花序　　　总状花序

Plant types and flower shapes

Flower pigments

Chromogenic substance
显色物质

chlorophylls
叶绿素

flavones
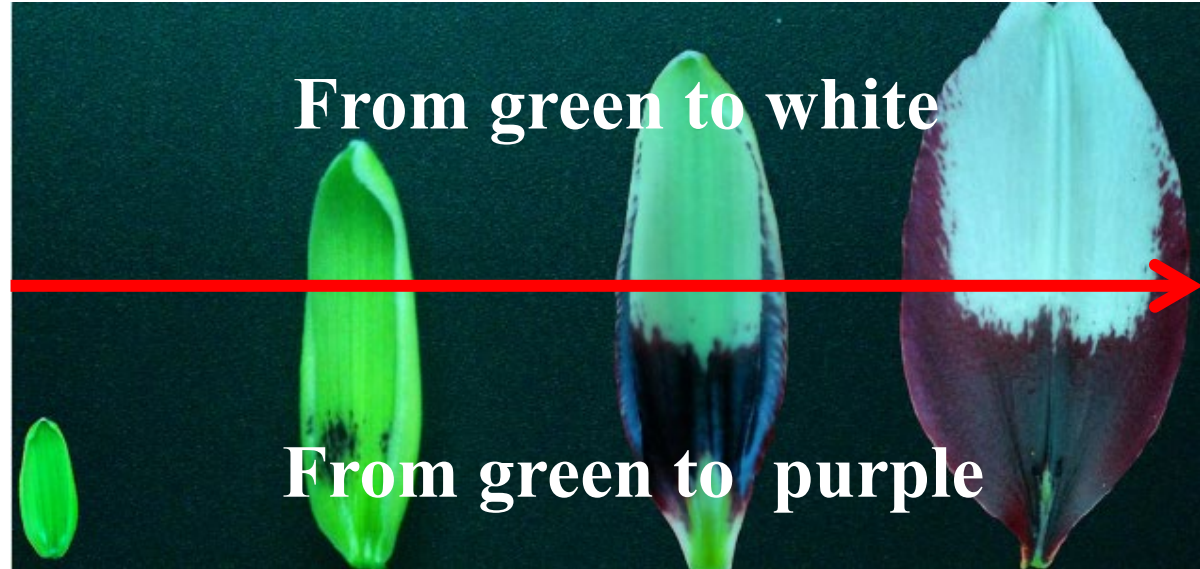类黄酮

carotenoids
类胡萝卜素

anthocyanins
花青苷

betalains
甜菜色素

Background

From green to white

From green to purple

The Asiatic lily cultivar

Background

Based on the functional annotation and gene expression analyses of transcriptome sequencing data in Lilium 'Tiny Padhye', a unigene (>c117585_g2) was identified as a putative anthocyanin-related WRKY unigene (*LhWRKY44*), for which the expression pattern was positively correlated with those of anthocyanin biosynthetic genes (Xu et al. 2017).

Background

➢ Anthocyanins are responsible for the pink, red and purple pigmentation of fowers in Asiatic hybrid lilies, synthesized at the cytoplasmic surface of the endoplasmic reticulum and then transported to the vacuole.

➢ To date，the molecular mechanisms underlying these changes remain largely uncharacterized.

➢ How does WRKY TF participate in and regulate anthocyanins？

WRKY family transcription factor
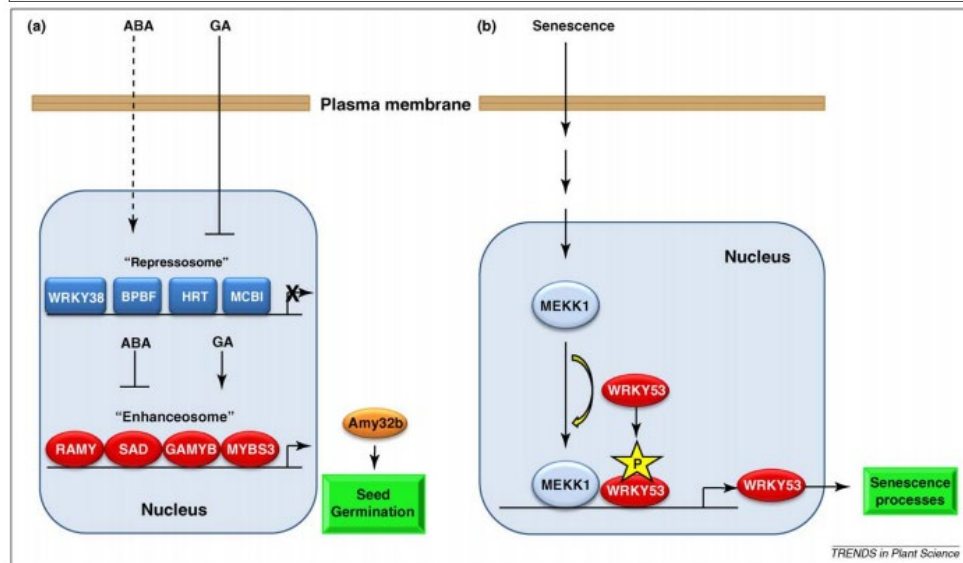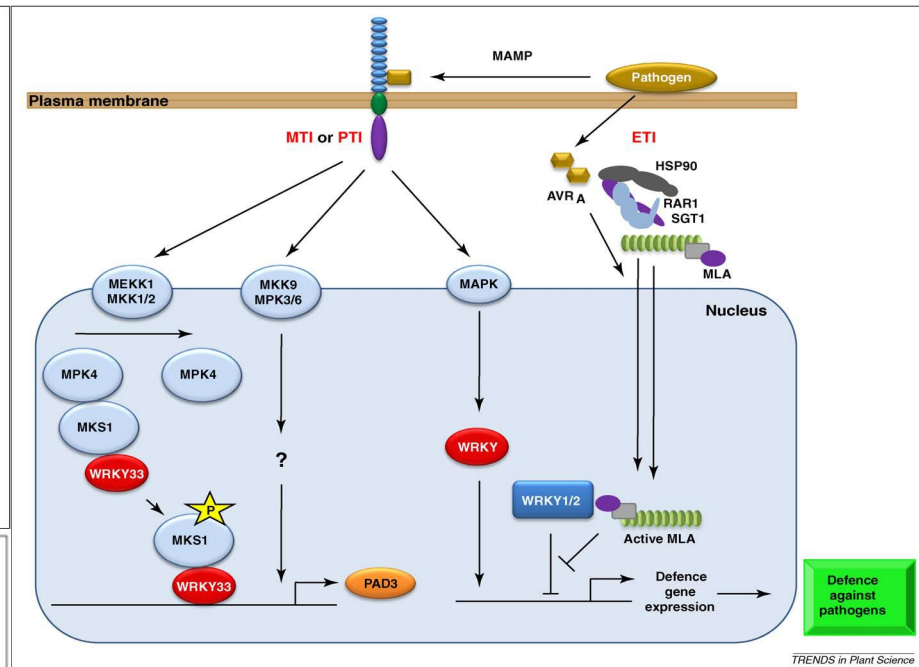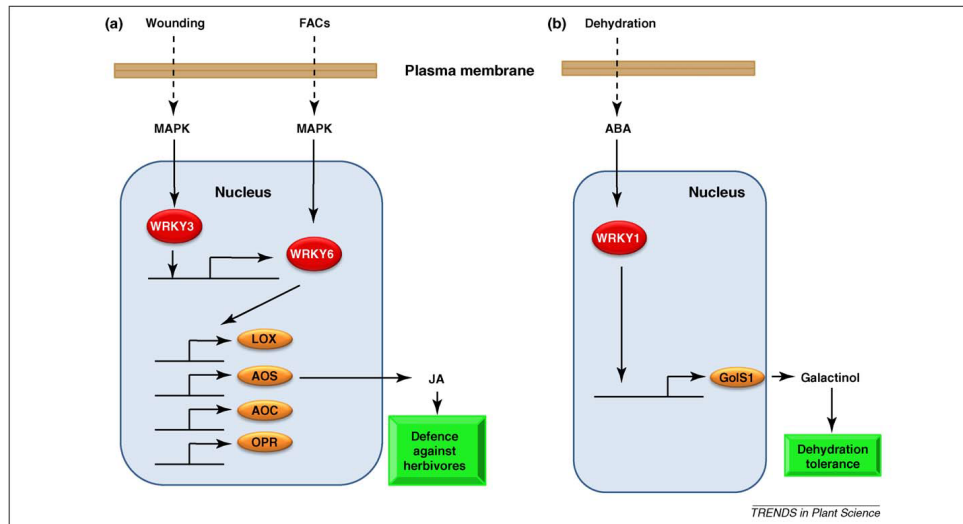
Plant growth and development
Biotic and abiotic stress responses
Leaf senescence
Anthocyanin, proanthocyanidin biosynthesis
Flowering
Seed development
Trichome patterning

(Paul J et al , 2010)

WRKY family transcription factor
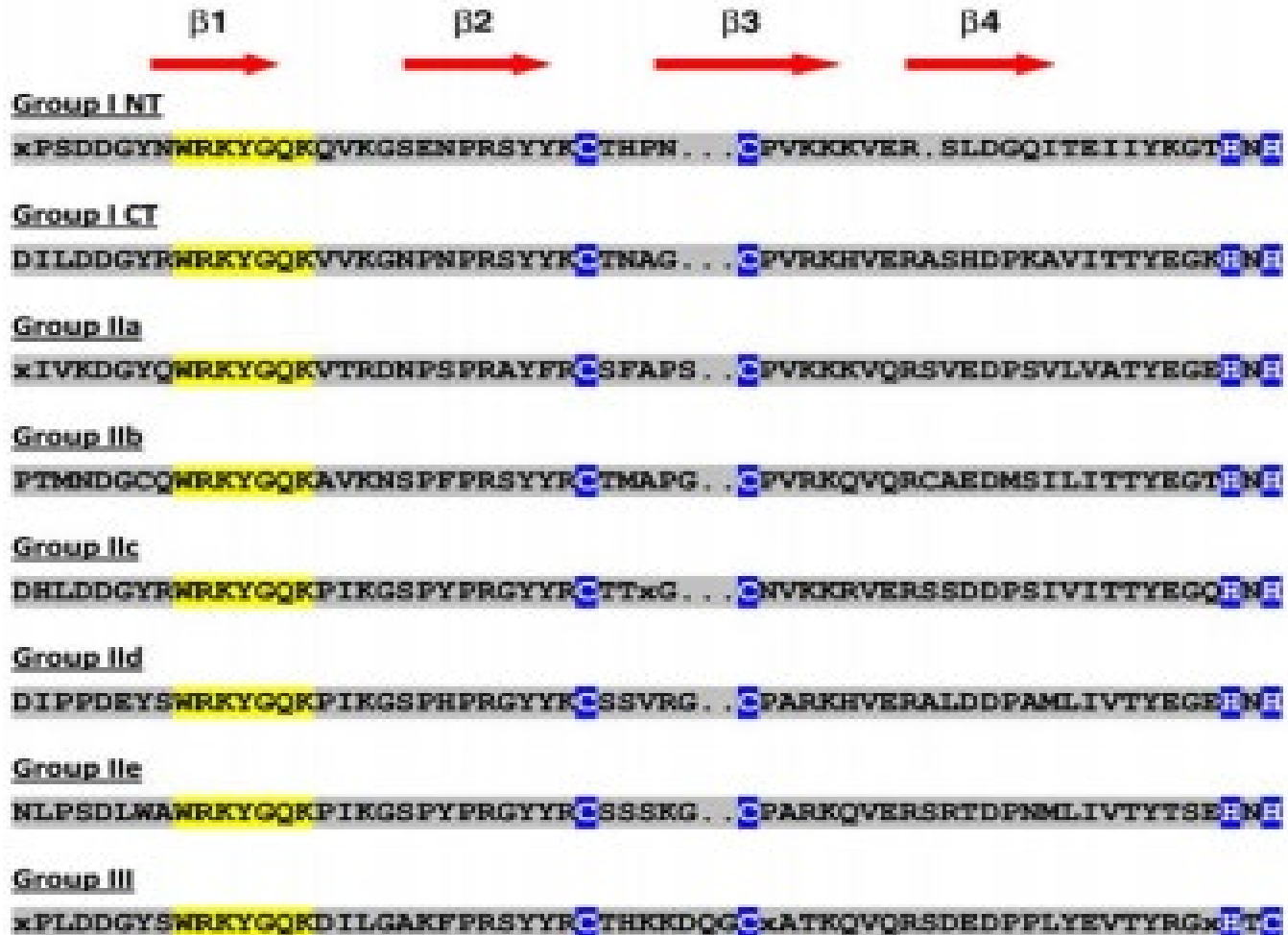
Fig.2   The WRKY domain.
The WRKY motif is highlighted in yellow and the cysteines and histidines that form the zinc finger are shown in blue. The four b-strands are shown in red. I CT and I NT denote the N-terminal and C-terminal WRKY domains from Group I WRKY proteins. (Paul J et al , 2010)

WRKY family transcription factor
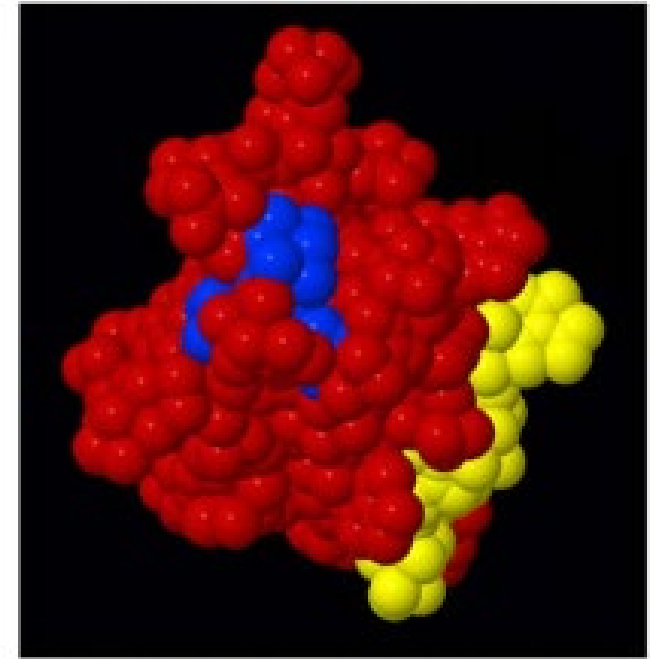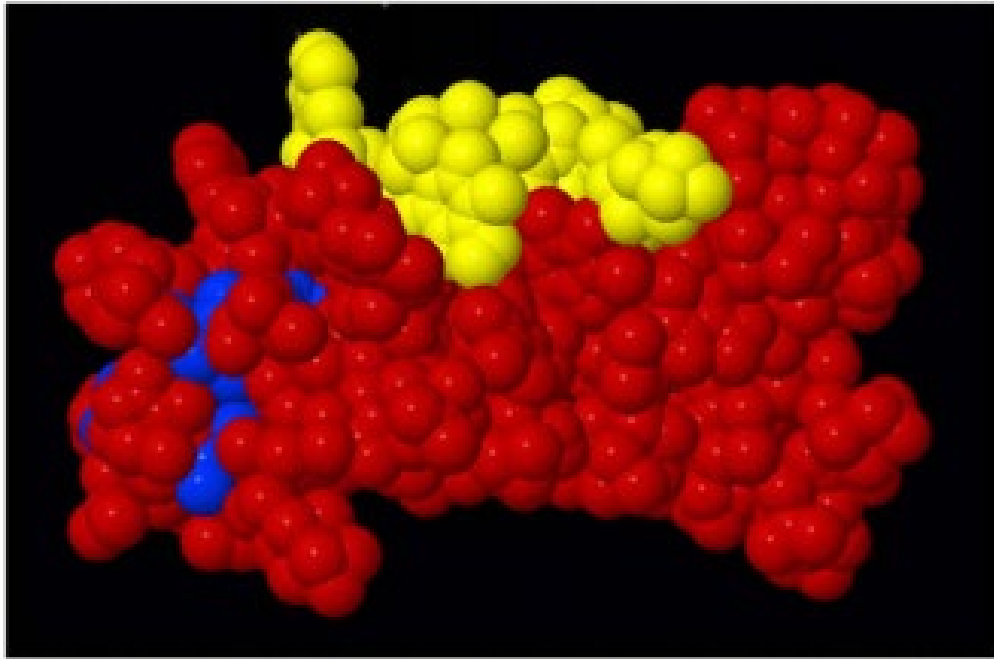
Q9XI90-WRKY4_ARATH   AT1G13960

Fig. 3 Two views of a spacefill structural model of the C-terminal
WRKY domain from AtWRKY4
the WRKYGQK motif is shown in yellow and the cysteines and
histidines that form the zinc_x0002_binding pocket are shown in
blue
(Paul J et al , 2010)
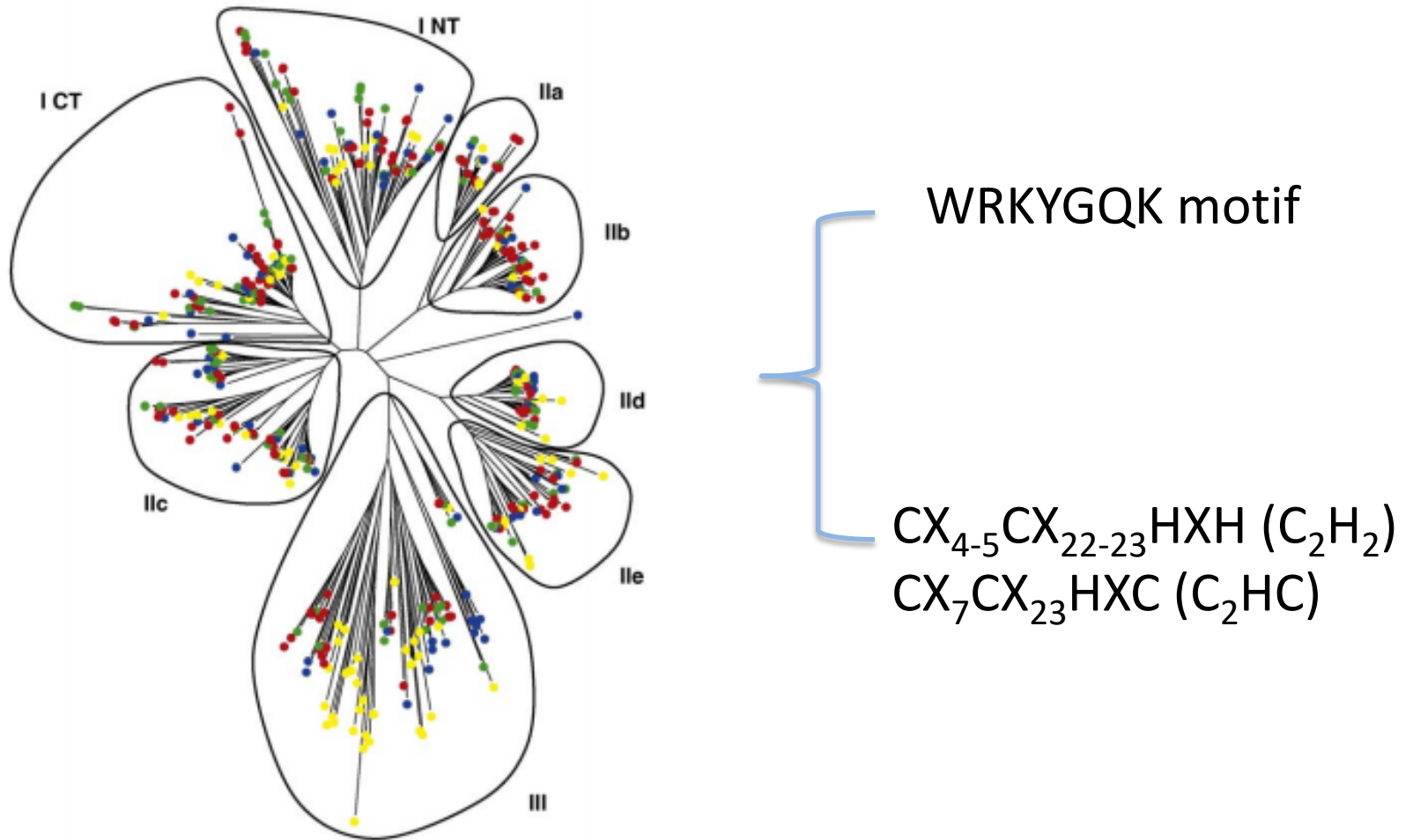
WRKY family transcription factor

WRKYGQK motif

$CX_{4-5}CX_{22-23}HXH$ (C_2H_2)
$CX_7CX_{23}HXC$ (C_2HC)

Fig. 4   Phylogenetic tree of the WRKY family in higher plants

(Paul J et al , 2010)

Plant Transcription Factor Database

Previous version: v3.0 v4.0

Home   TFext   BLAST   Prediction   Download   Help   About   Links   PlantRegMap   Search   (e.g., LFY)

Program blastp - (protein vs protein) ▾

Species All Species ▾

Enter a FASTA sequence (or paste an example)

```
>LhWRKY44
MPSQNNPPAEPASSMDTVSESPIFRSSDNLSIPTSREDSRSNVIYKPTAKVVSRTTASLLAKMGNDICHQQPLEEQTQ
FQVPNQVNQIQSTFDIYQNMPSDMLMNQVYEPNISESSNLMAPNTELDNIRDRPSYDGYNWRKYGQKQVKGSEYPRSY
YKCTHPICPVKRMVERSIDGKIAEIVYKGEHNHPKPQPPKRLSSASQGQTSVASEQAREIEEPLWSDLLIEKNLSDSR
MDGHIDVNFYGALDSPKVQCSHDSLIGATYSSGIMAPDPSQRLGADLEARSKGTVVDDGLHRSKRRKNLNLDSEAGAS
RNITAELHSVFQAPTESDVSADGFRWRKYGQKVVKGNSYPRCYYRCTSPKCSVRKYVERASGDTRSFVTTYEGKHDHD
KPEKKVDLMHKNERKLG
```

Scoring Matrix BLOSUM62 ▾

Color key for alignment score

| < 40 | 40 - 50 | 50 - 80 | 80 - 200 | > 200 |
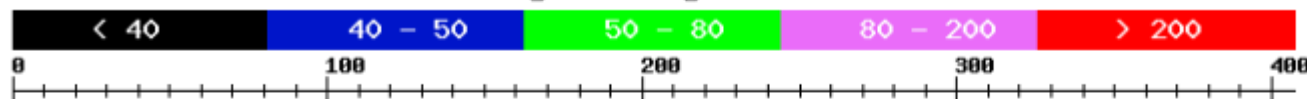
Query: LhWRKY44

XP_010908965.1
XP_010908962.1
XP_010922488.1
XP_010922486.1
XP_010922487.1
PDK_30s1053181g004
FANhyb_rscf00001029.1.g000
mrna18152.1-v1.0-hybrid
Pbr008278.1
PDK_30s880731g006
XP_010908967.1
XP_010908966.1
XP_008242045.1
Jcr4S02163.20
MDP0000268364
WALNUT_00025280-RA
GSMUA_Achr4P16840_001
29805.m001504

Sequencecs Alignment analysis

> PlantTFDB
> blastp
> All species
> LhWRKY44

Hit ID	Description	Score	E-Value
XP_010908965.1	Elaeis guineensis WRKY family protein	975	1e-128
XP_010908962.1	Elaeis guineensis WRKY family protein	975	1e-128
XP_010922488.1	Elaeis guineensis WRKY family protein	964	1e-126
XP_010922486.1	Elaeis guineensis WRKY family protein	964	1e-126
XP_010922487.1	Elaeis guineensis WRKY family protein	964	1e-126
PDK_30s1053181g004	Phoenix dactylifera WRKY family protein	892	1e-114
FANhyb_rscf00001029.1.g00002.1	Fragaria x ananassa WRKY family protein	879	1e-113
mrna18152.1-v1.0-hybrid	Fragaria vesca WRKY family protein	873	1e-112
Pbr008278.1	Pyrus bretschneideri WRKY family protein	865	1e-111
PDK_30s880731g006	Phoenix dactylifera WRKY family protein	857	1e-111
XP_010908967.1	Elaeis guineensis WRKY family protein	853	1e-110
XP_010908966.1	Elaeis guineensis WRKY family protein	853	1e-110
XP_008242045.1	Prunus mume WRKY family protein	857	1e-110
Jcr4S02163.20	Jatropha curcas WRKY family protein	856	1e-110

Fig. 5 Align sequences of LhWRKY44

Sequencecs Alignment analysis

> NCBI Global Align（Needle） LhWRKY44- EgWRKY

NW Score	Identities	Positives	Gaps
890	219/487(45%)	264/487(54%)	88/487(18%)

```
Query    1    M-------------------------------------------------------PSQNN    6
              M                                                        PS N
Sbjct    1    MQDAEKIVVAKPVASRPYSNLRSFSELLADAMNASPVTSFAEATVTIRPKTVRFKPSPNA    60

Query    7    PPAEPASSMDTVSESPI-FRSSDNLSIPTSREDSRSNVIYKPTAKVVSRTTASLLAKMGN    65
              AE   SS+D +SES + F    SD  S    S+ D +++V+YKP A+VVSRTTAS LA +GN
Sbjct    61   ALAE-VSSLDNISESRVAFNISDKASTTISKADYKTSVLYKPAARVVSRTTASRLANLGN    119

Query    66   -DICHQQPLEEQTQFQVPNQVN-------QIQSTFDIYQNMPSDMLMNQVYEPNISESSN    117
              D  HQ P    T  Q P V        Q+Q T    QN  S + MN +Y+P      SN
Sbjct    120  FDAGHQNPA---TDVQAPVHVIGQEKHQFQLQPTLSCQQNSFSHIDMNHIYKP-----SN    171

Query    118  LMAPNTELDNIR------DRPSYDGYNWRKYGQKQVKGSEYPRSYYKCTHPICPVKRMVE    171
              +  +TE D  +        DRPSYDGYNWRKYGQKQVKGSE PRSYYKCTHP CPVK+ VE
Sbjct    172  SVPESTEQDTRKQPATGGDRPSYDGYNWRKYGQKQVKGSECPRSYYKCTHPNCPVKKKVE    231

Query    172  RSIDGKIAEIVYKGEHNHPKPQPPKRLSSASQGQTSVASEQAREIEEPLWSDLLIEKNLS    231
              RS DG+IAEIVYKGEHNHPKPQPPKRLSS SQGQT V+    +E + PL S  L+  N+S
Sbjct    232  RSFDGQIAEIVYKGEHNHPKPQPPKRLSSGSQGQTVVSEGHGKESDNPLRSHHLVWGNVS    291

Query    232  DSRMDGHIDVNFYGALDS-PKVQCSHDSLIGATYSSGIMAPDPSQRLGADLEARSKGTVV    290
              D +++  I+++        K   SHD   +Y+S    PD + RL    E  +     +
Sbjct    292  DGQVEKGIEISLPDTTSFLGKAYFSHDPPFATSYNSVARNPDSAGRLSGASEFGNATASI    351

Query    291  DDGLHRSKRRKNLNLDSEAGASRNITAELHSVFQAPTESDVSADGFRWRKYGQKVVKGNS    350
              D+      RRKN N  S AGA     AE HS  Q     ESDVS DGF WRKYGQKVVKGN
Sbjct    352  DNDKPNYNRRKNENQVSGAGAVMEGVAEPHSATQTMVESDVSGDGFHWRKYGQKVVKGNL    411

Query    351  YPRCYYRCTSPKCSVRKYVERASGDTRSFVTTYEGKHDHDKPEKKVDL---------MHK    401
              YPR YYRCT+P C+VRKYVE+AS D+ SFVTTYEGKH+H   P KK +L          ++
Sbjct    412  YPRSYYRCTTPNCNVRKYVEKASDDSGSFVTTYEGKHNHALPAKKTNLAASDPDAVALNN    471

Query    402  NERKLG    407
              +   L
Sbjct    472  SRHNLNT   478
```

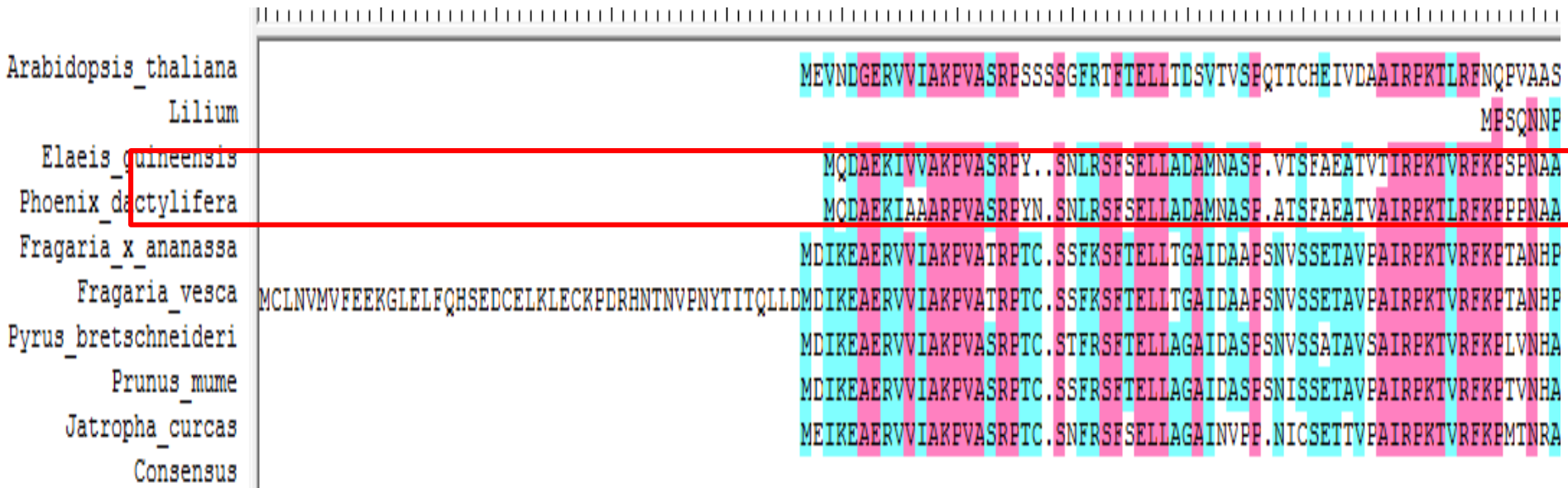Fig. 6 Align two sequences of LhWRKY44 and EgWRKY

Sequencecs Alignment analysis

Fig. 7 Multiple WRKY sequence alignment

反应问题：转录组测序拼接结果可能有误

Sequencecs Alignment analysis

NW Score	Identities	Positives	Gaps
1122	254/487(52%)	305/487(62%)	34/487(6%)

```
Query    1    MHDSEKIAVAKPVATRPYSSFRSFSDFLAGAIDASPTDSLAETTVAIRPKTVRFMSSQNN   60
              M D+EKI VAKPVA+RPYS+ RSFS+ LA A++ASP  S AE TV IRPKTVRF  S N
Sbjct    1    MQDAEKIVVAKPVASRPYSNLRSFSELLADAMNASPVTSFAEATVTIRPKTVRFKPSPNA   60

Query   61    PPAEPASSMDTVSESPI-FRSSDNLSIPTSREDSRSSVIYKPMAKVVSRTTASLLAKMGN  119
               AE  SS+D +SES + F  SD  S   S+ D ++SV+YKP A+VVSRTTAS LA +GN
Sbjct   61    ALAE-VSSLDNISESRVAFNISDKASTTISKADYKTSVLYKPAARVVSRTTASRLANLGN  119

Query  120    -EICHQQPLEEQTQFQVPDQVN-------QSQSTFDIYQNMPSDMLMNQVYEPNISESSN  171
               +  HQ P   T Q P V       Q  Q T    QN  S + MN +Y+P     SN
Sbjct  120    FDAGHQNPA---TDVQAPVHVIGQEKHQFQLQPTLSCQQNSFSHIDMNHIYKP-----SN  171

Query  172    LMAPNTDLDNIR------DRPSYDGYNWRKYGQKQVKGSEYPRSYYKCTHPICPVKRMVE  225
               +  +T+ D +         DRPSYDGYNWRKYGQKQVKGSE PRSYYKCTHP CPVK+ VE
Sbjct  172    SVPESTEQDTRKQPATGGDRPSYDGYNWRKYGQKQVKGSECPRSYYKCTHPNCPVKKKVE  231

Query  226    RSIDGKIAEIVYKGEHNHPKPQPPKRLSSASQGQTSVASEQAREIEEPLWSDLLIEKNLS  285
              RS DG+IAEIVYKGEHNHPKPQPPKRLSS SQGQT V+   +E + PL S  L+  N+S
Sbjct  232    RSFDGQIAEIVYKGEHNHPKPQPPKRLSSGSQGQTVVSEGHGKESDNPLRSHHLVWGNVS  291

Query  286    DSRMDGHIDVNFYGALDS-PKVQCSHDSLIGATYSSGIMAPDPSQRLGADLEARSKGTVV  344
              D +++  I+++         K    SHD   +Y+S   PD + RL    E +      +
Sbjct  292    DGQVEKGIEISLPDTTSFLGKAYFSHDPPFATSYNSVARNPDSAGRLSGASEFGNATASI  351

Query  345    DDGLHRSKRRKNLNLDSEAGASRNITAELHSVFQAPTESDVSADGFRWRKYGQKVVKGNS  404
              D+      RRKN N  S AGA     AE HS  Q   ESDVS DGF WRKYGQKVVKGN
Sbjct  352    DNDKPNYNRRKNENQVSGAGAVMEGVAEPHSATQTMVESDVSGDGFHWRKYGQKVVKGNL  411

Query  405    YPRCYYRCTSPKCSVRKYVERASGDTRSFVTTYEGKHDHDKPEKKVDL---------MHK  455
              YPR YYRCT+P C+VRKYVE+AS D+ SFVTTYEGKH+H  P KK +L         ++
Sbjct  412    YPRSYYRCTTPNCNVRKYVEKASDDSGSFVTTYEGKHNHALPAKKTNLAASDPDAVALNN  471

Query  456    NERKLG     461
              +  L
Sbjct  472    SRHNLNT    478
```

Fig. 8 Align two sequences of LhWRKY44 and EgWRKY
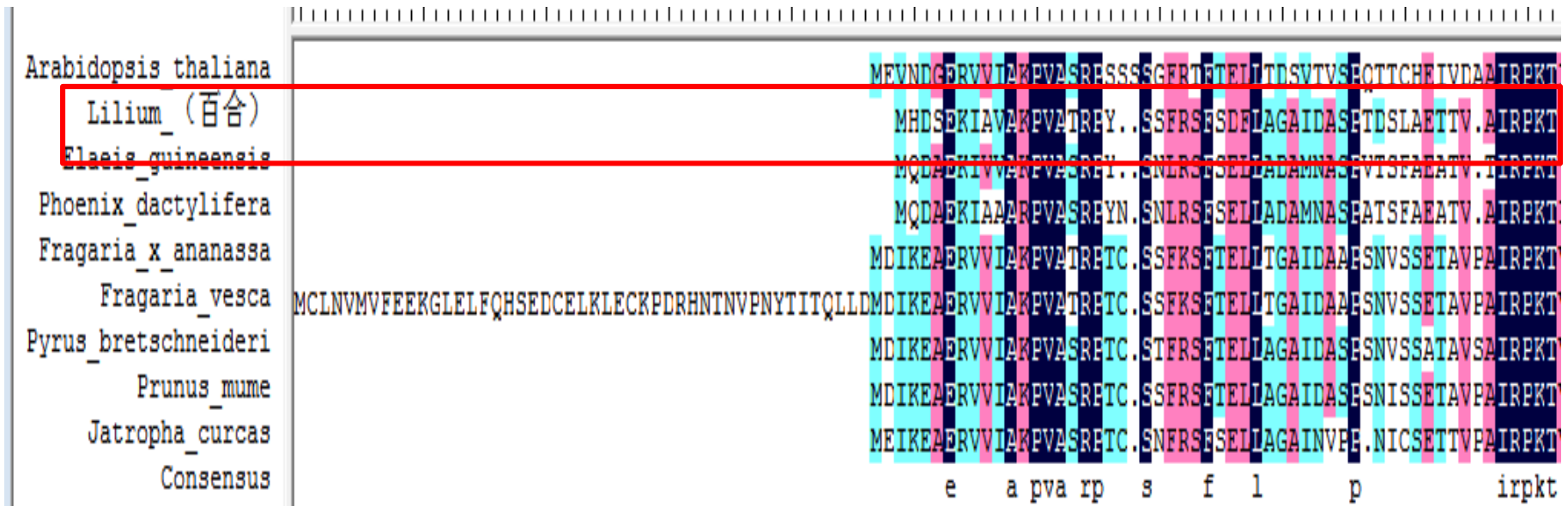
Sequencecs Alignment analysis

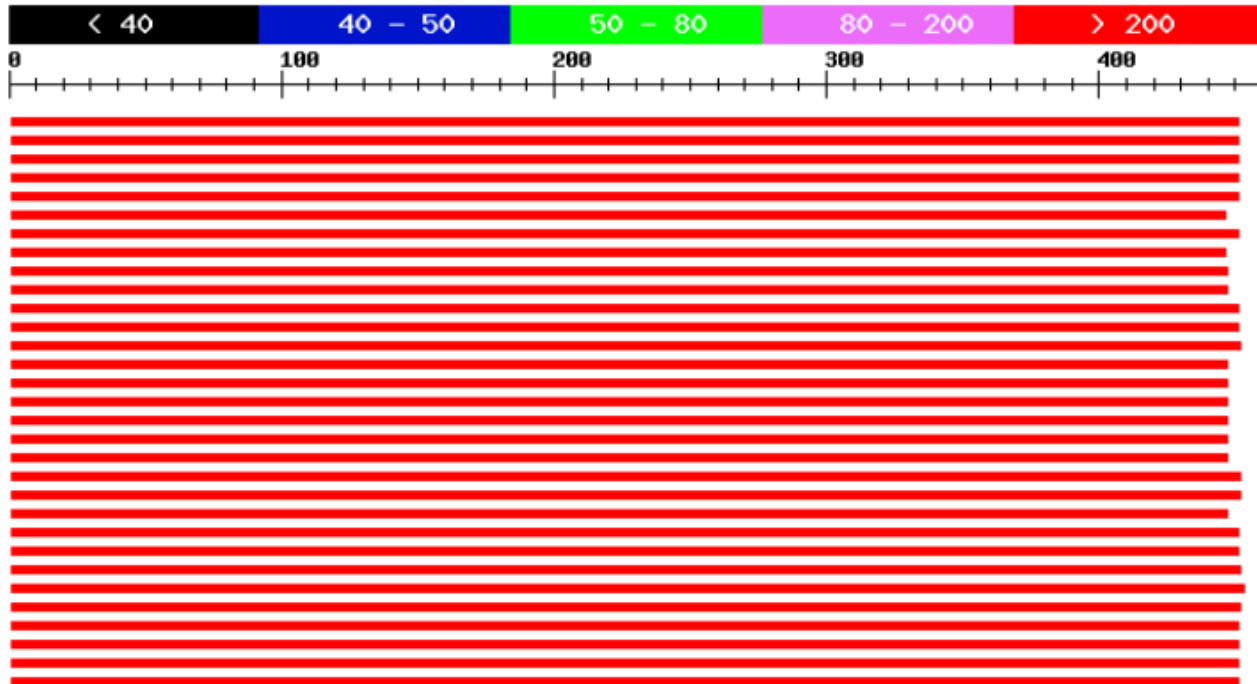Fig. 9 Multiple WRKY sequence alignment

解决办法：重新拼接转录组结果或Race

Sequencecs Alignment analysis

Sequence Analysis

➢ NCBI ORF Finder（https://www.ncbi.nlm.nih.gov/orffinder/）

LhWRKY44 contains a 1386-bp open reading frame and encodes a protein of 461 amino acid residues

Sequence Analysis

➢ NCBI CDD（https://www.ncbi.nlm.nih.gov/cdd/?term=）

The predicted **LhWRKY44** protein contains two typical WRKY domains, which include two WRKYGQK motifs .

Sequence Analysis

Two WRKYGQK motifs distributed between amino acids 193 and 199, 392 and 398. One C2-H2 zinc finger motif, which is hallmarks of class I WRKY transcription factor.

Fig. 10 Full-length CDS and predicted amino acid sequences of LhWRKY44
the amino acids highlignted are WRKYGQK conserved domains, the amino acids underlined is C_2-H_2 zinc finger.

```
              10        20        30        40        50        60
1    ATGCATGACTCAGAAAAGATTGCTGTGGCTAAACCTGTTGCTACAAGACCTTATTCCAGT
1      M  H  D  S  E  K  I  A  V  A  K  P  V  A  T  R  P  Y  S  S
              70        80        90       100       110       120
61   TTCAGGTCCTTCTCAGATTTCCTGGCAGGTGCTATTGATGCTTCTCCTACAGATTCTCTT
21      F  R  S  F  S  D  F  L  A  G  A  I  D  A  S  P  T  D  S  L
             130       140       150       160       170       180
121  GCCGAGACGACAGTAGCCATCAGGCCAAAGACTGTCAGGTTCATGTCCTCACAAAACAAT
41      A  E  T  T  V  A  I  R  P  K  T  V  R  F  M  S  S  Q  N  N
             190       200       210       220       230       240
181  CCCCCAGCTGAGCCAGCTTCATCCATGGATACTGTATCTGAATCGCCTATTTTTCGTTCA
61      P  P  A  E  P  A  S  S  M  D  T  V  S  E  S  P  I  F  R  S
             250       260       270       280       290       300
241  TCTGATAATCTATCTATCCCTACTTCAAGGGAAGACTCCAGATCAAGTGTGATATACAAA
81      S  D  N  L  S  I  P  T  S  R  E  D  S  R  S  S  V  I  Y  K
             310       320       330       340       350       360
301  CCTATGGCAAAAGTTGTTTCAAGGACAACTGCCTCTCTCTTGGCCAAGATGGGCAATGAA
101     P  M  A  K  V  V  S  R  T  T  A  S  L  L  A  K  M  G  N  E
             370       380       390       400       410       420
361  ATATGTCATCAACAACCATTAGAGGAGCAGACTCAGTTTCAGGTTCCTGACCAAGTAAAC
121     I  C  H  Q  Q  P  L  E  E  Q  T  Q  F  Q  V  P  D  Q  V  N
             430       440       450       460       470       480
421  CAAAGCCAGTCTACTTTCGATATTTATCAGAATATGCCATCAGACATGTTGATGAACCAA
141     Q  S  Q  S  T  F  D  I  Y  Q  N  M  P  S  D  M  L  M  N  Q
             490       500       510       520       530       540
481  GTATATGAGCCTAATATAAGTGAGTCTTCAAATTTGATGGCTCCAAATACAGATCTAGAC
161     V  Y  E  P  N  I  S  E  S  S  N  L  M  A  P  N  T  D  L  D
             550       560       570       580       590       600
541  AATATAAGAGATCGGCCTTCGTACGATGGATATAATTGGAGAAAGTACGGGCAAAAGCAA
181     N  I  R  D  R  P  S  Y  D  G  Y  N  W  R  K  Y  G  Q  K  Q
             610       620       630       640       650       660
601  GTAAAAGGAAGTGAATATCCACGAAGTTACTATAAATGCACGCATCCAATTTGTCCTGTG
201     V  K  G  S  E  Y  P  R  S  Y  Y  K  C  T  H  P  I  C  P  V
             670       680       690       700       710       720
661  AAGAGGATGGTTGAGAGATCAATTGATGGGAAAATAGCTGAAATTGTCTACAAGGGTGAA
221     K  R  M  V  E  R  S  I  D  G  K  I  A  E  I  V  Y  K  G  E
             730       740       750       760       770       780
721  CACAACCACCCTAAGCCTCAGCCTCCTAAGCGTTTATCGTCAGCTTCACAAGGACAAACT
241     H  N  H  P  K  P  Q  P  P  K  R  L  S  S  A  S  Q  G  Q  T
             790       800       810       820       830       840
781  TCTGTTGCTAGTGAGCAAGCGAGAGAGATTGAGGAACCCTTGTGGAGTGATCTTCTCATT
```

Sequence Analysis

Two WRKYGQK motifs distributed between amino acids 193 and 199, 392 and 398. One C2-H2 zinc finger motif, which is hallmarks of class I WRKY transcription factor.

Fig. 11 Full-length CDS and predicted amino acid sequences of LhWRKY44 the amino acids highlignted are WRKYGQK conserved domains, the amino acids underlined is C_2-H_2 zinc finger.

```
  S  V  A  S  E  Q  A  R  E  I  E  E  P  L  W  S  D  L  L  I
     850      860      870      880      890      900
GAGAAGAACTTGTCTGATAGTAGAATGGATGGCCACATTGATGTTAATTTTTATGGTGCT
  E  K  N  L  S  D  S  R  M  D  G  H  I  D  V  N  F  Y  G  A
     910      920      930      940      950      960
CTAGATTCACCAAAGGTGCAATGCTCTCATGACTCACTAATAGGTGCAACATATTCTAGC
  L  D  S  P  K  V  Q  C  S  H  D  S  L  I  G  A  T  Y  S  S
     970      980      990     1000     1010     1020
GGGATCATGGCTCCTGATCCTTCTCAAAGACTTGGCGCGGACTTAGAAGCACGAAGCAAA
  G  I  M  A  P  D  P  S  Q  R  L  G  A  D  L  E  A  R  S  K
    1030     1040     1050     1060     1070     1080
GGAACTGTTGTCGATGATGGCTTGCATAGGAGTAAGCGAAGGAAAAACCTTAATCTAGAT
  G  T  V  V  D  D  G  L  H  R  S  K  R  R  K  N  L  N  L  D
    1090     1100     1110     1120     1130     1140
AGTGAAGCAGGTGCATCAAGGAACATTACTGCAGAGCTACATTCTGTATTTCAAGCTCCG
  S  E  A  G  A  S  R  N  I  T  A  E  L  H  S  V  F  Q  A  P
    1150     1160     1170     1180     1190     1200
ACAGAATCTGATGTCTCAGCAGATGGTTTCCGTTGGCGAAAATATGGACAGAAGGTTGTG
  T  E  S  D  V  S  A  D  G  F  R  W  R  K  Y  G  Q  K  V  V
    1210     1220     1230     1240     1250     1260
AAGGGAAATTCATATCCAAGGTGTTACTACAGATGCACGAGTCCTAAATGTAGCGTTAGG
  K  G  N  S  Y  P  R  C  Y  Y  R  C  T  S  P  K  C  S  V  R
    1270     1280     1290     1300     1310     1320
AAATATGTCGAAAGGGCATCTGGCGATACAAGATCTTTCGTTACAACTTACGAAGGAAAG
  K  Y  V  E  R  A  S  G  D  T  R  S  F  V  T  T  Y  E  G  K
    1330     1340     1350     1360     1370     1380
CACGACCATGATAAGCCAGAAAAGAAAGTCGATTTGATGCACAAGAACGAGAGAAAGCTT
  H  D  H  D  K  P  E  K  K  V  D  L  M  H  K  N  E  R  K  L

GGTTAA
  G  *
```

Phylogenetic tree

➢ PlantTFDB(http://planttfdb.gao-lab.org/index.php)

Browse by Family

AP2 (4461)	ARF (4578)	ARR-B (2354)	B3 (10609)	BBR-BPC (1256)
BES1 (1549)	C2H2 (17740)	C3H (9693)	CAMTA (1343)	CO-like (2125)
CPP (1612)	DBB (1651)	Dof (5655)	E2F/DP (1781)	EIL (1234)
ERF (21129)	FAR1 (7527)	G2-like (9874)	GATA (5335)	GRAS (9304)
GRF (1876)	GeBP (1564)	HB-PHD (477)	HB-other (2277)	HD-ZIP (8602)
HRT-like (249)	HSF (4574)	LBD (7216)	LFY (253)	LSD (957)
M-type_MADS (7541)	MIKC_MADS (6918)	MYB (22032)	MYB_related (15369)	NAC (19997)
NF-X1 (403)	NF-YA (2461)	NF-YB (3099)	NF-YC (2446)	NZZ/SPL (109)
Nin-like (2766)	RAV (690)	S1Fa-like (359)	SAP (164)	SBP (4168)
SRS (1327)	STAT (214)	TALE (4433)	TCP (4187)	Trihelix (6256)
VOZ (635)	WOX (2358)	WRKY (14549)	Whirly (530)	YABBY (1719)
ZF-HD (2589)	bHLH (28698)	bZIP (15498)		

➢ TAIR(https://www.arabidopsis.org/index.jsp)

Phylogenetic tree

Distribution of WRKY family in different species

(G)-species with genome sequence

Species	Count
Actinidia chinensis (G)	106
Aegilops tauschii (G)	92
Aethionema arabicum (G)	58
Amaranthus hypochondriacus (G)	54
Amborella trichopoda (G)	32
Ananas comosus (G)	56
Aquilegia coerulea (G)	56
Arabidopsis halleri (G)	60
Arabidopsis lyrata (G)	79
Arabidopsis thaliana (G)	90
Arabis alpina (G)	58
Arachis duranensis (G)	87
Arachis hypogaea	28
Arachis ipaensis (G)	95
Artemisia annua	36

WRKY Family Introduction

WRKY transcription factors are one of the largest families of transcriptional regulators in plants and form integral parts of signalling webs that modulate many plant processes. Here, we review recent significant progress in WRKY transcription factor research. New findings illustrate that WRKY proteins often act as repressors as well as activators,

Phylogenetic tree

Arabidopsis thaliana WRKY Family

- WRKY Family Introduction
- Download Sequences
- Multiple Sequences Alignment
- Phylogenetic Tree

Species	TF ID	Description
	AT1G13960.1	WRKY DNA-binding protein 4
	AT1G13960.2	WRKY DNA-binding protein 4
	AT1G18860.1	WRKY DNA-binding protein 61

Transcription Factor Information

Basic Information | Signature Domain | Sequence | Protein Features | 3D Structure | Gene Ontology | Plant Ontology | Expression | Function | Cis-element | Regulation | Interaction | Phenotype | Orthologous Group | Publication

Basic Information	*? help*	*Back to Top*
TF ID	AT2G37260.1	
Common Name	ATWRKY44, DSL1, F3G5.5, TTG2, WRKY44	
Organism	*Arabidopsis thaliana*	
Taxonomic ID	3702	

Phylogenetic tree

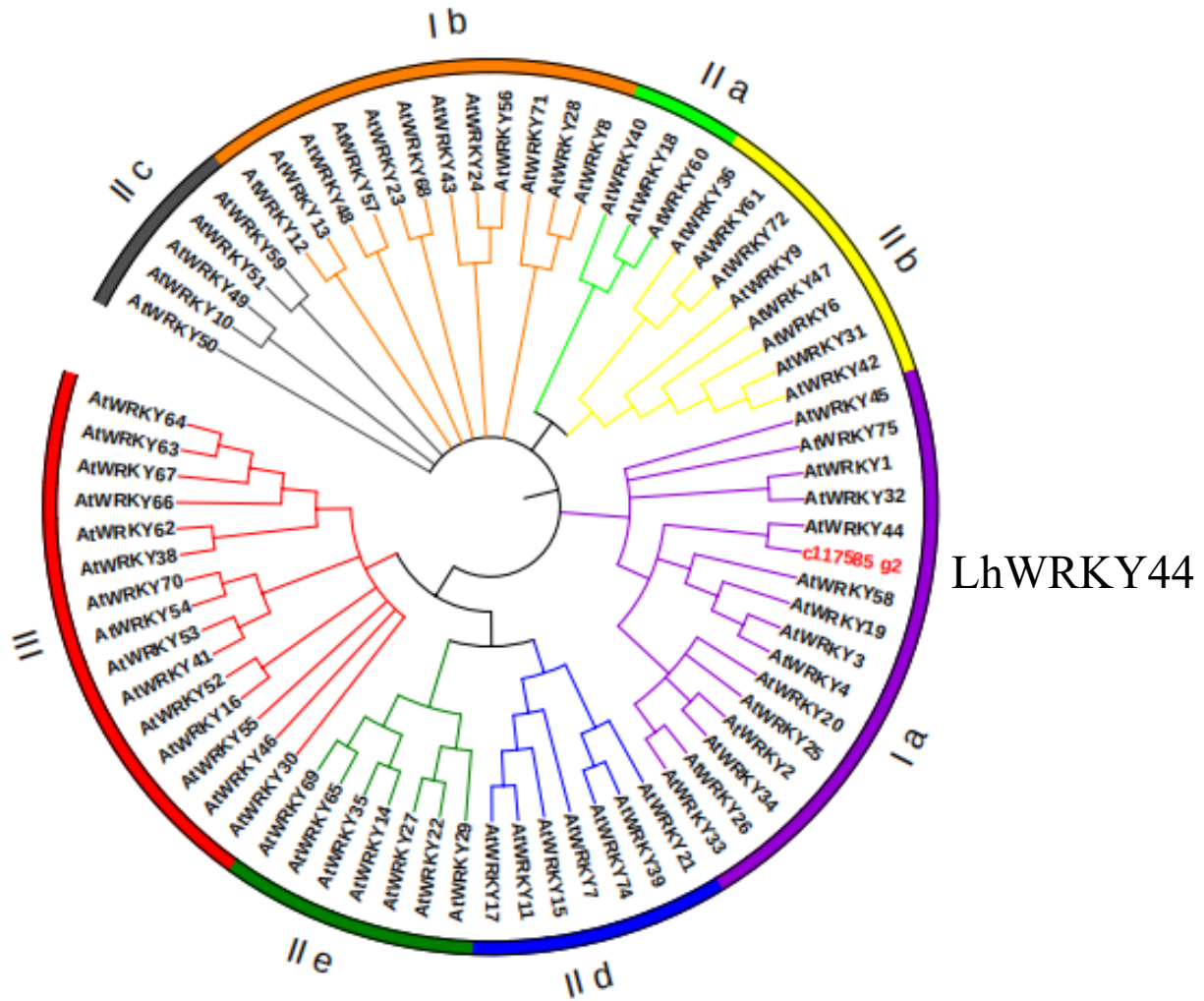➢ iTOL(https://itol.embl.de/itol.cgi)
➢ MEGA

Fig. 12 Phylogenetic tree analysis of LhWRKY44

Phylogenetic **tree**
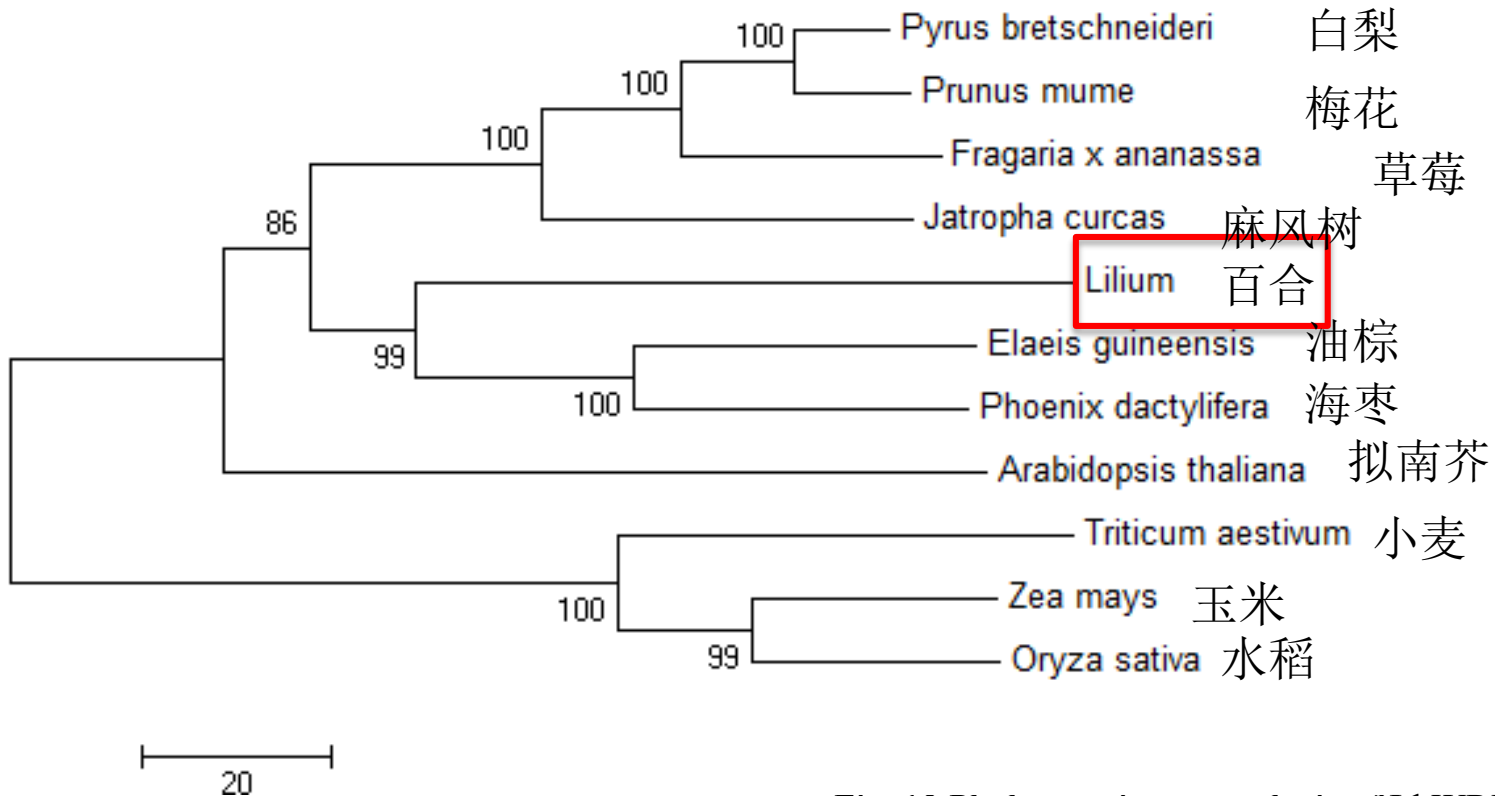
Fig. 13 Phylogenetic tree analysis of LhWRKY44
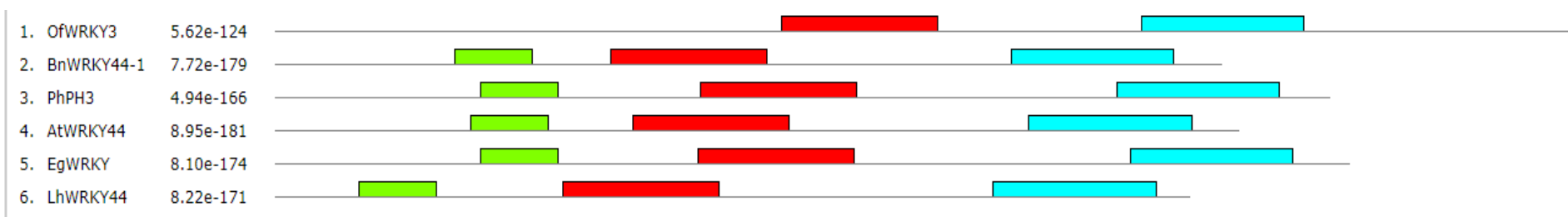
Subcellular localization

➤ PSORT（https://psort.hgc.jp/）
➤ Lhwrky44 is most likely located in the nucleus.

Table 1 Subcellar location prediction of LhWRKY44

Gene expression position in cells	Rate (%)
Nucleus	88.1029%
Vacular membane	11.8971%

Motifs Prediction

MEME(http://meme-suite.org/tools/meme)

Primer Design

Primer Design - The online primer design tool at BioWeb.

Primer Blast - NCBI primer design tool with Blast search.

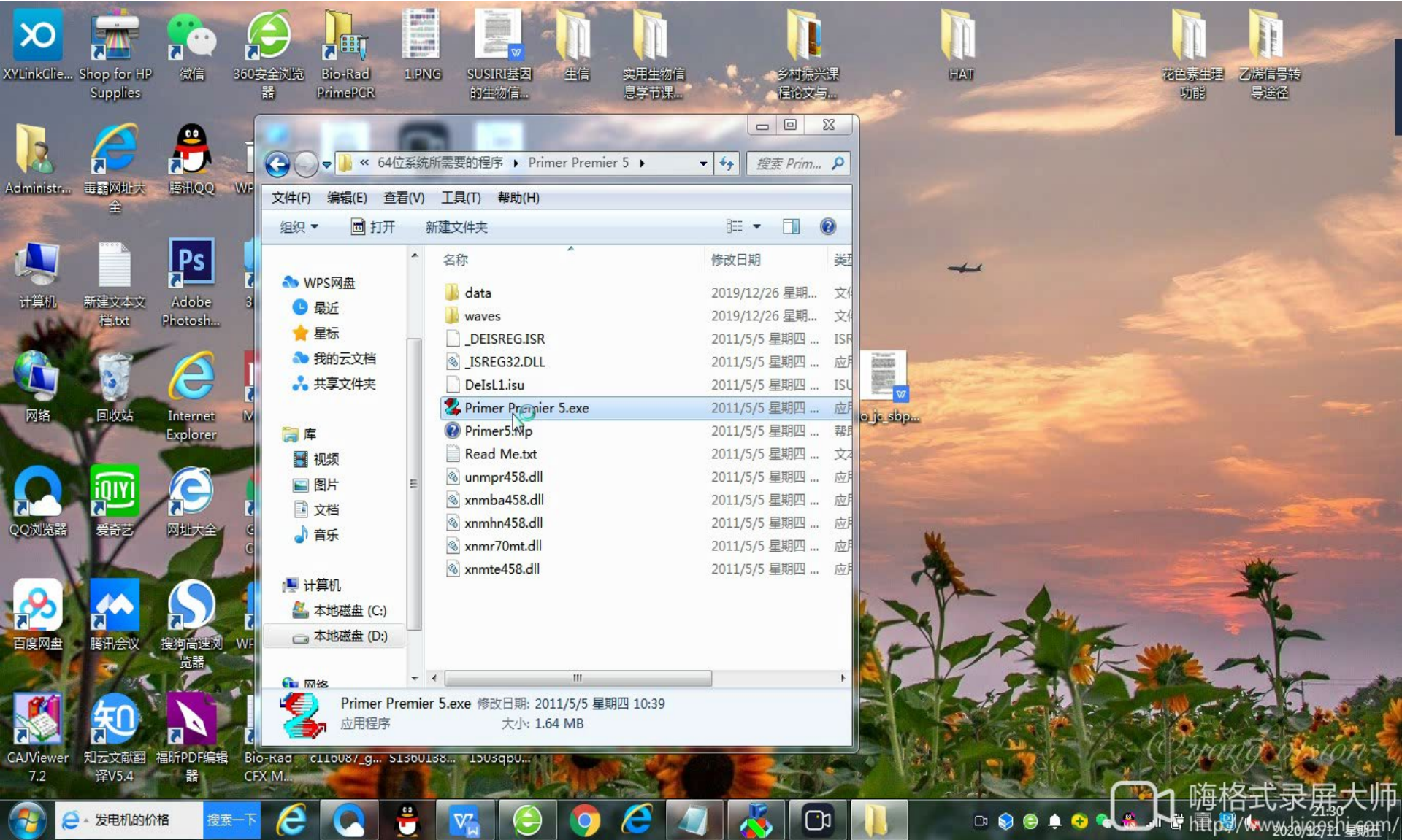**Web Primer** - Online primer design tool at the yeast genome database.

Snapgene

Primer Premier 5.0

引物设计原则

1. 长度一般为15-30bp，常用的为18-25bp。
2. GC含量为40-60%，以45-55%为宜，上下游引物GC含量和Tm值要保持接近。
3. 引物与模板结合自由能（ΔG值）相对要低。
4. 错配率一般不要超过100，否则会出现非目的条带。
5. 避免反向重复和自身互补序列。
6. 二聚体及发卡结构能量绝对值一般不要超过4.5，否则容易产生二聚体而且会降低引物浓度，导致PCR不能正常进行。
7. 3'端连续碱基GGG或CCC会导致错误引发；3'端最后一个碱基A或T会导致错配。

Primer Design

参考文献

1. Xu L, Yang P, Feng Y, et al. Spatiotemporal Transcriptome Analysis Provides Insights into Bicolor Tepal Development in Lilium "Tiny Padhye". Front Plant Sci. 2017.
2. Paul J, Rushton, Imre E, et al. WRKY transcription factors. Trends in Plant Science. 2010.

*Thanks to the photos provider Pro.Wang in 11th slide .

Thank you for your attention!