

# **LEB22 MYBDB project demo**

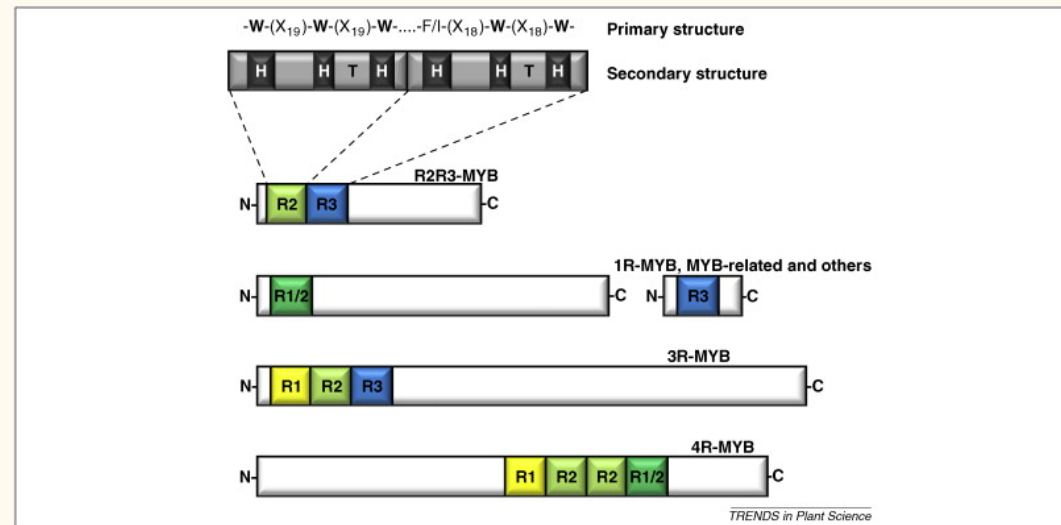
Yung-Cheng Chiang, LEB

2022.06

## MYB Family Introduction

MYB factors represent a family of proteins that include the conserved MYB DNA-binding domain. The first MYB gene identified was the "oncogene" v-Myb derived from the avian myeloblastosis virus. Evidence obtained from sequence comparisons indicates that v-Myb may have originated from a vertebrate gene, which mutated once it became part of the virus. Many vertebrates contain three genes related to v-Myb c-Myb, A-Myb and B-Myb and other similar genes have been identified in insects, plants, fungi and slime moulds. The encoded proteins are crucial to the control of proliferation and differentiation in a number of cell types, and share the conserved MYB DNA-binding domain. This domain generally comprises up to three imperfect repeats, each forming a helix-turn-helix structure of about 53 amino acids. Three regularly spaced tryptophan residues, which form a tryptophan cluster in the three-dimensional helix-turn-helix structure, are characteristic of a MYB repeat. The three repeats in c-Myb are referred to as R1, R2 and R3; and repeats from other MYB proteins are categorised according to their similarity to either R1, R2 or R3. MYB proteins can be classified into three subfamilies depending on the number of adjacent repeats in the MYB domain (one, two or three). We refer to MYB-like proteins with one repeat as "MYB1R factors", with two as "R2R3-type MYB" factors, and with three repeats as "MYB3R" factors.

In contrast to animals, plants contain a MYB-protein subfamily that is characterised by the R2R3-type MYB domain, which is a unique MYB transcription factor in plants and closely related to anthocyanin synthesis and regulation.



# LEB22 Project MYBDB

[Home](#) | [About](#) | [Download](#) | [Database](#) | [Search](#)

## About

### Group List

We have a strong team behind the MYBDB project!

No	Name	Gender
0001	Yung-Cheng Chiang	M
0002	Danni Fan	F
0003	Zeyang Qu	M
0004	Yunjia Lu	M
0005	Haobo Zhang	M
0006	Chunyang Jiang	F

### Sequence alignment and data analysis

- Selected *Malus x Domestica* GDDH13 v1.1 Whole Genome Assembly as reference.
- Performed HMMER search using MYB\_DNA\_binding (PF00249) model. Acquired 475 alignment results.
- Performed PSI-BLAST using MYB\_ARATH reviewed sequences. Acquired 175 alignment results.
- Considered 122 overlapping alignments as verified.

### Database construction and visualization

- Basic information: Constructed mySQL table and then extracted information through php-mysql connection.
- Alignment results: Manipulated text files directly through php scripts.
- Real-time MYB pages coding: Passed variables through php url.

---

Last modified: 22 June, 2022. The database was constructed and maintained by the **LEB MYBDB team**. One could refer to **SBP webpage** or **PlantTFDB MD MYB webpage** for detailed source code.

## *Malus x domestica* MYB and MYB\_related Family

TF ID	Domain number	Description
MD14G1233900	5	Myb domain protein 4r1
MD04G1037700	4	Myb family transcription factor
MD12G1141900	3	Myb domain protein 3r-5
MD04G1129200	3	Myb domain protein 3r-5
MD08G1117700	3	Homeodomain-like protein
MD15G1097000	3	Homeodomain-like protein
MD08G1117600	3	Homeodomain-like protein
MD04G1064500	3	Myb domain protein 3r-4
MD16G1218900	3	Myb domain protein 15
MD06G1172900	2	Myb domain protein 98
MD05G1089600	2	Myb domain protein r1
MD09G1183800	2	Myb domain protein 66
MD02G1179000	2	Myb domain protein 73
MD00G1092300	2	High response to osmotic stress 10
MD15G1288600	2	Myb domain protein 73
MD15G1323500	2	Myb domain protein 93
MD15G1076600	2	Myb domain protein 73
MD08G1092000	2	Myb domain protein 73
MD14G1179000	2	Myb domain protein 98

## MD15G1076600

Basic Information <a href="#">? help</a>		Back to Top					
TF ID	MD15G1076600						
PSI-BLAST	Unverified						
Organism	<i>Malus x domestica</i>						
Taxonomic ID	3750						
Family	MYB						
Protein Properties	Length: 317 aa MW: 34446.89 Da PI: 7.7243						
Description	Myb domain protein 73						
Gene Model	Gene Model ID	Type	Source	Coding Sequence			
	MD15G1076600	genome	GDR	View CDS			
Signature Domain <a href="#">? help</a>		Back to Top					
No.	Domain	Score	E-value	Start	End	HMM Start	HMM End
1	PF00249	60.2	4.1e-18	12	58	1	46
<pre> SSSBHHHHHHHHHHHHHT.TCHHHHHHHHTSSHHHHHHHHHHHT CS PF00249 1 rgkwteeEdeLLleavkklG.nrwkIaerlggRtekqcrsrwnyl 46 +g+W++eEde L ++vk +G +W++I++ ++gR++k+cr+rw n l MD15G1076600 12 KGPWSPEEDEALQNLVKIYGpRNWLSLISKSIIPGRSGKSCRLRWCNQL 58 79*****999*****9975 PP </pre>							
2	PF00249	55.4	1.4e-16	66	108	3	45
<pre> SSBHHHHHHHHHHHHHTTCHHHHHHHHTSSHHHHHHHHHHHT CS PF00249 3 kwteeEdeLLleavkklGnrwkIaerlggRtekqcrsrwnyl 45 ++++eEd+ L++a++++Gn+W++Ia++lgRt+++++w++ MD15G1076600 66 PFSPEEDDTLIRAHARFGNKWATIARLLNGRDTNAIKNHWNST 108 89*****999*****9975 PP </pre>							
Sequence <a href="#">? help</a>		Back to Top					
Protein Sequence Length: 317 aa		<a href="#">Download sequence</a>					
<pre> MASTKKVVDRIKGPWSPEEDEALQNLVKIYGPRNWSLISKSIIPGRSGKSCRLRWCNQLSP 60 EVEHRPFSPEEDDTLIRAHARFGNKWATIARLLNGRDTNAIKNHWNSTLKRKCSSMSDEL 120 CSDVHDHPPHKRSASVAVSGTGLYFNPGSPSGDLSDSLHGGVSPSSQVFTPIARVPV 180 PPIAPPMEAAATSAAVDPPTSLSLSLPGSESESCDGSNHMGSGFGSNPIVGPQMVVEQP 240 HEVAAPPVVGLLPPRQSNLNNQFFSFEFLDVMQEMIRKEVRNYMTGIEQKGCQMGT 300 EAIRNAVIKRIGISKIE </pre>							

header

TFs basic  
information

Hmsearch results

Protein sequence

foot

# Helping Pages

## 1. TFs information



**Transcription Factor Information**

**TF ID**  
We chose *Malus x Domestica* genome GDDH13 v1.1 as our target. TF ID here is identical to gene model ID.

**Taxonomy**  
The taxonomic ID and lineage for each organism was collected from NCBI [Taxonomy](#).

**Gene Model**  
The gene (data source) coding for this transcription factor.

**Gene Model ID**  
The ID of gene model, which was extracted from the original data source. Gene model ID can be searched in advanced search page.

**Gene Model Type**  
The type of gene model. MYBDB gene models came from genome.

**Source**  
The source where gene model was got

**Signature Domain**  
The Domain used to identify and classify transcription factors.

## 2. TFs searching

**Search**

Search by

Search by  
ID  
Domain number  
Description

...e 2022, by [Group 3+7](#). One could refer to [S](#)

# **Supplementary**

Method and Technical Details

# Sequence Alignment and Data Analysis

## HMMER search

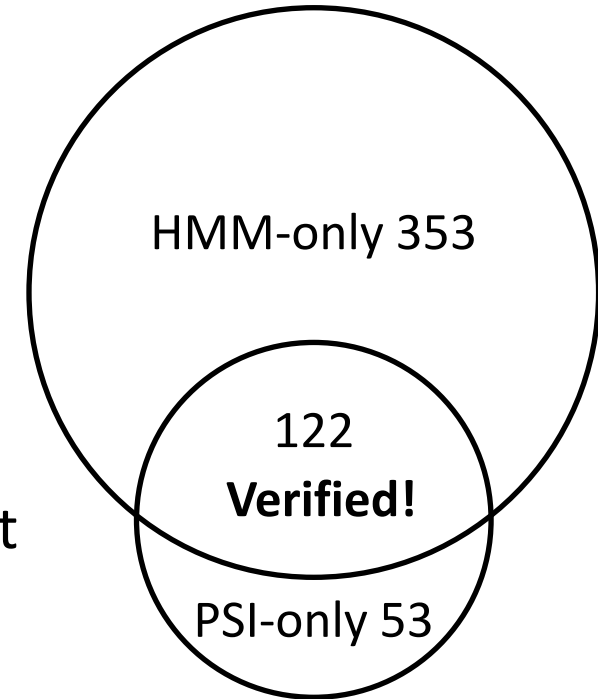
Database: *Malus x domestica* GDDH13 v1.1

HMM Profile: Pfam PF00249

## PSI-BLAST

Database: *Malus x domestica* UniProtKB/Swiss-Prot

Query: *Arabidopsis thaliana* MYB sequences





# Webpage Design

PHP (后端, 编程语言)  
HTML (前端, 标签语言)

```
(base) leb3a@bbt:/rd1/www/group3/myb_tfdb$ ll
```

```
total 64
```

```
drwxrwxr-x  7 leb3a leb 4096 Jun 21 10:55 ./
drwxrwxr-x 11 root  leb 4096 Jun 21 10:55 ../
-rwxrwxr-x  1 leb3a leb 1362 Jun 18 22:45 about.php*
drwxr-xr-x  3 leb3a leb 4096 Jun 21 10:55 archive/
drwxr-xr-x  5 leb3a leb 4096 Jun 17 18:52 data/
-rwxrwxr-x  1 leb3a leb  858 Jun 19 21:54 data_db.php*
-rwxrwxr-x  1 leb3a leb  527 Jun 20 10:26 download.php*
-rwxrwxr-x  1 leb3a leb  292 Jun 19 21:09 foot.html*
-rwxrwxr-x  1 leb3a leb  223 Jun 19 21:08 head.html*
-rwxrwxr-x  1 leb3a leb 2210 Jun 17 20:06 help_info.php*
drwxr-xr-x  2 leb3a leb 4096 Jun 17 18:52 images/
-rwxrwxr-x  1 leb3a leb 2450 Jun 19 15:10 index.php*
drwxr-xr-x  2 leb3a leb 4096 Jun 19 19:55 scripts/
-rw-r--r--  1 leb3a leb 3285 Jun 20 11:35 search.php
drwxr-xr-x  2 leb3a leb 4096 Jun 17 18:52 styles/
-rwxrwxr-x  1 leb3a leb  622 Jun 17 19:52 test_db.php*
```

- about.php: 项目信息
- archive/: 旧代码
- data/: 序列数据和比对结果
- data\_db.php: 数据库子页
- head/foot.html: 页首页尾
- index.php: 主页, Myb基本信息
- scripts/: 可运行脚本
- styles/: 网页样式
- test\_db.php: 数据库页

# Passing variables to multiple pages

```
echo
"<tr>
<td align=center><a href=../data_db.php?ID={$ID}>$ID</a></td>
<td align=center>$DN</td>
<td align=left>$DP</td>
</tr>";
```

```
<?php
    $ID = $_GET['ID'];
    echo "<h2>$ID</h2>";
    include './scripts/show_myb_basic_info.php';
    include './scripts/show_myb_sig_domain.php';
    include './scripts/show_myb_prot_seq.php';
?>
```

```
<?php
    $fname = $_GET['ID'];
    //echo "$fname";
    //build a connection
    //echo '<p>Connecting to MYB mySQL database...';
    $dbhost = 'localhost';           // MySQL server
    $dbuser = 'leb3a';               // MySQL user
    $dbpass = 'leb2022';
    $conn = mysqli_connect($dbhost, $dbuser, $dbpass);
```

Variable passing  
[http://group3.pku.edu.cn/..../data\\_db.php?ID={\\$ID}](http://group3.pku.edu.cn/..../data_db.php?ID={$ID})



*One code, multiple pages*

[http://group3.pku.edu.cn/.../data\\_db.php?ID={\\$MD15G1076600}](http://group3.pku.edu.cn/.../data_db.php?ID={$MD15G1076600})

[http://group3.pku.edu.cn/.../data\\_db.php?ID={\\$MD14G1233900}](http://group3.pku.edu.cn/.../data_db.php?ID={$MD14G1233900})

[http://group3.pku.edu.cn/.../data\\_db.php?ID={\\$MD08G1117700}](http://group3.pku.edu.cn/.../data_db.php?ID={$MD08G1117700})

*...475 php pages, right on click!*

# Drawing HMM Alignment Results Using PHP

```
echo
"  <tr>
    <td colspan='8'>
      <img src='./scripts/draw_hmm_fig.php?AAL={AAL}&num_dom={num_dom}&HMM_result=" . urlencode(serialize($
    </td>
  </tr>
"

//img x1 y1 x2 y2 color
for ($x = 1; $x <= $num_dom; $x++) {
  imageline($img, $left_bound, 25+25*$x, $right_bound, 25+25*$x, $black);
  imagefilledrectangle($img, $left_bound+($right_bound-$left_bound)*$s[$x-1]/$AL, 25+25*$x-8, $left_bound+($right_bound-
  imagestring($img, 4, $left_bound+($right_bound-$left_bound)*($s[$x-1]+$e[$x-1])/(2*$AL)-2, 25+25*$x-7, $x, $black);
  imagestring($img, 8, $right_bound+25, 25+25*$x-7, 'PF00249', $black);
  //imagestring($img, 8, 675, 25+25*$x-7, $e[$x-1], $black);
}
header("Content-type: image/png");
imagepng($img);
```

