转录因子数据库构建实例

PlantRegMap/PlantTFDB v5.0

Plant Transcription Factor Database

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

Note:

- The new version PlantTFDB 5.0 has been incorporated into PlantRegMap.

Browse by Species

open all | close all

Taxonomic Group (165 species) (G)-species with genome sequence

- Chlorophytae (16 species)
- Charophyta (1 species)
- Marchantiophyta (1 species)
- Bryophyta (2 species)
- Lycopodiophyta (1 species)
- Coniferophyta (5 species)
- Basal Magnoliophyta (1 species)
- Monocots (38 species)
- Eudicots (100 species)

Quick links: Arabidopsis thaliana, Glycine max, Oryza sativa, Populus trichocarpa, Solanum lycopersicum, Zea mays

Couldn't find the species interested? Please try the extended TF repertoires or identify TFs by the prediction server.

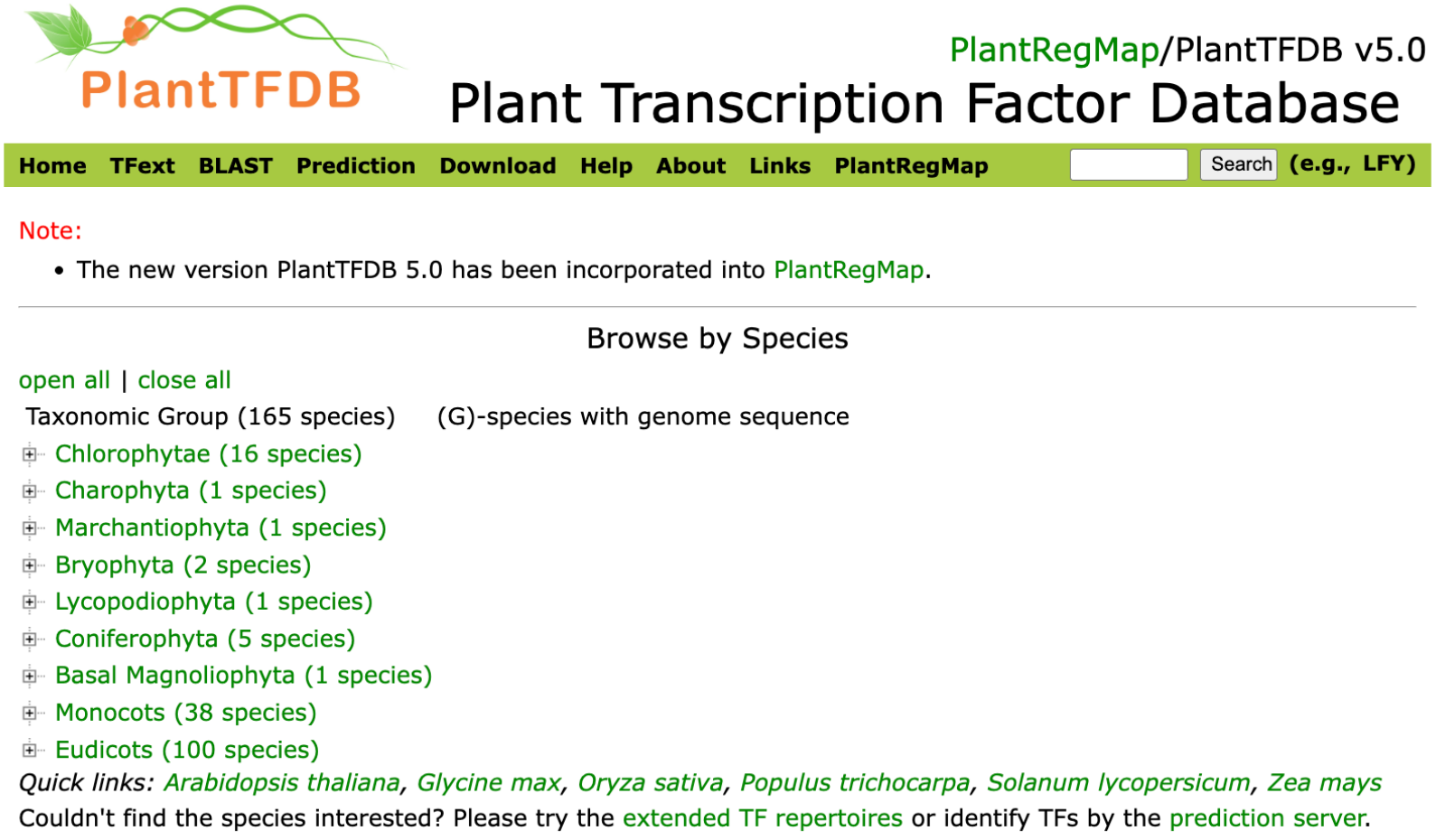
Browse by Family

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)		

https://planttfdb.gao-lab.org/

网站首页:
index.html 或 index.php

功能:
浏览、查询和分析服务



静态网页

HTML文件在服务器上已存在，
收到请求时直接调用

planttfdb.gao-lab.org/tf.php?sp=Ath&did=AT5G61850.1

PlantTFDB

PlantRegMap/PlantTFDB v5.0

Plant Transcription Factor Database

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

Transcription Factor Information

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

Basic Information	***? help***			***Back to Top***
TF ID	AT5G61850.1			
Common Name	LFY, LFY3, MAC9.18			
Organism	*Arabidopsis thaliana*			
Taxonomic ID	3702			
Taxonomic Lineage	cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis			
Family	LFY			
Protein Properties	Length: 420aa MW: 46582.3 Da PI: 7.1121			
Description	floral meristem identity control protein LEAFY (LFY)			
Gene Model	***Gene Model ID***	***Type***	***Source***	***Coding Sequence***
	AT5G61850.1	genome	TAIR	View CDS

URL: https://planttfdb.gao-lab.org/tf.php?sp=Ath&did=AT5G61850.1

动态网页

HTML文件在服务器上并不存在，收到请求时通过php查询数据库生成相应的html文件

用户交互：表单(form)

```
<form action="search_result.php" method="post">
 <input type="text" name="input" id="input"><br><br>
 <input type="submit" value="Submit">
</form>
```

练习：https://www.w3schools.com/Html/tryit.asp?filename=tryhtml_form_text

相关介绍：https://www.w3schools.com/Html/html_forms.asp

参数传递与接收

- **GET** 适合传输不敏感、少量数据（如搜索查询），且该数据可公开。
- **POST** 适合传输大量数据和敏感信息（如用户登录、提交表单等）。

$id = $_GET[‘TF_ID’];

$id = $_POST[’ input’];

https://planttfdb.gao-lab.org/tf.php?sp=Ath&did=AT5G61850.1

```
<form action="search_result.php" method="post">
  <input type="text" name="input" id="input"><br><br>
  <input type="submit" value="Submit">
</form>
```

创建数据库连接

```php
<?php

// 参数
$host = ‘localhost‘;                    //数据库服务器IP
$user = ‘username’;                     //用户名
$passwd = ‘password’;                   //密码
$db = ‘database’;                       //数据库名称

// 创建数据库连接
$conn = mysqli_connect($host, $user, $passwd, $db);

// 数据库连接是否成功
if (!$conn) {
   die('Connection failed: ' . mysqli_connect_error());
}

// 关闭数据库连接
mysqli_close($conn);
?>
```

数据库检索

```php
<?php
$TF_ID = $_GET['TF_ID‘]; // 获取 GET 请求中的 TF_ID 参数

$sql = "SELECT * FROM Ath_TF_list WHERE TF_ID = '$TF_ID’"; // 创建 SQL 查询语句

$result = mysqli_query($conn, $sql); // 执行查询

?>
```

数据：https://planttfdb.gao-lab.org/download/TF_list/Ath_TF_list.txt.gz
表格：Ath_TF_list

TF_ID	Gene_ID	Family
AT3G25730.1	AT3G25730	RAV
AT1G68840.1	AT1G68840	RAV
AT1G68840.2	AT1G68840	RAV
……		

使用预处理语句防止SQL注入：https://www.php.net/manual/en/mysqli.prepare.php

检索结果展示

```php
<?php
// 检查查询是否失败
if (!$result) {
    die('Query error: ' . mysqli_error($conn));
}
// 检查结果是否为空
elseif (mysqli_num_rows($result) == 0) {
    die('Invalid ID');
}
// 如果查询成功并且有数据
else {
    // 输出表头
    echo '<tr><th>TF ID</th><th>Family</th></tr>';

    // 遍历结果并输出
    while ($row = mysqli_fetch_assoc($result)) {
        echo '<tr>';
        echo '<td>' . $row['TF_ID'] . '</td>';
        echo '<td>' . $row['Family'] . '</td>';
        echo '</tr>';
    }
}
?>
```

预测分析服务

输入与提交页面示例

Input protein or nucleic acid sequences in FASTA format (Max file size: 75M):

Or load it from disk: Choose File No file chosen

Link: Best hit in *Arabidopsis thaliana*

Reset Prediction

输入与提交 代码示例

```
<form action="prediction_result.php?task_id=<?php echo($task_id); ?>" method="post" enctype="multipart/form-data">
 Input protein or nucleic acid sequences in FASTA format (Max file size: 75M):<br>
 <textarea id="input_seq" name="input_seq" rows="10" cols="100"></textarea><br><br> <!-- 输入框 -->

 Or load it from disk:
 <input type="file" id="input_file" name="input_file" size="20"/><br><br> <!-- 上传文件 -->

 Link:
 <input type="checkbox" name="best1_ath" id="best1_ath"/> <!-- 复选框 -->
 <label for="best1_ath">Best hit in <i>Arabidopsis thaliana</i></label><br><br>

 <input type="reset" name="reset" value="Reset" /> <!-- 重置按钮 -->
     
 <input type="button" name="prediction" id="submit_button" value="Prediction" onclick="check()" /> <!-- 提交按钮 -->
</form>
```

创建任务号

```
//定义创建随机字符串的函数
function create_random($length) {
    $chars = 'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789'; //字符集
    $id = '';
    for ($i = 0; $i < $length; $i++) {
        $id .= $chars[rand(0, strlen($chars) - 1)]; //从字符集中随机选择字符
    }
    return $id;
}
```

```
$task_id = create_random(10);
```

任务接收与后台处理

```php
<?php
$tmp_name = $_GET['task_id'];
$tmp_path = "TF_prediction/tmp/" . $tmp_name;
$tmpf = $tmp_path . '/input_seq.fas';

if (!is_dir($tmp_path)) {
    mkdir($tmp_path); // 创建临时文件夹

    if (!empty($_FILES["input_file"]["name"])) {  // 检查是否有文件上传
        move_uploaded_file($_FILES["input_file"]["name"], $tmpf);
    } else {
        $seq = $_POST['input_seq']; // 检查是否有序列输入
        if (isset($seq[0]) && $seq[0] == ">") { // 检查序列格式
            $safe_seq = escapeshellarg($seq); // 对字符串中的特殊字符进行转义，防止命令注入
            $cmd = "echo $safe_seq > $tmpf"; // 准备执行的系统命令
            exec($cmd); // 执行命令
        }
    }
}
?>
```

展示结果文件

```php
<?php
$result_file = "results.txt"; // 结果文件

if (!file_exists($result_file)) {
    die("结果文件不存在！ ");
}

// 打开文件读取内容
$lines = file($filepath, FILE_IGNORE_NEW_LINES | FILE_SKIP_EMPTY_LINES);
if (!$lines) {
    die("无法读取文件内容！ ");
}

//遍历每一行并输出
foreach ($lines as $line) {
    $columns = explode("\t", $line); // 使用制表符分隔

    echo "<tr>";
    foreach ($columns as $col) {
        echo "<td>" . htmlspecialchars($col) . "</td>";
    }
    echo "</tr>";
}
?>
```