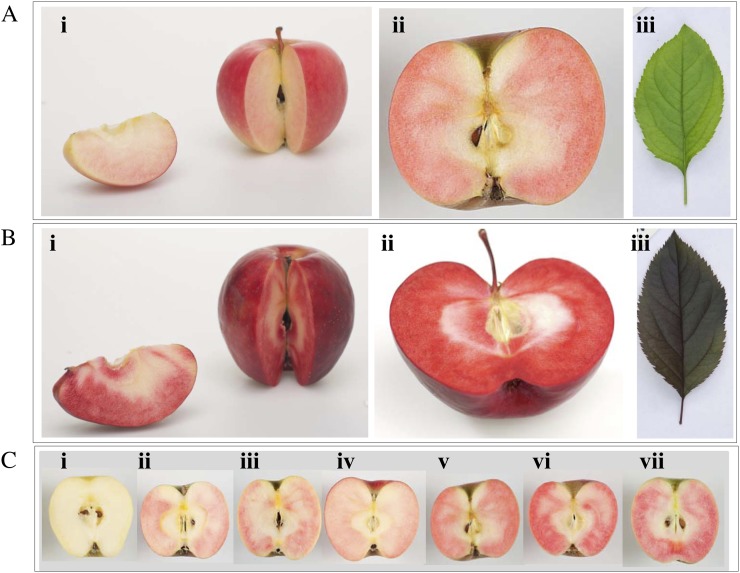
1、转录因子数据库简介

Plant TFDB是专门的植物转录因子数据库（http://planttfdb.gao-lab.org/），里面包括了来自于165个物种的转录因子数据，并分为了58个转录因子家族。在库中对确定这些TFs属于哪一个家族的分类原则是根据结构域、辅助域、禁止域三种来确定的。大多数通过结构域就可以分类，禁止域可以将含有DNA结合结构域但不具有转录活性的蛋白质剔除。关于具体分类规则可以去库中结合那个图看，或者我们讨论的时候我再说一说。

2、我的课题简介

日常生活中见到的苹果大都是红皮白色果肉的，如下图这样红色果肉的苹果是由于MYB10转录因子的激活转录，使苹果果肉中花青素大量合成积累，促成苹果的红肉表型。食用高花青素含量的红肉苹果中对保持人体健康具有重要意义。



课题背景是基于目前我导师选育出的两个红肉苹果新种质，一个果肉颜色鲜红、肉质脆、极早熟并且极耐贮运，另一个除了不耐贮运其他都相同。目前已将这两个品种的转录组和代谢组测序，基因组测序也正在进行中。通过两个品种的全基因组比较分析，从分子水平上对苹果的花青素色素积累、香味、早熟性、耐贮性之间的发育关联性进行了解。这里面涉及的转录因子家族有MYB及与其互作共同与调控苹果的花青素合成的bHLH、WRKY等，还有与成熟期有关的乙烯合成转录的ERF转录因子家族等。

3、MYB转录因子家族简介

我对MYB转录因子家族较熟悉一下，所以接下来就介绍一下我目前了解的相关信息。

MYB是成髓细胞瘤myeloblastosis（成髓细胞瘤）的缩写，因为第一个鉴定的MYB转录因子是第一个鉴定的MYB基因源自禽成髓细胞瘤病毒的“癌基因”。

在MYB转录因子家族中，根据其DNA结合结构域数量的多少分为了MYB和MYB-related两个亚家族，只含有一个结构域归类为MYB-related，含有两个及两个以上的DNA结构域归类为MYB家族，SWIRM是一个禁止域，不论是哪个亚家族中都不允许出现SWIRM结构域。

**MYB\_related**

**MYB**

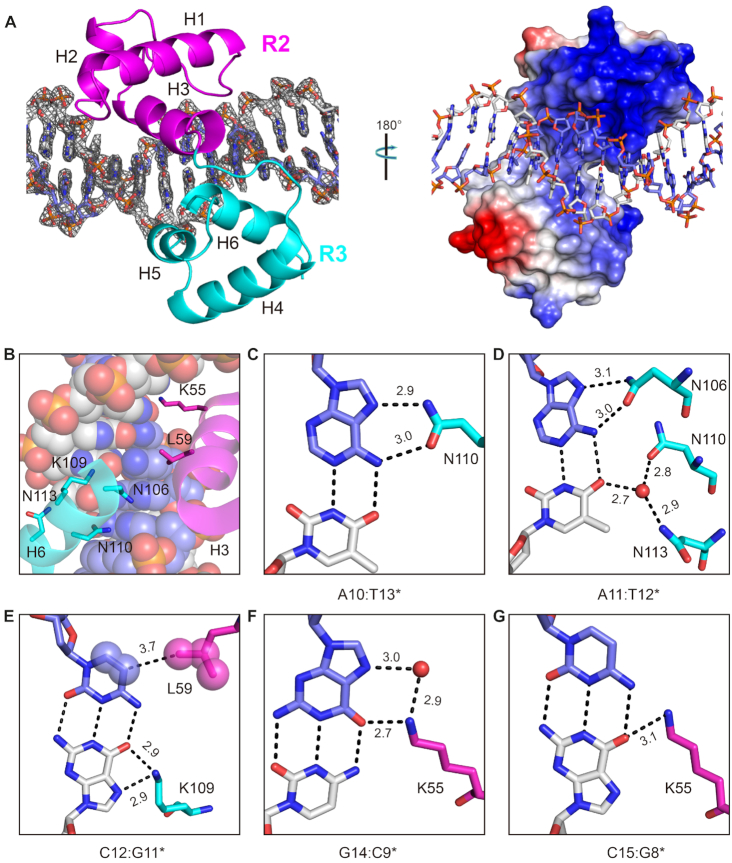
**Myb\_dna\_bind**

**SWIRM**

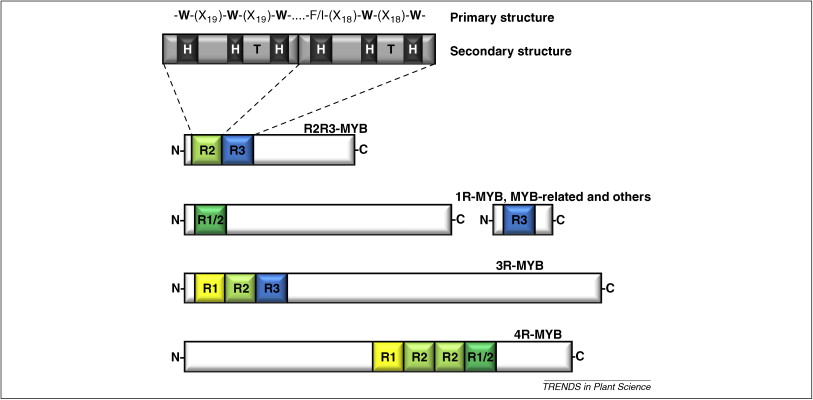
1

2

**MYB**

MYB结构域通常是由约 53 个氨基酸组成的的螺旋-转角-螺旋结构。三个规则间隔的色氨酸残基在三个α-螺旋中结构中形成色氨酸簇，构成疏水核心，稳定其结构。

MYB共有四种类型的转录因子，并且根据TF中含有的结构域的数量不同，分为1R-MYB、R2R3-MYB、3R-MYB、4R-MYB，如下图。除了1R-MYB属于MYB-related家族，另外几种类型都含有两个及以上的DNA结合结构域，属于MYB家族。R2R3-MYB是植物中特有的一种MYB转录因子类型，与植物的花青素合成调控等密切相关。

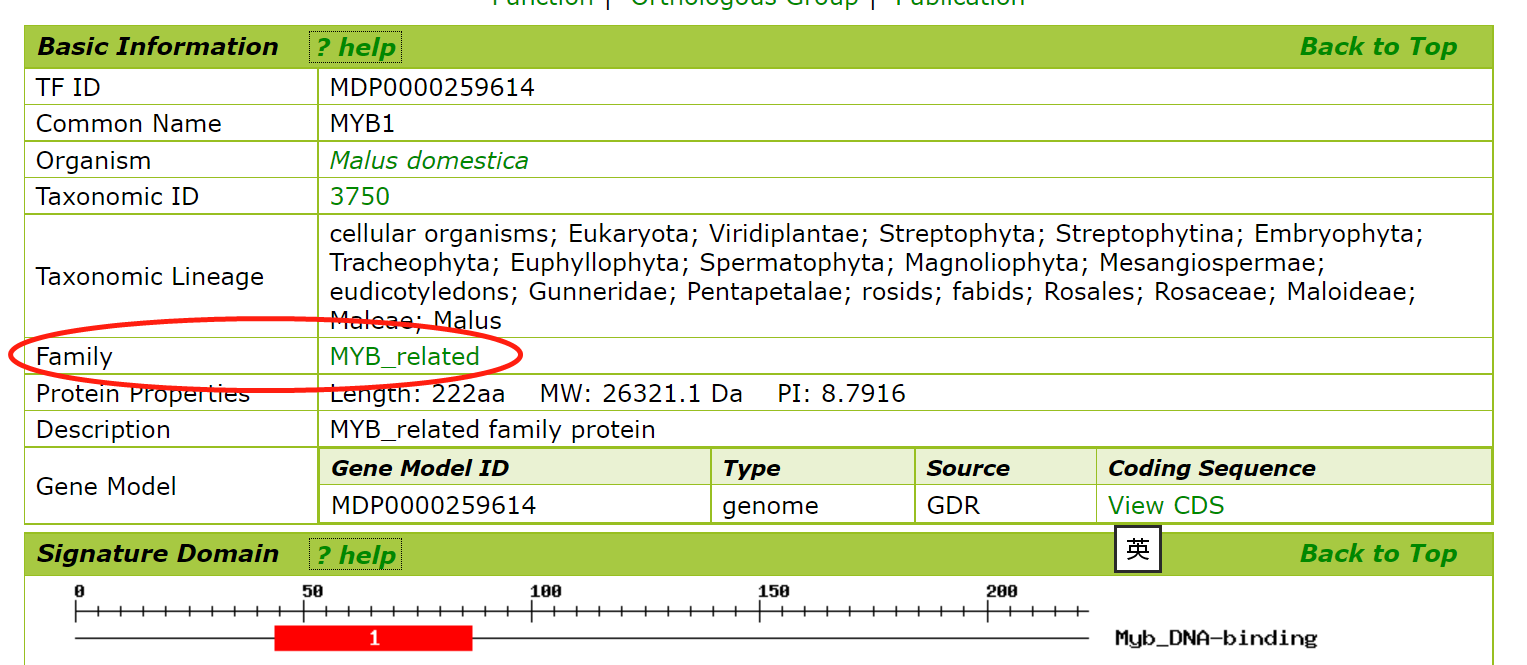


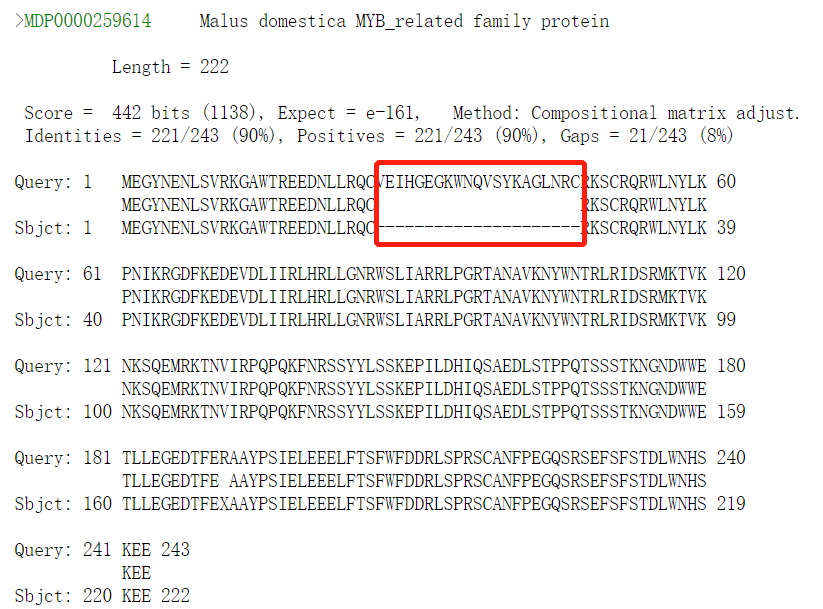
<https://www.uniprot.org/uniprot/A0SQ35>

<https://pfam.xfam.org/family/PF00249>

这两个链接是MYB10转录因子在UniProt库中的注释信息以及在Pfam中MYB转录因子家族的注释信息，之前罗老师课上用hmmsearch来进行演示，就是用的从Pfam中下载的MYB家族的数据，将所有序列中含有的结构域的个数进行了初步的扫描。

我之前在用这个转录因子库的过程中发现了一些苹果转录因子的注释信息有错误，家族分类也不正确，可能跟这个测序的片段由于外显子可变剪切等出现了缺失，导致扫描时缺少结构域，就被错误归类。比如前面提到的MYB10转录因子，在库中的注释信息是属于MYB-related家族，只有一个DNA结合结构域，与UniProt库中的MYB10序列进行blast就发现缺失一段序列。





在Plant TFDB中，用的是苹果第一版2012年测序组装的基因组数据，目前苹果中注释信息最全面、最常用的是于2017年用金冠苹果的双单倍体样品进行从头测序组装的GDDH13这个版本。

（https://www.rosaceae.org/species/malus/malus\_x\_domestica/genome\_GDDH13\_v1.1）

所以接下来用苹果的基因组序列数据来建库的话，我们就用这个啦。