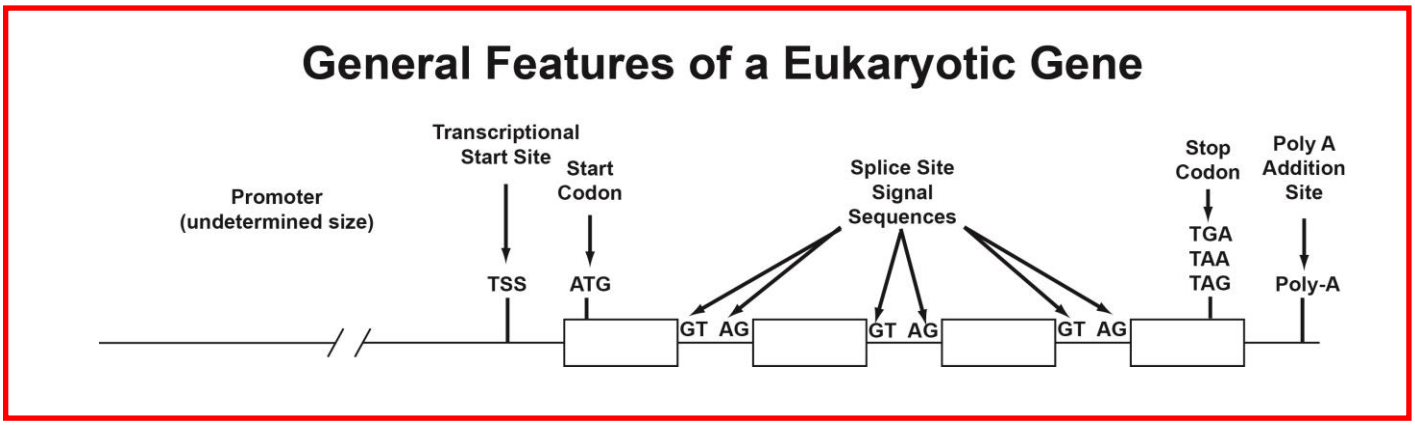
Transcription, Transcription Factors, and Chromatin Remodeling



The Gene

- Complex collection of sequences that
  - ***Controls a phenotype***
    - Individually
      - ***OR***
    - Complexed with the action of other genes
- Size varies
- Structural features vary
- Encode for a protein(s) that is translated from a mRNA
- **Expression**
  - **Requires many associated factors**

The Genome Is Significantly Involved in Gene Regulation

- The number of promoter sequences is equal to the number of protein coding sequences
- **Transcription regulation is a major function of the genome**

Transcription - the synthesis of RNA from a DNA template

Three Main Transcription Events
For prokaryotic or eukaryotic organisms

1. Initiation

- Binding of RNA polymerase to double-stranded DNA
  - This step involves a transition to single-strandedness in the region of binding
    - RNA polymerase binds at a sequence of DNA generally called the **promoter**
- ***Initiation is the most important step in gene expression!!!***

2. Elongation

- The covalent addition of nucleotides to the 3' end of the growing polynucleotide chain
  - Involves the development of a short stretch of DNA that is transiently single-stranded

3. Termination

- The recognition of the transcription termination sequence
  - Release of RNA polymerase

Product of Transcription

Transcription Unit

- Extends from the transcription start site (TSS) to the termination sequence
- The product is called the
    - **Primary Transcript**
        - Immediate transcription product

Other Critical Sequences for Transcription

- **Upstream Sequences**
    - Sequences before the mRNA transcription start site
    - Necessary for building the transcription apparatus for transcription
- **Downstream Sequences**
    - Sequences after the start site
    - Can also have a regulatory effect

Eukaryotic RNA Polymerase

1. Three types of RNA Polymerase exist

- **Each with a distinct function**

Type of Polymerase	Product	Location	Size	Subunits
RNA Polymerase I	rRNA	nucleolus	590 kDa	14
RNA Polymerase II	hnRNA	nucleoplasm	550 kDa	12
RNA Polymerase III	tRNA	nucleoplasm	700 kDa	17

2. RNA Polymerase II is key to mRNA synthesis

- ~550 kd in size
- Two large subunits
- <10 small subunits
    - **Many non-polymerase factors required for binding of the enzyme to DNA**

Steps in Model Eukaryotic Transcription

Model: Adenovirus late promoter

- Requires four accessory factors and RNA Polymerase II added in a defined manner
- These steps are common across eukaryotes

Order	Factor	Length of promoter covered (bp)
1.	TFIID	-42 to -17 (binds TATA box)
2.	TFIIA	-80 to -17
3.	TFIIB	-80 to -17 and -10 to +10
4.	RNA Polymerase II	-80 to +15
5.	TFIIE	-80 to +30

The Transcription Product

Heterogeneous nuclear RNA

- **hnRNA**
- ***Complexity of hnRNA is 4x the mRNA pool***
  - Splicing of introns from the primary transcript
    - Average hnRNA size = 8000 - 10,000 nucleotides
    - Range = 2000 - 14,000 nucleotides

Splicing

- Removes the introns from the hnRNA
- Alternate splicing
  - **An intron is skipped; or**
    - Uses signals other than the GT/AG associated with introns signals
  - **Result**
    - **Multiple transcripts and proteins can be synthesized from a single gene sequence**

Finishing the mRNA

A. 5' Capping Step

- **Protects the transcript**
  - Added immediately after the start of transcription
  - The original 5' base of the mRNA is rarely seen
- Unique nucleotide
  - **5' methyl guanosine**
- Sequence linkage
  - 5' methyl guanosine 5'-5' linkage
    - Not the typical 5'-3' linkage
- Enzyme
  - Guanylyl transferase.

B. 3' Polyadenylation Step

- Enzymatic action of **Poly (A) Polymerase**
  - Adds a ***Poly-A tail*** (many adenines) to end of transcript
  - Found in all eukaryotic mRNA
- Sequence signal for adding the poly-A tail
  - 5'-AAUAAA-3'
  - Sequence is located
    - About **10-30 bp upstream of the poly A tail**.

Transcription Factors: General Terms and Concepts

Promoter

- Difficult to define
- General definition
    - ***All the DNA sequences containing binding sites for RNA polymerase and the transcription factors necessary for normal transcription***

Transcription Factor

- ***Any protein other than RNA polymerase that is required for transcription***

Functions of Transcription Factors

- Bind to RNA Polymerase
- Bind another transcription factor
- Bind to cis-acting DNA sequences

Basal Transcription Apparatus

- RNA polymerase + General transcription factors
- Both needed to initiate transcription
    - ***These steps are the minimum requirement for transcription***

Upstream Transcription Factors

- Ubiquitous factors that increase the efficiency of transcription initiation
    - ***Set of factors necessary to for expression of each gene***

Inducible Transcription Factors

- Act in the same manner as an upstream factor
    - **BUT**
        - ***Their synthesis is regulated in a temporal or spatial manner***

Early Research on Plant Regulatory Regions

Temporal regulation

- ***Gene only expressed a specific time in development***
  - **Examples:**
    - Genes that are only expressed in **day light**
    - Genes that are only expressed during **flower development**

Spatial regulation

- ***Gene only expressed in a specific location in the plant***
  - Examples:
    - **Seed storage proteins**
    - **Leaf or root specific genes**

Technical Approach to Studying Gene Regulation

Dissect the promoter region and determine effect on gene expression

- **"Promoter Bashing"**
  - **Analyzing effects of upstream regions on gene expression**
- **Steps in "Promoter Bashing"**
  1. Determine sequence and identify of the promoter region of a gene
     - Usually ~1500 – 2000 bp upstream of coding region
  2. Sequentially remove portions of the promoter
  3. Develop ***expression construct*** with
     - "Truncated" promoter segment fused to a reporter gene
  4. Introduce construct into plant tissue
     - Transgenic plant or cell culture
  5. Expose biological unit (plant or culture) to a biological treatment
     - ***Measure expression level of reporter gene for each construct***

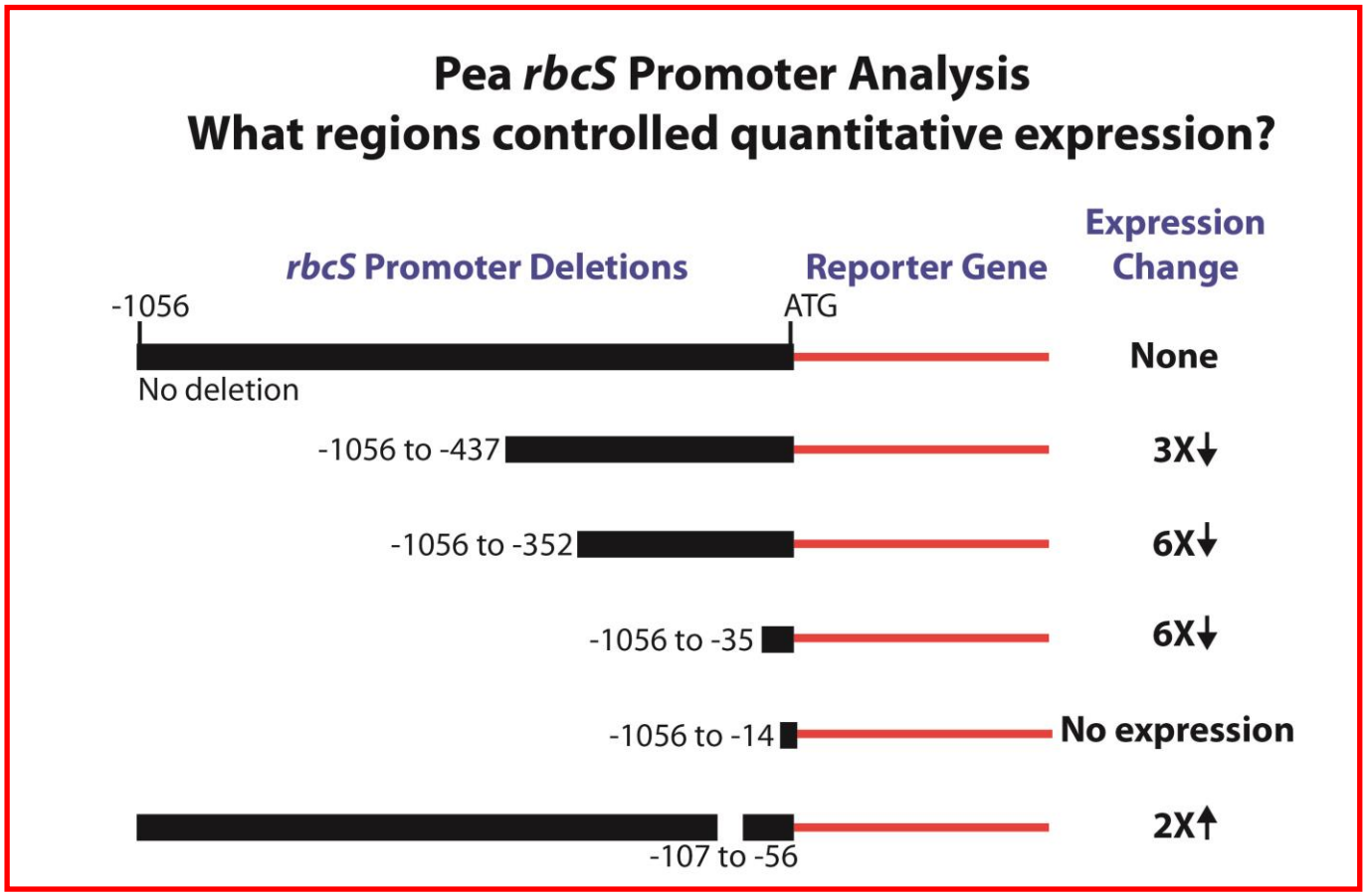
Light Regulation of Gene Expression

- **Morelli et al: Nature (1985) 315:200**

Example: *rbcS*

Ribulose bisphosphate carboxylase small subunit

- **Deletions of the promoter region studied for effect on reporter gene expression**



Promoter region	Effect on expression	Implication
-1052 to -437	3X reduction	Sequences between -1052 to -437 **increases expression 3X**
-1052 to -352	6X reduction	Sequences between -437 and -352 **increase expression 2X**
-1052 to -35	6X reduction	No sequence between -352 and -35 controls level of expression
-1052 to -14	no expression	Sequence between -35 and -14 (TATA) absolutely required for expression
-107 to -56	2x increase	Sequences between -107 and -56 decreases 2x

Conclusion

- ***Specific regulatory sequences in the upstream ("promoter") modulate the level of gene expression in light conditions***

Hormone Regulation of Gene Expression

Bean Chitinase
Broglie et al. 1989. Plant Cell 1:599

Chitinase Gene

- Defense gene against fungal pathogens
  - Induced by ethylene
    - 20-50X in expression
- Bean chitinase transgenic tobacco plants developed
  - Promoter deletions constructs developed
- Promoter contains
  - **Suppressor elements**
  - **Enhancer elements**
  - **Ethylene response elements**

Deleted region	Expression w/ ethylene	Type of cis element
-1057 to -846	3x increase	Suppressor
-1057 to -422	20x decrease	Enhancer
-1057 to -195	No ethylene induction	Ethylene response

Transcription Factors – Big Picture in Eukaryotes

Talbert et al. 2019. Nat. Rev. Genet. 20:283

Example of Evolutionary Conserved Complex

- **Polycomb Repressive Complexes PCR1 and PCR2**
  - **"…**essential roles in controlling cell- type-specific developmental gene expression in multicellular eukaryotes. Diversification of these complexes may have facilitated the advent of cell differentiation in multicellular organisms by serving as a ***flexible, modular silencing apparatus*** **that selectively *inactivates* a range of *cis* elements in response to developmental cues**."
- **PCR2s: Complex methylates H3K27 (histone 3, lysine 27)**

PCR Complex
**Different proteins
**Different growth states
**Histone methylation control

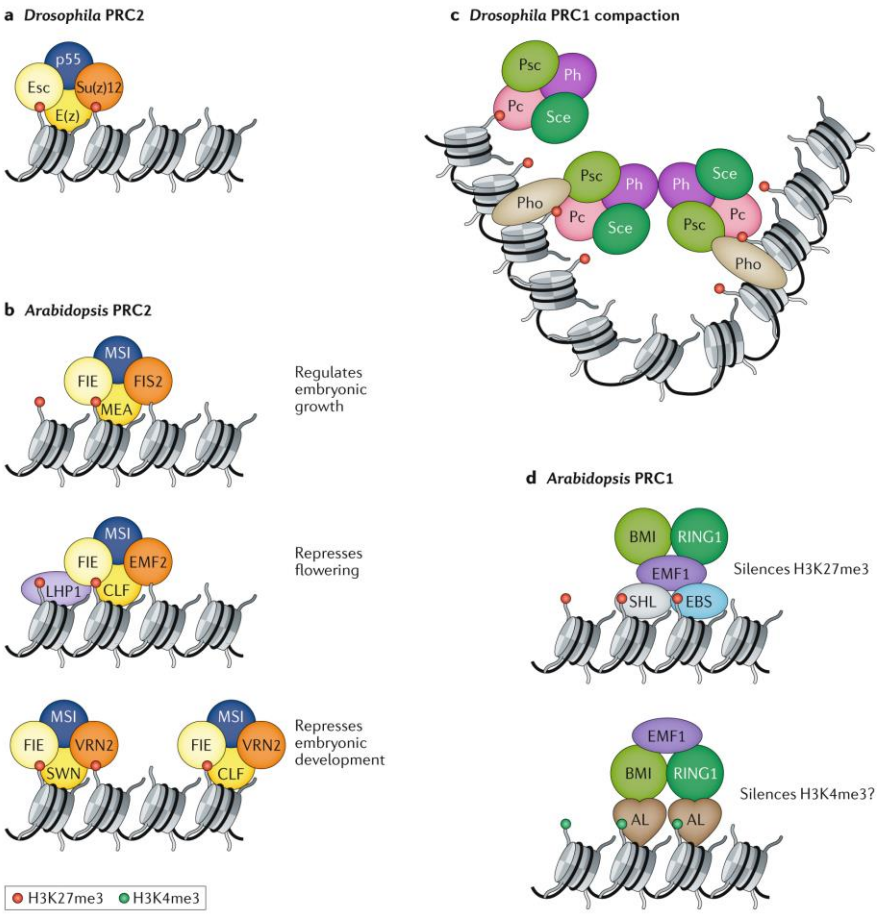


Fig. 3 (edited) | PRc1 and PRc2 in animals and plants. b | The *A. thaliana* chromodomain protein **LHP1 binds to H3K27me3** and together with the histone methyltransferase curly leaf (CLF) acts to ***spread H3K27me3***. **d |** In *A. thaliana*, PRC1 complexes are not well characterized, but two complexes have been proposed containing **BMI and RING1**, homologues of **Psc and Sce**, respectively , along with plant- specific components with PHD fingers that can bind to H3K27me3 (SHL and EBS) or H3K4me3 (AL). The latter complex is proposed to shut off active genes to transition to repressed chromatin marked with H3K27me3 and H2AKub. **Shapes coloured identically represent homologous proteins.**

Major Principle of Gene Regulation

- Remember this quote:
  - **"Trans-acting factors bind to Cis-acting elements"**
    - ***These interaction controls gene expression***
- **Trans-acting factors**
  - Commonly called: Transcription Factors
    - TFs are the product of a gene different (usually) than the one that it regulates
  - Function in the nucleus of the cell
  - Bind upstream of the start site of transcription
- **Cis-acting Elements**
  - Short conserved (relatively) DNA motifs (or sequence)
    - Located upstream of the transcription start site
  - **TFs bind to the element to control gene expression**

Transcription Factors

- **Wray et al Mol Biol Evol 2003. The Evolution of Transcriptional Regulation in Eukaryotes. 20:1377**

Phenotype is Affected by Mutations In:

- **Structural region of a gene**
  - Function of a protein is modified (structure/function relationship)
- **Regulatory region of a gene**
  - When/where/how much the protein is expressed
  - **Gene regulation**

Considerations of Gene Regulations

1. ***Changing the regulation pattern = can change phenotype***
2. ***One transcription factor (TF) can affect multiple genes in a pathway***
3. ***TF orthologs regulate different organisms differently***
4. ***Promoter contains module that affect expression***

Approaches to Studying Gene Regulation

- **Mutants**
  - Do induced mutants represent natural variation?
- **Expression patterns**
  - Expression patterns of orthologs can differ among species
- **Expression levels**
  - Phenotypic differences result from changes in the amount of protein

Effect of Varying Expression level

- **Spatial effects**
  - Varying the amount of expression in a tissue can change phenotype
- **Cis-effects**
  - Variation in expression level often related to changes in cis-element sequence
- **Inducibility**
  - Alleles can be induced differentially

Levels of Expression Can Vary at the:

- **mRNA level**
- **Protein level**

What Amount of the Gene Expression Variation is the Result of "Controlling Region" Variation???

- **Natural variation exists in promoters**
  - Associated with phenotypic changes
- **Artificial selection of promoter sequences can change expression**
  - Maize *tb* locus is an example
- **Promoter "elements" are conserved among species**
  - Specific sequences important for gene expression
- **Variation in promoter sequence related to human disease susceptibility**
  - Susceptibility to specific pathotypes related to promoter sequences

Transcription Patterns are Variable

- **Transcription initiation is the most important step in phenotypic expression**
- Regulation is at the gene not gene family level
  - **Paralogs are independently regulated**
- **Transcription is dynamic**
  - **Expression levels vary**
  - **Expression can fluctuate rapidly**
  - **Expression in neighboring cells can differ**
- **Expression profiles vary among genes**
  - Regulatory gene expression profile is inducible and highly variable
  - Housekeeping gene expression is generally constitutive but varies in response to stimuli and by cell type

Role of Controlling Regions (=Promoters) in Gene Expression

- **Promoters**
  - Contain sequence motifs that bind factors that modulate gene expression
- **Constitutive (housekeeping) promoters**
  - On by default
  - Turned off in response to stimuli
- **Inducible promoters**
  - Off by default
  - Turned on in response to stimuli
- ***TF determine if genes are turned on or off***

Promoters

- Universal conserved features are not found
- Common sequence motifs not found

Basal Gene Expression

- **Basal promoter**
  - RNA polymerase complex binding site
    - Contains TATA box or initiator element
  - Null promoters exist
    - Lacks TATA box or initiator element
  - Multiple basal promoters can exist for some genes
- **TATA-box binding protein (TBP)**
  - First protein to bind the basal promoter
    - Other proteins guide TBP to the binding site
- **RNA polymerase holoenzyme complex**
  - Complex interactions of proteins builds the transcription complex

- **Basal promoters provides for minimal, low level of expression**
  - ***Expression mediated by constitutively expressed general transcription factors***

Modifying Basal Gene Expression Levels

- **TF binding to controlling regions required for full gene expression**
  - TF are specific to cell types and stimuli conditions
    - ***Interaction of controlling regions and TF controls gene expression***

Controlling Region TF Binding Sites

- **Binding sites are isolated in controlling region**
  - Binding sites are embedded in regions to which **no other TFs bind.**
- **Binding sites numbers**
  - 10 – 50 binding sites for 5 –15 TF
- **Role of other sequences**
  - Local, sequence-specific conformational changes can affect TF binding
  - AT-rich regions
- **Spacing of binding sites**
  - Partial overlap to
  - 10s of kilobases apart

Features of TF Binding Sites

1. Size

- Footprint (sequences covered by TF) is 10-20 bp
- Direct binding site is 5-8 bp
- Essential sequence is 4-6 bp

2. Site definition

- Consensus sequence (although not all consensus sequences bind TF)

3. Binding sites can overlap

- TF pool determines which site is bound
    - Binding sites compete for a limited TF pool

4. Location

- 100 basepairs to 100 kilobases from transcription start site

5. Functional TF binding site locations

- >30 kb 5' of basal promoter
- few kb of basal promoter
- in 5' UTR
- in introns
- >30 kb 3' of basal promoter
- exon
- other side of adjacent gene

Features of TF Binding Sites (cont.)

6. Location constraints

- Some sites are constrained to specific positions relative to transcription start site

7. Isolating binding sites effects

- **Insulator sequences** limit TF interactions to specific basal promoters
    - TATA or TATA-less TF interaction specificity
- Specific recruitment of TF at a specific sequence to interact with basal promoter

Abundance of Transcription Factors

- **TF are members of small to large multi-gene families**
    - Arabidopsis
        - LFY and SAB Families
            - **One member**
        - βHLH Family
            - **225 members**
        - Variation in family size is a result of gene duplication events
- **12-15 unique DNA binding domains**
    - Evolutionary conservation

Modular Domain Structure of Transcription Factors

1. **DNA binding domain**
   - Localized
     - MADS-box or homeo domains
   - Dispersed
     - Zn-finger or leucine zipper domains

bHLH
MYB

2. **Protein-protein interaction domain**
   - Binding to other proteins necessary for activation
3. **Intracellular trafficking domains**
   - Nuclear localization signal
4. **Ligand binding domain**
   - Steroid or hormone-binding domains
5. **Evolutionary domain shuffling has occurred**
   - Protein-protein interaction domain lost but DNA binding domain maintained

Transcription Factor DNA Binding Domain

1. **Most bind the major groove of DNA**
2. **Domain sequence is highly conserved**
   - Single amino acid mutations can alter significantly TF binding
3. **TF binding specificity ranges from 3-5bp**
4. **Specificity may be increased by**
   - Multiple binding domains
   - Domains that bind minor groove
   - Dimerization of two proteins, (homomeric or heteromeric)
5. **Binding is strong and highly specific**
   - 5000 – 20,000 copies of TF needed for high binding specificity
6. **Cofactor interactions increase specificity**
   - Phosophorylation
7. **Paralogs may have unique binding specificities**

MADS Box Binding Example

- **Smaczniak et al. 2012. Development 139, 3081-3098 (2012)**

MADS box genes

- **Key regulatory of growth**

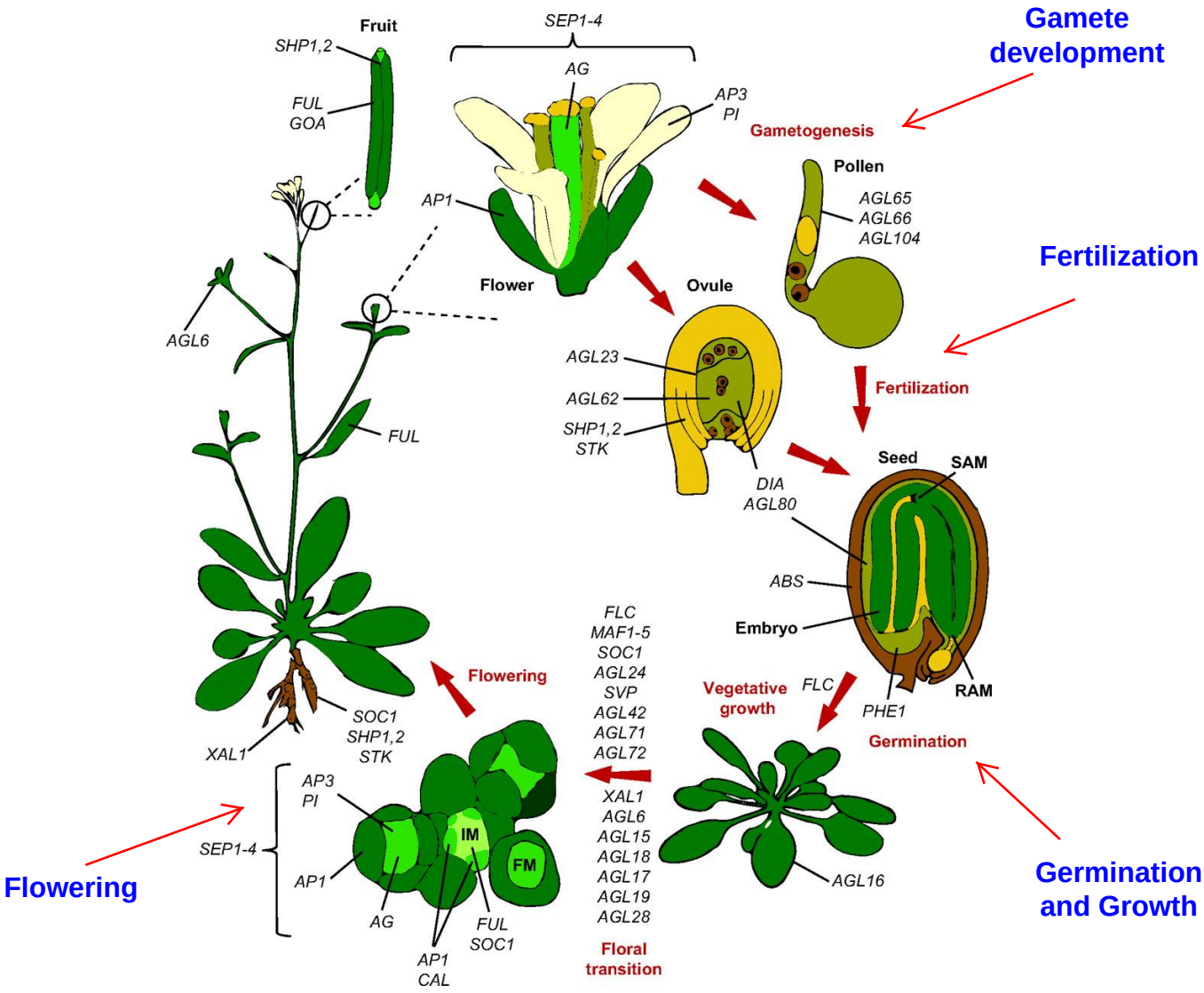


Fig. 2. Functions of MADS-box genes throughout the life cycle of *Arabidopsis thaliana*. *Arabidopsis* progresses through several major phase changes during its life cycle and MADS box genes play distinct roles in the various developmental phases and transitions. **Reproductive development** starts with the generation of male and female haploid gametes (gametogenesis) and, after double fertilization, this results in a developmentally arrested embryo that possesses a root apical meristem (RAM) and a shoot apical meristem (SAM), enclosed within a seed. Under favorable conditions, **seeds germinate** and young plants go through the **vegetative phase** of development in which leaves are formed and plants gain size and mass. Finally, the plant is ready to flower and the **floral transition stage** results in the conversion of vegetative meristems into inflorescence meristems (IMs) and floral meristems (FMs) that produce floral organs. Subsequently, **gametes are formed** within the inner flower organs, thus completing the cycle. The **MADS box genes that are involved in each of the various stages of development are indicated.**

- **MADS box protein act in a regulator complex**

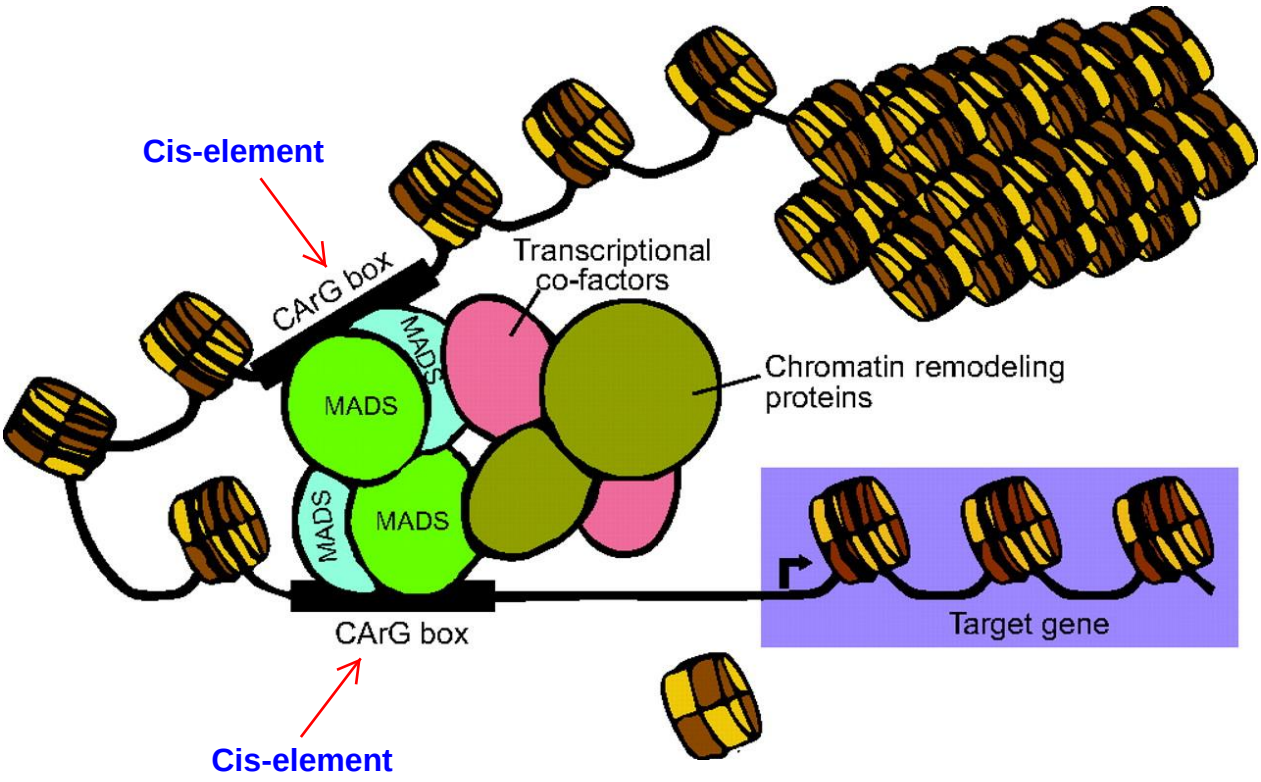


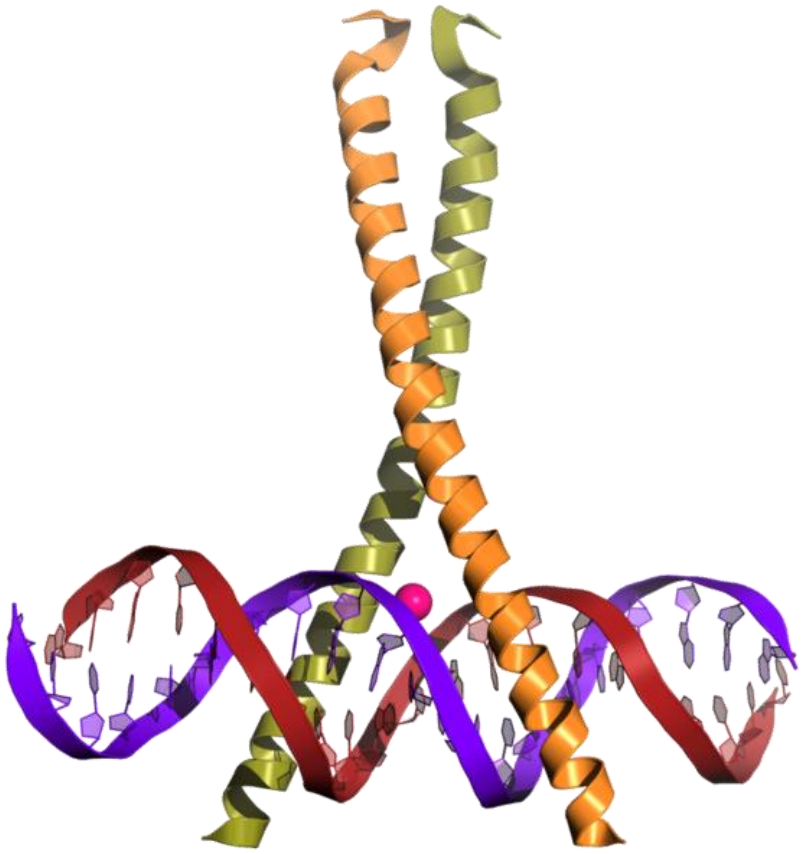
Fig. 3. Model for the action of MADS-domain protein complexes. Shown is a model of MADS-domain protein complex formation and a hypothesized mechanism of regulatory action. In this model, **MADS domain proteins (green and blue) form quaternary complexes** according to the 'floral quartet' model and **interact with two DNA binding sites** (CArG boxes; black) in close proximity, resulting in **DNA looping**. Subsequently, MADS-domain proteins **recruit transcriptional co-factors** (pink), which mediate transcriptional regulation and may influence target gene specificity, as well as chromatin remodeling proteins (brown), which relax the chromatin structure at the target gene transcription start site allowing for the initiation of transcription. **Depending on the selection of transcriptional co-factors and chromatin remodeling factors, the complex may also play a role as a transcriptional repressor.**

From: Smacznaik et al. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development 139:3081.

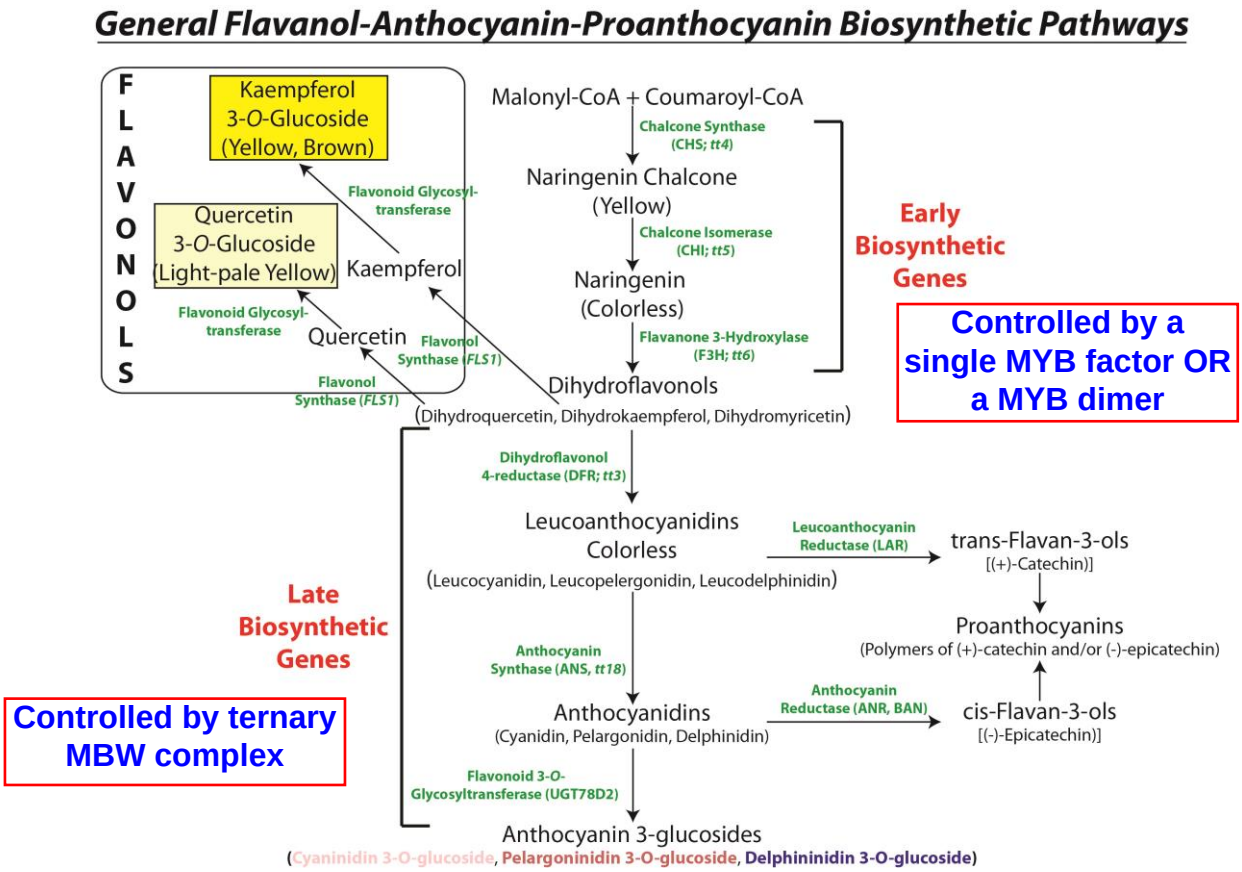
- **Transcription Factor Can Function As Dimers**

bZIP TFs

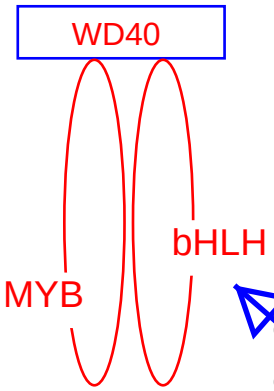
- A transcription factor found across all taxonomic domains of eukaryotes
- bZIP proteins act as dimers
    - Two bZIP work together to regulate gene expression
        - Basic region binds DNA
        - Acid regions binds together the two proteins



Regulation of Flavonoid (Pigment Molecules) in Plants: A Conserved System



- A ternary (three proteins) complex is required to activate expression of the late biosynthetic genes in the flavonoid pathway
  - **MBW**
    - **M** = **Myb protein**
      - TT8 gene in Arabidopsis
    - **B** = **basic Helix-Loop-Helix protein (bHLH)**
      - TT2 and other Arabidopsis gene
    - **W** = **WD40 protein**
      - TT2 gene in Arabidopsis
  - ***Members of the complex change by tissue type***
  - The functional components of the complex conserved throughout plants
    - Mendel A gene = TT8 gene
      - Green vs yellow (recessive) seed color



MBW Activities

- **Xu et al. 2015. Trends in Plant Science 20:176)**

- **MBW complex TF family members change during the life cycle of the plant**

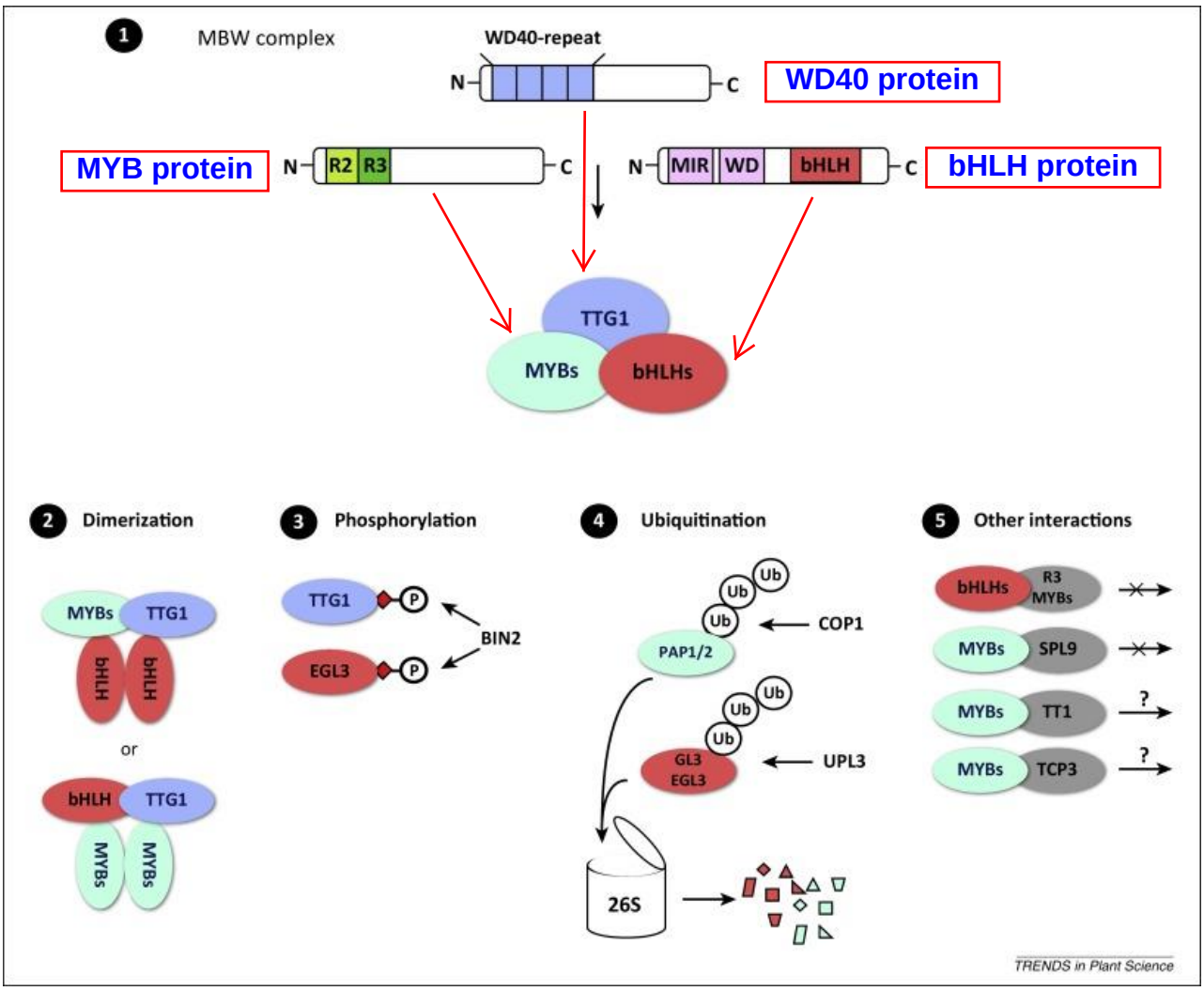


Figure 1. MBW (MYB–bHLH–WDR) complexes and post-translational regulation. The bHLH proteins of the IIIf subgroup (TT8, GL3, EGL3, and AtMYC1) can interact with R2R3-MYBs from various subgroups such as TT2, PAP1, or PAP2, and form ternary complexes with TTG1 **(1)**. The interactions involved the R3 repeat of the MYB and the N-terminal MYB-interacting region (MIR) of the bHLH. The specific role of each partner in the complex is not yet fully understood. The activity of the MBW complexes can be regulated through different post-translational modifications including **dimerization (2)**, **phosphorylation (3)**, **protein degradation (4)**, and various **protein interactions (5)**.

- **Different members of bHLH and MYB protein families interact in different tissues**

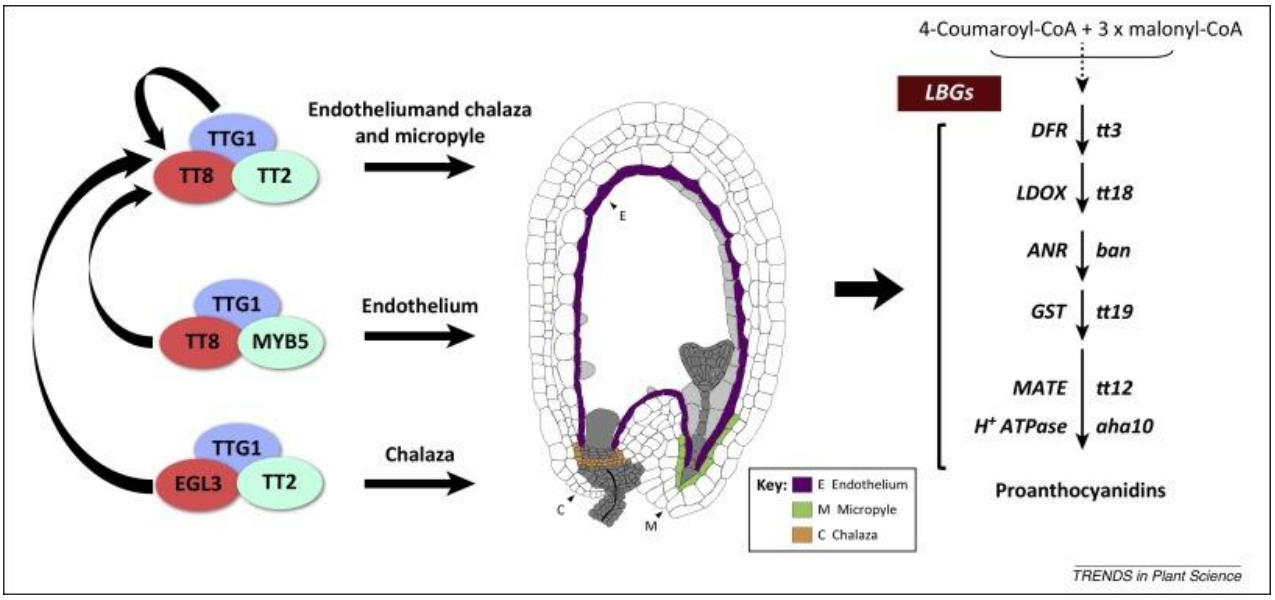


Figure 2. MBW regulation of proanthocyanidin biosynthesis in the seed coat. The schematic representation of a developing seed is adapted from [85]. PA-accumulating cells are localized in the most inner cell layers of the integuments (i.e., E, endothelium; C, chalaza; and M, micropyle area). Names of genes and proteins are indicated in capital letters (with italics for genes), and corresponding mutants in lower-case italics. Abbreviations: **DFR**, dihydroflavonol-4-reductase; **EGL3**, enhancer of glabra3; **LDOX**, leucoanthocyanidin dioxygenase; **ANR**, anthocyanidin reductase; **GST**, glutathione-*S*-transferase; LBG, late biosynthetic gene; MATE, multidrug and toxic efflux transporter; **MBW**, MYB–bHLH–WDR; **PA**, proanthocyanidin; **TT1/2/8/16**, transparent testa 1,2,8,16; **TTG1,2**, transparent testa glabra 1,2. Curved arrows indicate the cell-specific induction of *TT8* expression by MBW complexes.

- **Hormonal, developmental, and environmental regulation of the MBW complex genes**

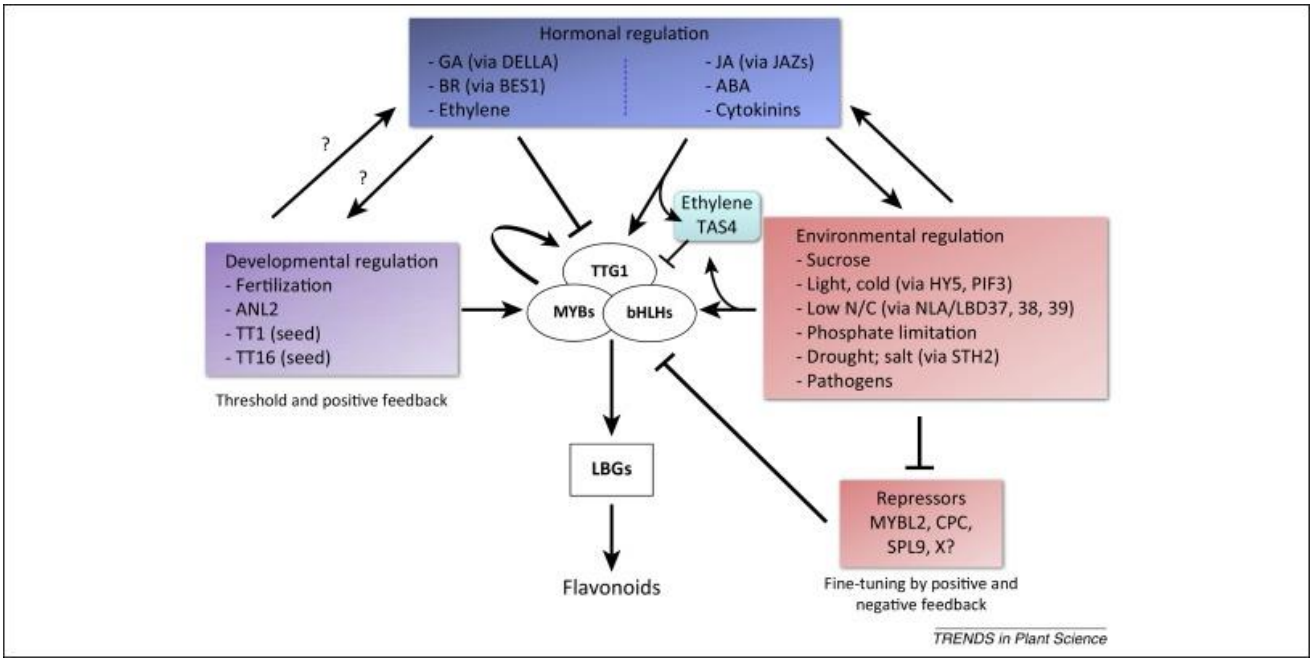


Figure 4. **MBW complexes** are involved in both types of developmental and environmental **regulation of flavonoid biosynthesis** through the **activation of late biosynthetic gene (LBG) expression.** The complexity of these transcriptional regulatory networks is remarkable. It allows **cell specific accumulation** of various flavonoids to fulfill their different functions. Developmental regulation in the seed involves a positive feedback loop allowing high-level and **specific expression of PA genes in a single cell layer** of the seed coat. By contrast, environmental regulation involving diverse negative feedbacks **allows fine-tuned and reversible expression** of flavonoid genes and flavonoid accumulation **depending on the physiological status** of the plant tissues and the environmental conditions.

Transcription Factor Protein-Protein Interactions Modulate Gene Expression

1. **Increase (or decrease) the frequency in which the transcription apparatus is built**
    - Can recruit (or prevent recruitment) of apparatus components
2. **Specific interactions necessary to regulate gene expression**
    - As homodimers
    - As heterodimers
    - As solo proteins
3. **Neighboring effects**
    - TF at one site can prevent cofactor from interacting with a neighboring site
4. **Altering chromatin structure**
    - Recruit other complexes that
        - Acetylate, deacetylate, methylate, or demethylate histones
        - Methylate or demethylate DNA
5. **Create physical bends**
    - Facilitates binding of other TF
6. **Cofactors can bring TF and transcriptional apparatus together**

Role of Functional Modules

- **Functional modules**
  - **Collection of proteins that collaborate to control gene expression**
- **Module functions**
  1. ***Initiate*** transcription
  2. ***Enhance*** transcription rate
  3. ***Repress*** transcription rate
  4. ***Mediate*** extracellular signals
  5. ***Insulate*** one module from another
     - Insulator function
  6. ***Tethered*** to cellular structure
     - Membrane tethered
     - Released by signal and activate module
  7. ***Bring*** other modules into contact with basal promoter

Additive and Epistatic Interactions of Transcription Factors

1. **Modifying** one TF and its module interaction can ***additively reduce*** the phenotype
2. **Modifying** insulator or tethering TF functions is ***epistatic***
   - **Proper expression, recruitment, and modular association of TF is necessary for full phenotypic expression**

A Transcription Family Has Multiple Target Genes

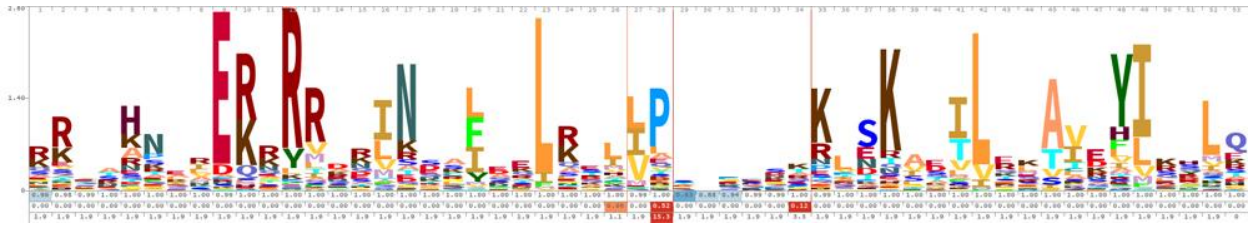
- The function of TF networks ***affect many genes***
- Because of the limited number of TF, ***a single TF may interact with 10s to 100s of genes***
    - **Drosophila *eve* and *ftz* regulate the majority of genes in the genome**
- Mutations can be modulated by the effects of other downstream genes

Transcription Factors Defined by Conserved Pfam Sequence Motifs (mostly)

- **(Pfam: accepted motif sequence definitions; http://pfam.sanger.ac.uk/)**

WHAT IS Pfam?

- **A database of specific domains found in proteins**
- Example: **HLH domain**
    - Family members have the **HLH** (**H**elix-**L**oop-**H**elix) consensus DNA-binding domain amino acid sequence
        - **Pfam number**: PF00010
    - HMM (hidden Markov model) amino acid sequence logo
        - The larger the letter, the more frequently the amino acid appears in the proteins with the function

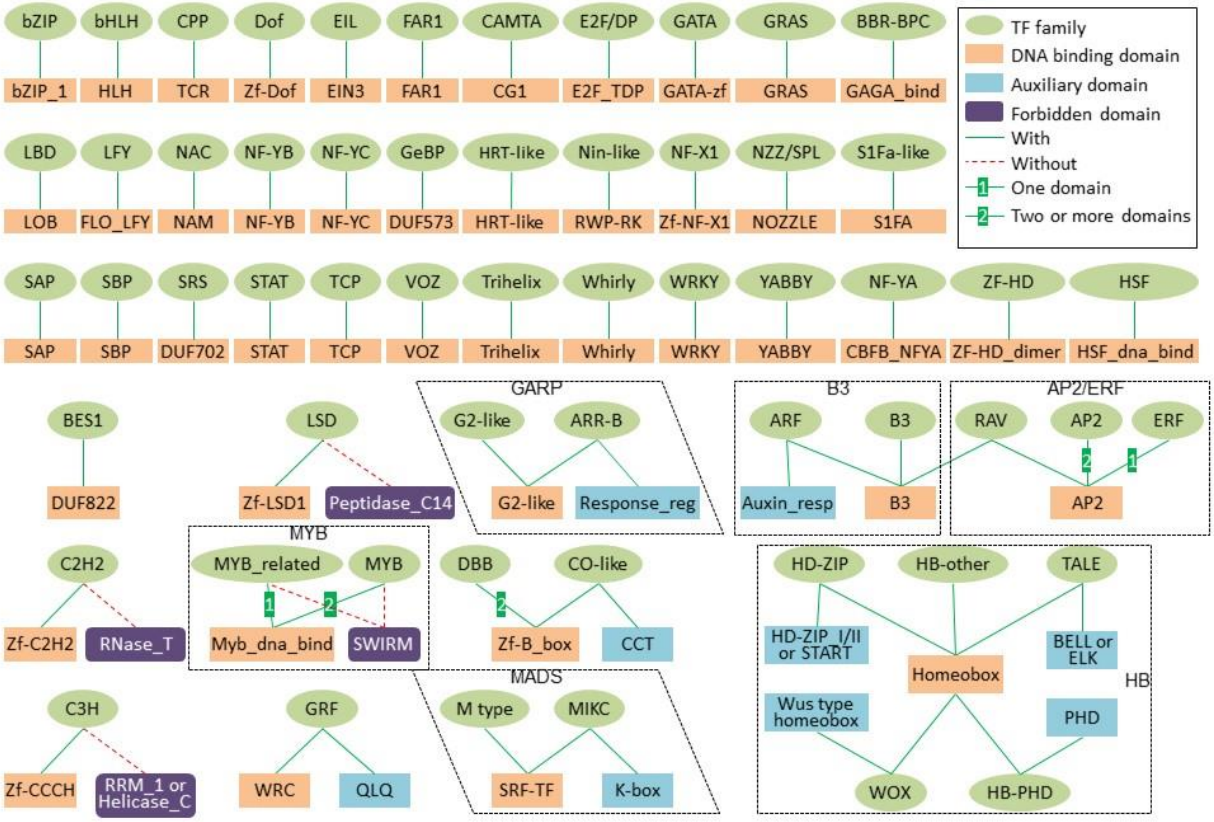


Plant Transcription Factor Database: Plant TFDB

http://planttfdb.cbi.pku.edu.cn/index.php

- How transcription factors are defined
  - ***Some*** Pfam domains have ***DNA binding functions***
  - The DNA domain they bind to is the ***cis-acting element***
    - **Proteins with Pfam-defined DNA binding domains are considered *TRANSCRIPTION FACTORS***
- **Family assignment rules:**

  http://planttfdb.cbi.pku.edu.cn/help_famschema.php)



TF Family Examples

- **bHLH Family** (2nd family, first green family row)
  - Has the HLH (Pf00010) domain
- **NAC Family** (3rd family, second green family row)
  - Has the NAM (Pf02365) domain
- **ARF Family** (5th family, fourth green family row)
  - Has the Auxin_resp auxillary domain (Pf06507) **AND** the B3 (Pf02362) domain

Distribution of Transcription Factors Among Dicot Genomes

- **(family assignment rules: http://planttfdb.cbi.pku.edu.cn/help_famschema.php)**

Family	Grape (3x)	Papaya (3x)	Arabidopsis (3x + 2x)	Tomato (3x + 3x)	Soybean (3x + 2x + 2x)
AP2	19	17	30	27	76
ARF	17	10	37	22	85
ARR-B	12	12	21	21	42
B3	29	34	77	73	112
BBR-BPC	5	3	17	6	22
BES1	6	6	14	9	19
C2H2	64	76	116	99	267
C3H	43	28	66	48	136
CAMTA	4	4	10	7	23
CO-like	6	9	22	13	32
CPP	6	4	9	4	19
DBB	7	6	14	10	36
Dof	22	20	47	33	93
E2F/DP	7	6	16	8	28
EIL	2	4	6	9	12
ERF	80	77	139	137	330
FAR1	18	19	26	28	103
G2-like	40	51	64	59	164
GATA	19	23	41	30	70
GRAS	43	42	37	54	139
GRF	8	7	9	13	31
GeBP	1	4	23	11	11
HB-PHD	2	1	3	2	11
HB-other	7	8	11	16	31
HD-ZIP	33	29	58	58	140
HRT-like	1	2	2	1	1
HSF	19	18	25	26	61
LBD	44	35	50	47	111
LFY	1	1	1	1	2
LSD	3	2	12	3	17
M-type	18	225	70	67	88
MIKC	36	20	76	32	160
MYB	138	98	168	140	369
MYB_related	57	51	97	79	265
NAC	71	82	138	101	247
NF-X1	3	1	2	2	8
NF-YA	7	5	21	10	57
NF-YB	17	11	27	29	46
NF-YC	8	4	21	20	35
NZZ/SPL	1	1	1	1	0
Nin-like	8	6	17	10	45
RAV	1	2	7	3	5
S1Fa-like	2	1	4	1	4
SAP	1	2	1	3	2
SBP	19	11	30	17	73
SRS	5	4	16	9	33
STAT	1	1	4	1	1
TALE	21	11	33	21	101
TCP	15	22	33	36	71
Trihelix	26	29	34	31	93
VOZ	2	2	3	2	20
WOX	11	11	18	10	42
WRKY	59	49	90	81	233
Whirly	2	2	4	2	13
YABBY	7	9	8	9	34
ZF-HD	10	10	18	22	54
bHLH	115	105	225	161	480
bZIP	47	46	127	70	266
Total	**1276**	**1379**	**2296**	**1845**	**5069**

Distribution of Transcription Factors Among Monocot Genomes

- **(family assignment rules: http://planttfdb.cbi.pku.edu.cn/help_famschema.php)**

Family	Japonica rice (2x)	Brachypodium (2x)	Sorghum (2x)	Corn (2x + 2x)	Arabidopsis (3x + 2x)
AP2	22	29	32	54	30
ARF	48	36	33	62	37
ARR-B	11	9	13	13	21
B3	65	45	86	77	77
BBR-BPC	7	4	6	9	17
BES1	6	7	9	16	14
C2H2	135	93	122	179	116
C3H	74	53	55	111	66
CAMTA	7	10	10	10	10
CO-like	21	14	14	18	22
CPP	20	11	12	17	9
DBB	13	11	11	20	14
Dof	37	27	35	51	47
E2F/DP	10	7	13	24	16
EIL	11	6	10	9	6
ERF	163	120	165	205	139
FAR1	133	69	62	25	26
G2-like	62	61	56	89	64
GATA	32	30	34	54	41
GRAS	69	48	86	104	37
GRF	19	14	11	32	9
GeBP	13	15	15	29	23
HB-PHD	1	5	3	4	3
HB-other	17	12	8	28	11
HD-ZIP	61	43	47	97	58
HRT-like	1	1	1	0	2
HSF	38	26	25	49	25
LBD	39	24	36	60	50
LFY	2	1	1	4	1
LSD	12	7	6	20	12
M-type	35	24	46	47	70
MIKC	61	51	47	90	76
MYB	130	98	132	203	168
MYB_related	106	77	116	169	97
NAC	170	109	141	190	138
NF-X1	2	1	3	4	2
NF-YA	25	12	16	36	21
NF-YB	16	17	16	28	27
NF-YC	19	15	18	25	21
NZZ/SPL	0	0	0	0	1
Nin-like	15	15	16	23	17
RAV	4	4	4	3	7
S1Fa-like	2	2	2	5	4
SAP	0	0	0	0	1
SBP	29	18	22	55	30
SRS	6	5	6	11	16
STAT	1	1	1	2	4
TALE	45	30	28	52	33
TCP	23	21	21	52	33
Trihelix	40	32	36	59	34
VOZ	2	2	2	10	3
WOX	17	9	12	30	18
WRKY	128	87	110	163	90
Whirly	2	2	2	6	4
YABBY	15	13	10	31	8
ZF-HD	15	15	18	26	18
bHLH	211	158	233	308	225
bZIP	140	95	123	218	127
Total	**2408**	**1751**	**2198**	**3316**	**2296**

Distribution of Transcription Factor Families between *P. vulgaris* (common bean) and *G. max* (soybean)

- **Soybean has undergone a genome duplication since its split from common bean**
- **Soybean 2x the number of TFs per TF family**
- **(family assignment rules from: http://plntfdb.bio.uni-potsdam.de/)**

TF family	Pv count	Gm count	Ratio
ABI3VP1	41	90	2.2
Alfin-like	24	38	1.6
AP2-EREBP	179	363	2.0
ARF	27	60	2.2
ARID	12	26	2.2
ARR-B	15	31	2.1
AUX/IAA	30	66	2.2
BBR/BPC	5	18	3.6
BES1	7	16	2.3
bHLH	155	359	2.3
BSD	10	24	2.4
bZIP	78	204	2.6
C2C2-CO-like	8	26	3.3
C2C2-Dof	42	81	1.9
C2C2-GATA	32	64	2.0
C2C2-YABBY	8	18	2.3
C2H2	10	62	6.2
C3H	44	153	3.5
CAMTA	8	15	1.9
CCAAT	55	253	4.6
Coactivator p15	3	9	3.0
CPP	6	20	3.3
CSD	5	8	1.6
DBP	2	4	2.0
DDT	11	20	1.8
E2F-DP	7	16	2.3
EIL	7	12	1.7
FAR1	25	80	3.2
FHA	19	39	2.1
G2-like	49	131	2.7
GeBP	5	19	3.8
GNAT	38	58	1.5
GRAS	55	119	2.2
GRF	10	24	2.4
HB	119	203	1.7
HMG	9	24	2.7
HRT	1	1	1.0
HSF	30	52	1.7
IWS1	10	22	2.2
Jumonji	21	40	1.9
LFY	1	8	8.0
LIM	9	20	2.2

TF family	Pv count	Gm count	Ratio
LOB	49	95	1.9
LUG	5	12	2.4
MADS	78	180	2.3
MBF1	3	4	1.3
MED6	1	1	1.0
MED7	1	3	3.0
mTERF	34	58	1.7
MYB	141	291	2.1
MYB-related	68	314	4.6
NAC	90	186	2.1
NOZZLE	5	6	1.2
OFP	20	47	2.4
PBF-2-like	3	7	2.3
PHD	32	270	8.4
PLATZ	14	34	2.4
Pseudo ARR-B	6	12	2.0
RB	1	3	3.0
Rcd1-like	2	8	4.0
RWP-RK	12	28	2.3
S1Fa-like	3	12	4.0
SAP	1	2	2.0
SBP	23	47	2.0
SET	44	82	1.9
Sigma70-like	9	13	1.4
SNF2	37	64	1.7
SOH1	1	2	2.0
SRS	10	22	2.2
SWI/SNF-BAF60b	18	31	1.7
SWI/SNF-SWI3	5	9	1.8
TAZ	4	5	1.3
TCP	27	56	2.1
Tify	13	33	2.5
TIG	5	1	0.2
TRAF	22	56	2.5
Trihelix	41	73	1.8
TUB	10	24	2.4
ULT	1	11	11.0
VARL	3	6	2.0
VOZ	5	8	1.6
WRKY	90	186	2.1
zf-HD	19	57	3.0
Zn-clus	0	0	
Total	**2188**	**5225**	

Cis-acting Elements Vary Among Gene Family Members

- **Example: *rbcS*: small subunit of RUBISCO**
- **Manzara et al. 1991. The Plant Cell 3:1305**

Conserved cis-acting element

Table 1. Conserved Sequence Motifs in *rbcS* Promoters

Number/ Symbol	Motif	Sequence	Binding Factor	Putative Function	References
	L box	AATTAACCAA		unknown	Ueda et al. (1989)
1		pyrimidine rich		unknown	
2	I box	GGATGAGATAAGATTA (C C)	GA-1	unknown	Ueda et al. (1989)
2	I box	GATAAG	GA-1	"enhancer"	Donald and Cashmore (1990)
2	GATA	GATGAGATA (T)	ASF-2 [a]	leaf-specific	Lam and Chua (1989)
3	G box	CACGTGGC (T)	GBF	"enhancer"	Donald and Cashmore (1990)
5		TTAAATAGAGGGCGTAA		unknown	
8	Box II	GA A A A TTGTGPuTAATPuT	GT-1	light regulation	Lam and Chua (1990)
9		A C TTTGTAATGTCAA		unknown	
10		A GAGCCACA		unknown	
12	CAAT box	T ATCCAAC		unknown	
13		GGTTAC		unknown	
15		AGATGAGG	LRF-1 (?)	light regulation	Buzby et al. (1990)
16		T A A TTTGTGTCCGTTAGATG		unknown	
18	"LRE"	CCTTATCAT		unknown	Grob and Stuber (1987)
19	TATA box	C T TTATATAAA	TFIID [a]	transcription initiation	Gasch et al. (1990)
	A-T rich	(varies)		unknown	
	A-T rich	AATATTTTTATT	AT-1	unknown	Datta and Cashmore (1989)
	A-T rich	AAATAGATAAATAAAAACATT	3AF-1 [a]	"enhancer"	Lam et al. (1990)

[a] These genes have been cloned.

Family member specific cis-acting elements

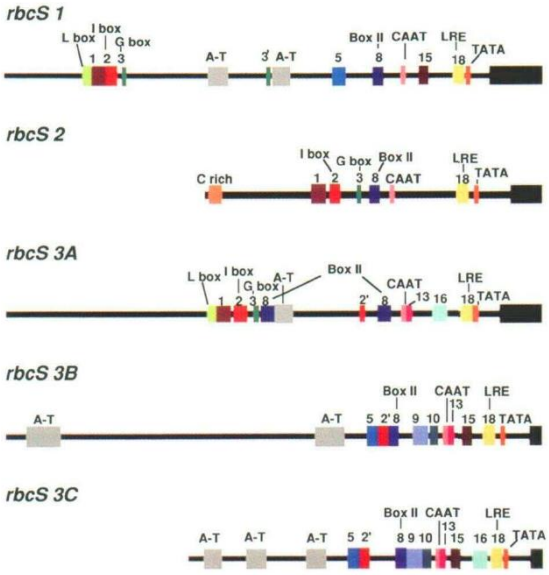


Figure 1. Conserved DNA Sequence Motifs in the 5′ Upstream Regions of the Tomato *rbcS* Genes.

Heavy black lines represent the 5′ upstream region of each gene. Conserved sequences are designated by colored boxes (for key, see Table 1). Solid black boxes at the 3′ end (right) of each upstream region represent the 5′ untranslated region of each gene. The numerical designations for conserved sequences are from Manzara and Gruissem (1988). L-box, I-box, and G-box are the terminology of Ueda et al. (1989), box II is the terminology of Fluhr et al. (1986), and LRE is the terminology of Grob and Stüber (1987).

Transcription Factors Bind to Different Domains of a Promoter in Different Tissues

- **The Plant Cell (1991) 3:1305**
- White ovals = cis-element bound by proteins (TFs)
  - **Binding varies by developmental stage**

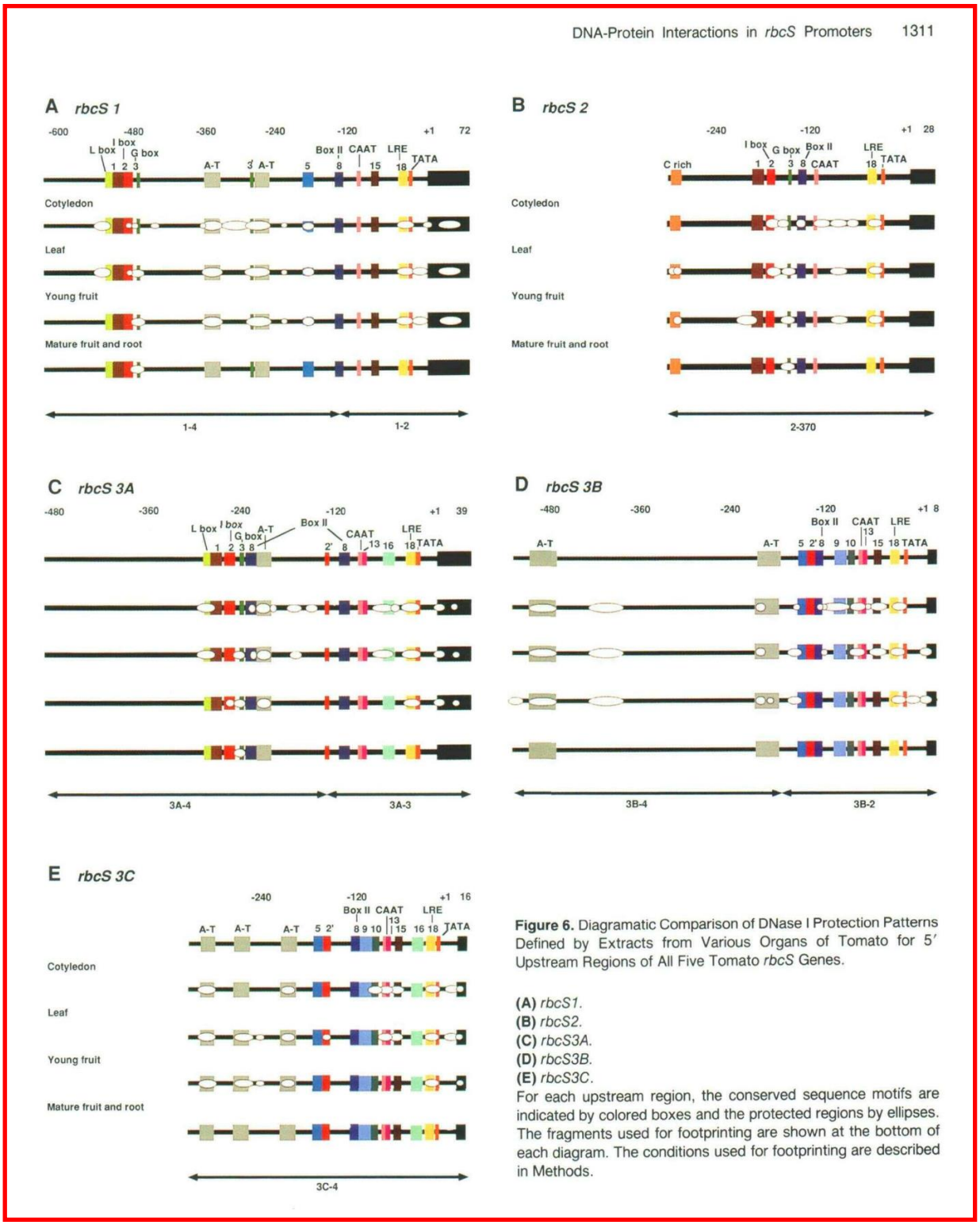


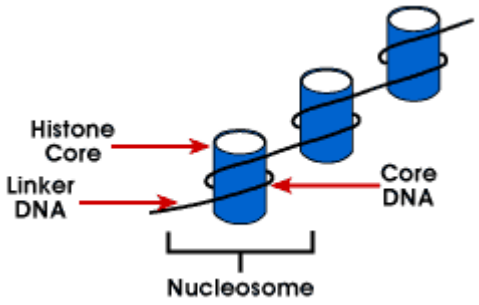
Figure 6. Diagramatic Comparison of DNase I Protection Patterns Defined by Extracts from Various Organs of Tomato for 5′ Upstream Regions of All Five Tomato *rbcS* Genes.

(A) *rbcS1.*
(B) *rbcS2.*
(C) *rbcS3A.*
(D) *rbcS3B.*
(E) *rbcS3C.*
For each upstream region, the conserved sequence motifs are indicated by colored boxes and the protected regions by ellipses. The fragments used for footprinting are shown at the bottom of each diagram. The conditions used for footprinting are described in Methods.

Chromatin Remodeling and Gene Expression

- **Nucleosome Structure is the Normal State**
  - In the nucleus, DNA is packed tightly
    - Histone proteins are organized into a structure called the **histone core**
      - Histone core
        - Two copies of
          - Histone H2a, H2B, H3, H4 each
    - **Core DNA**: ~ 146 bp (invariant_
      - Acts as a repressive state
      - Must be remodeled for active gene expression
    - Histones linked by **linker DNA**
      - Linker DNA ~8-114 bp
    - Packing of DNA into nucleosome
      - Reduces DNA length by six-fold

No remodeling of chromatin; ***No gene expression



Remodeling Process

- **Remodeling is:**
  - **Alterations in chromatin structure that *activates* or *deactivate* gene expression**
  - **Involves transcription factors that actively recruit remodeling complexes**
- May be coupled to DNA replication
- Involves two steps

1. **Histone modification**
   - Specific lysine residues are modified by
     - **Acetylation** [by histone acetylases (HATs)]
       - ***Loosens structure***
       - **Transcription apparatus has access to promoter**
     - **Methylation** (by methylases)
       - ***Tightens structure***
       - **Transcription apparatus has access blocked to promoter**
     - **Ubiquitination** mediated protein degradation
       - Ubiquitin
         - Small protein that is attached to tail of histone protein
       - ***Often marks that protein for degradation***

2. **Recruitment of remodeling complexes**
   - **Swi/Snf family**
     - Contains helicases that **twist DNA on the nucleosomes**
       - **DNA slides on the histones**
         - DNA is more accessible to the transcription factors
       - **Complexes with other proteins to repress a transcriptional unit**

Example of derepressing a transcriptional complex

- **Fernie and Tohge (2015) Location, location, location – no more! The unravelling of chromatin remodeling regulatory aspects of plant metabolic gene clusters. New Phytologist 205:458.**

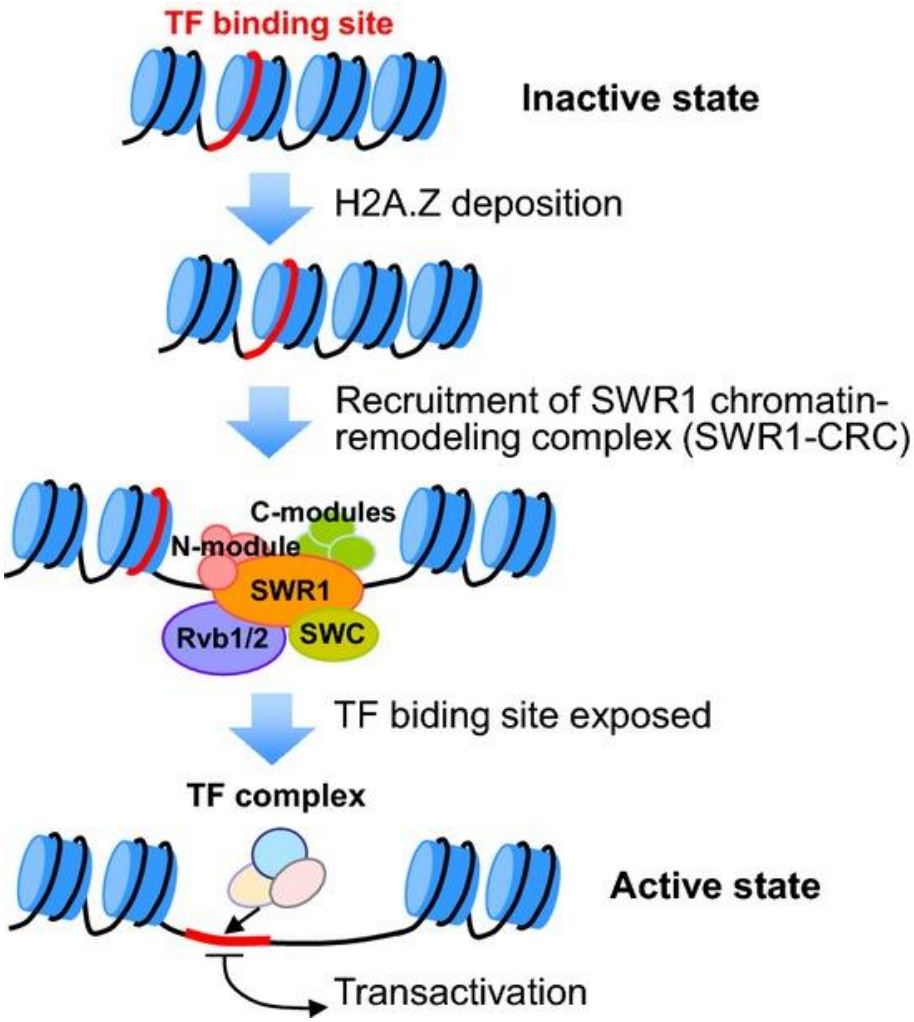


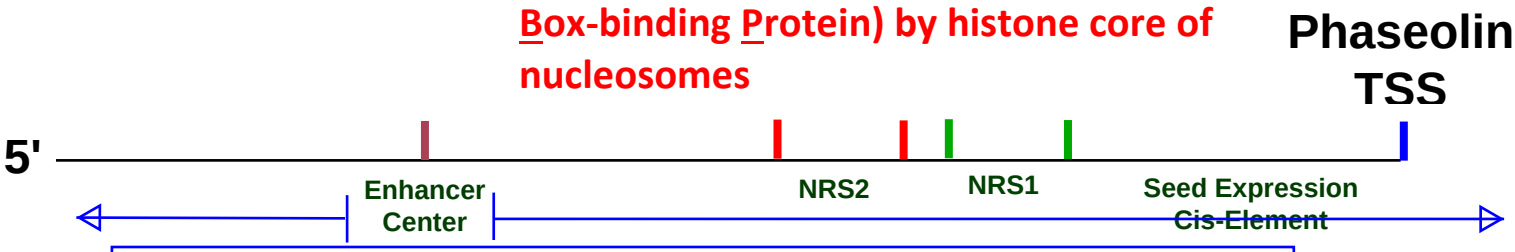
Figure 2. Schematic overview of chromatin remodeling following H2A.Z deposition. N-module and C-module indicate histone and H2A.Z bindings, respectively. Rvb, ruvb-like DNA helicase; SWC, subunit of the SWR1/SRCAP complex.

Steps

1. Novel histone H2A.Z incorporated into histone complex
2. The SWR1 chromatic remodeling complex recruited
3. Upstream region of gene exposed
4. TF complex binds
5. Transcription of gene occurs

The Phaseolin (Phas) Complex in Common Bean

- Phaseolin
  - **Major storage protein in bean seed**
    - Tandemly repeated complex at a single locus
    - Contains three TATA boxes
      - **TATA boxes are protected from the TBP(TATA Box-binding Protein) by histone core of nucleosomes**



Phas Cis-elements

- **Spatial (seed) expression requires 295 bp upstream of transcription start site**
  - Does not include other modulating sequences
- **Negative regulator of premature gene expression**
  - NRS1: -391 to -295
  - NRS2: -518 to -418
- **Matrix attachment region (MAR)**
  - Where DNA binds to the nuclear protein matrix
    - 5' centered at ~ -800 bp (acts as an enhancer)
    - 3' site: centered at ~ +2500 bp

Phaseolin Transcription Activation Steps

Potentiation step

- **Pv-ALF–initiated chromatin remodeling** of the TATA-box domain
  - **A B3-domain transcription factor**
- **Histone modification**
  - **A function of B3-domain transcription factors**

Activation step

- **Abscisic acid** regulated transcription of phaseolin mRNA

PvALF Activation of Phas Gene Expression

- **PvALF Activation**
  - **Protein acts as a transcription factor**
    - **Member of the VP1 and AB13 family of transcription factors**
  - **Modifies chromatin structure of the Phas promoter**
    - **TATA boxes become accessible to TBP**
      - Does not activate transcription by itself
      - Thought to prime the system for ABA-induced gene expression

The Active Phas Complex

1. **Experiments show that all major cis-elements occupied from early to mid-seed maturation**
2. **Model suggest that various cis-elements are occupied differentially**

Transition to the Silent State

- **Protein binding to the promoter decreased after mid-maturation**
- **ROM1**
  - bZIP factor that binds ACGT sequence
  - Probably antagonistic to PvALF
- **PvALF itself may be involved in stage specific developmental repression**
  - May have a role in histone **deacteylation**

Discovering the Transcriptional Regulation of the Phaseolin Gene

- **Li et al. 2001. Plant Molecular Biology 46:121**

A.Nucleosome structure prevents access to the promoter

B. Expression suppressed by negative regulators

C. PvALF recruits chromatic remodeling factors

D. ABA induces other factors to bind to promoter

E. Different tissues have different factors bound to promoter region

F. Promoter access repressed during seed maturation

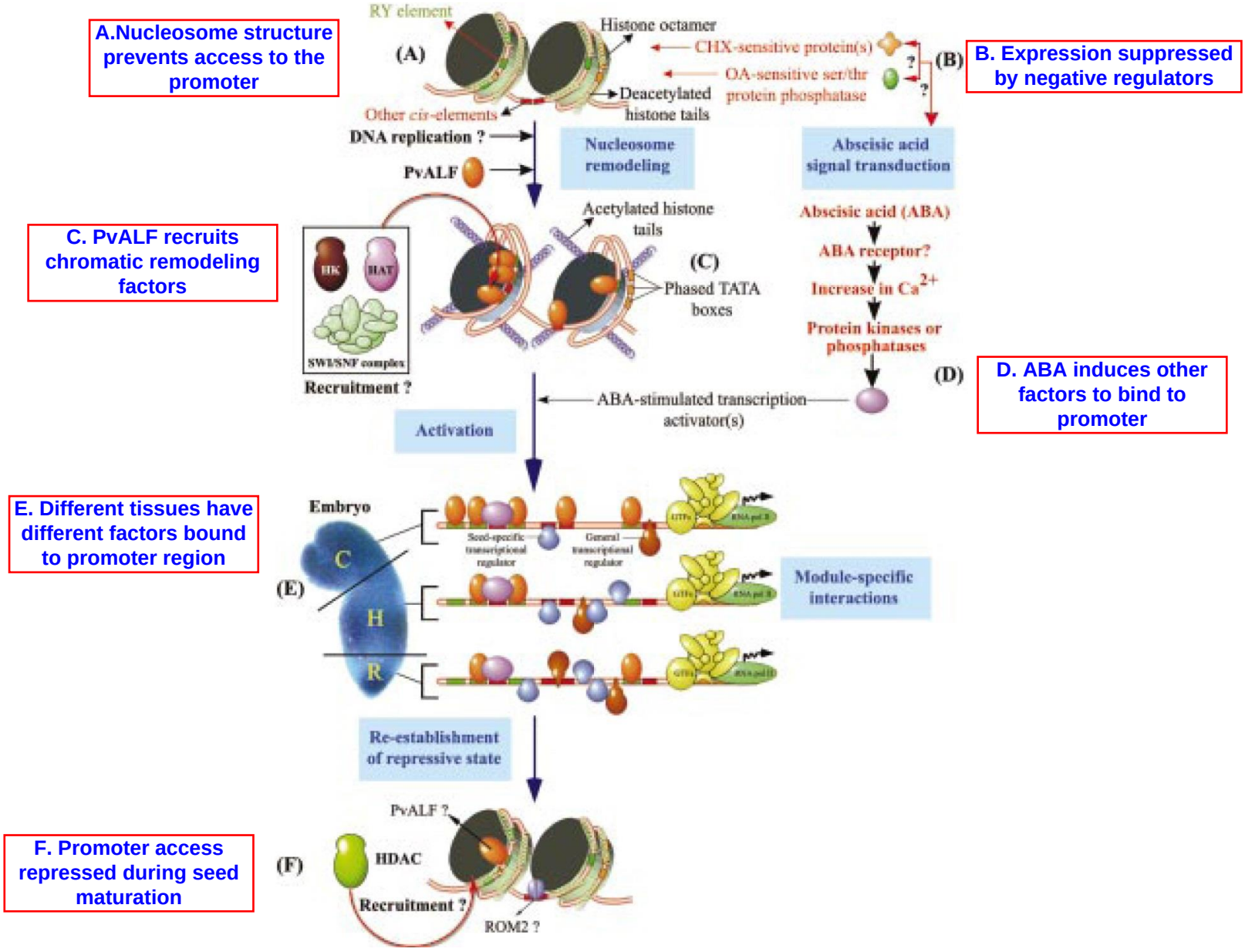


Figure 1. Proposed interactions between chromatin and transcription factors in *phas* activation. A. The **closed chromatin structure** over the *phas* promoter **prevents TBP access** in vegetative tissues. **B.** Non-histone **negative regulators reinforce the repressed status**. **C. PvALF-mediated recruitment of remodeling factors** results in a **relaxed structure** during embryogenesis. **D. ABA-mediated signal transduction** actuates transcription activators that mediate **recruitment of the basal transcription machinery to the phas promoter**. **E. Heterogeneous DNA-protein arrays** yield **module-specific expression** in the embryo (C, cotyledon; H, hypocotyl; R, radicle). **F. The repressive state is re-established during seed maturation.**

Histone Modification Activation of the Phaseolin Promoter

- **Ng et al. 2006. The Plant Cell 18:119**

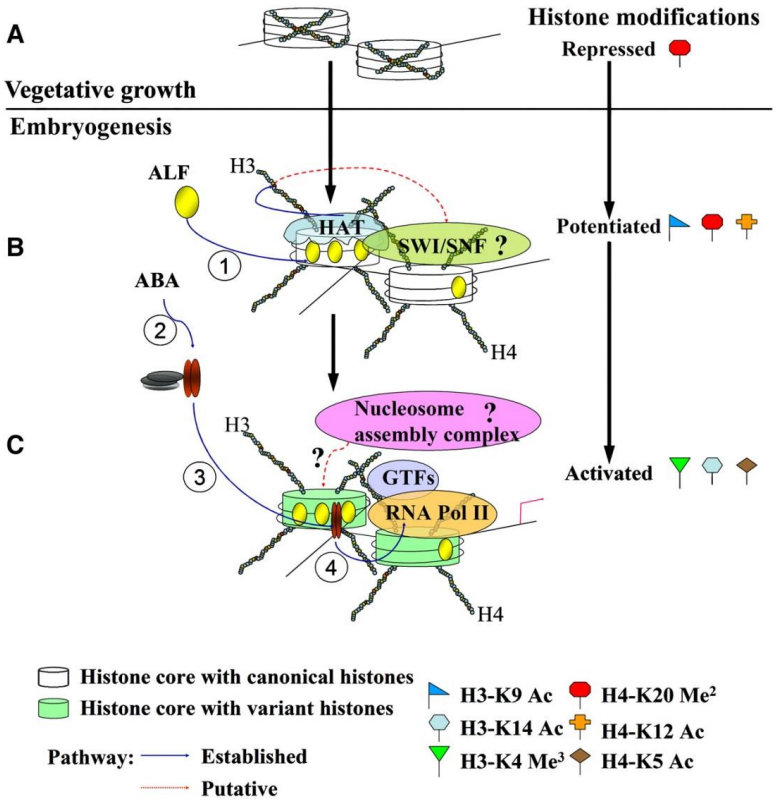


Figure 7. Model Depicting the Sequential Events and Ordered Modification of Chromatin over the phas Promoter during Potentiation and Activation. Histone modifications associated with various *phas* promoter states are shown as symbols at right. Experimentally verified and putative pathways leading to *phas* activation are shown as blue (solid) and red (dotted) lines, respectively. **(A)** In the **repressed state** during vegetative growth, the promoter is envisaged as being heterochromatic, with **nucleosomes** bearing **dimethylated H4-K20**. **(B) ALF-mediated potentiation** of *phas* (**1**), possibly through recruitment of a complex with histone acetyltransferase (HAT) activity; **H3-K9 and H4-K12 are acetylated**. Histone modifications may recruit a chromatin-remodeling complex such as SWI/SNF, resulting in a decrease in histone–DNA interactions. **(C)** Addition of **ABA triggers** the assembly of the **ABA signaling cascade** components (**2**) that **interact with the ABRE** within the *phas* promoter (**3**), leading to the **recruitment of RNA Pol II and GTFs** (**4**). **New histone code modifications** (H3-K4 trimethylation, H3-K14 and H4-K5 acetylation) are **incorporated in the actively transcribed phas chromatin** with the loss of histone H4-K20 dimethylation. During active *phas* transcription, histone displacement and redeposition of variant histones may take place that result in the deposition of new histone modifications at the *phas* chromatin. Although a marked increase in H4-K5 acetylation was evident during activation, a similar increase occurred when only ABA was added (see Figure 4E), suggesting that this modification may reflect events other than activation. The original repressive chromatin status of *phas* is restored at the end of seed maturation, and canonical histones are deposited into the *phas* chromatin through DNA replication during seed germination and vegetative growth.

Model for Phaseolin Activation Using Arabidopsis System

- **Sundaram et al. 2013. The Plant Cell 25:2601**

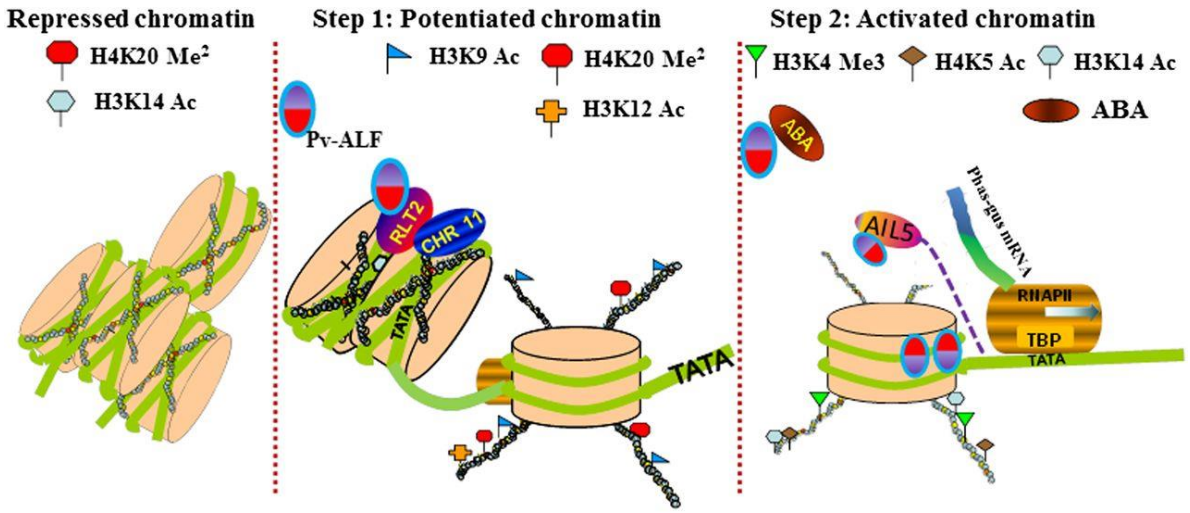


Figure 9. **Model Depicting Sequential Changes in Chromatin Modifications over the phas Promoter during Potentiation and Activation.** In the **repressed state** during vegetative growth, the promoter is repressed by **nucleosomes bearing dimethylated H4-K20**. **Pv-ALF–mediated potentiation** (Step 1) is predicted to **recruit RLT2**, a **component of ISWI chromatin-remodeling complex** that also contains the CHR11-like SWI2/SNF2 ATPase. During this stage, **ordered histone modifications** occur by **demethylation of histone H3-K4, acetylation of H3-K14 and H4-K5, and histone methylation** (Ng et al., 2006). As illustrated in Step 1, this results in **remodeling of the chromatin architecture over the TATA region** of the phas promoter but does not lead to transcriptional activation in the absence of ABA. During the ABA-dependent activation illustrated in Step 2, **Pv-ALF induces AIL5**, which **activates the expression of the phas promoter**.