Isolation and analysis of *MPF2*-like MADS-box genes from Physaleae and characterization of their *cis*-regulatory regions

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Simone Riß

aus Meppen

Köln 2009

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln durchgeführt.

Berichterstatter:	Prof. Dr. Heinz Saedler
	Prof. Dr. Martin Hülskamp
Tag der mündlichen Prüfung:	23.06.2009

Table of contents

1. Introduction

1.1 Morphological innovations

The origin of morphological and functional novelties is a problem that goes back to the early days of evolutionary biology (Darwin 1859). In 1960 Ernst Mayr re-emphasized its importance: ‘‘There are fashionable problems and there are neglected problems in any field of research. The problem of the emergence of evolutionary novelties has undoubtedly been greatly neglected during the past two or three decades, in spite of its importance in the theory of evolution.’’ However, the question how morphological novelties arise remains an unsolved problem in evolutionary biology (Theissen 2006). A promising approach to gain further knowledge on this subject relies on the exploration of developmental and genetic mechanisms that are responsible for these novel structures. For instance, one mechanism that can lead to novelties is heterochrony, i.e. changes in the rate or timing of developmental events (Wake and Roth 1989, Raff *et al.* 1990, Keys *et al.* 1999, Wang *et al.* 1999, Frary *et al.* 2000). Another mechanism is heterotopic expression of existing functions (Kanno *et al.* 2003, He and Saedler 2005), which refers to evolutionary change in spatial patterning. These two mechanisms typically involve either changes in *cis*-regulatory elements (Doebley and Lukens 1998), or changes in *trans*-acting transcriptional regulators (Vrebalov *et al.* 2002, Yu *et al.* 2004).

Changes in gene expression can lead to dramatic morphological changes. One well known example is the *TEOSINTE BRANCHED1* (*TB1*) gene from maize, which has played a major role during the morphological evolution of the modern crop plant *Zea mays ssp. mays* from its wild ancestor *Zea mays ssp. parviglumis*. Diversity analyses at *TB1* loci indicate a partial selective sweep in the *TB1*-promoter region of *Zea mays*, suggesting selection has acted on a regulatory region during domestication (Wang *et al.* 1999). In addition, many studies on animals support the contribution of alterations in the expression of developmental genes in morphological evolution (Carroll *et al.* 2001, Abzhanov *et al.* 2004, Marcellini and Simpson 2006).

1.2 Flower morphology

A key event in the evolution of plants was the emergence of angiosperms, the most diverse group of land plants. In general, the floral organs of most higher eudicots are arranged in four or sometimes more whorls: the calyx (the outer whorl of sepals, usually green organs to shelter the developing inner floral organs), the corolla (the whorl of petals, which are mostly specially shaped and colored to attract pollinators), the androecium (generally one or more whorls of stamens, the male sexual organs) and the gynoecium (one or more carpels, the female sexual organs). Flowers of higher eudicots exhibit a high degree of morphological diversity. In many families, innovations in flower morphology have occurred. Thus flowering plants are an interesting model system to study the molecular mechanisms underlying the origin of morphological novelties.

For example, TCP transcription factors have been shown to play an important role in the determination of flower symmetry, particularly concerning the symmetry of second whorl organs (Luo *et al.* 1996, Luo *et al.* 1999). Changes in these transcription factors can affect petal symmetry of the corolla and generate novel morphological structures. For instance, in the Scrophulariaceae, zygomorphy is under the control of *CYCLOIDEA* (*CYC*; Luo *et al.* 1996) and *DICHOTOMA* genes (*DICH*; Luo *et al.* 1999). The snapdragon *Antirrhinum majus* has zygomorphic flowers, but in case one of these genes is mutated, semi-radial flowers are generated. The *cyc dich* double mutant features radial (actinomorphic) flowers, where all petals have the same identity similar to its close relative *Mohavea confertiflora*, which possesses more actinomorphic flowers and features an altered expression domain of *CYC* (Hileman *et al.* 2003).

1.3 MADS-box genes and the molecular basis of flower morphology

While TCP genes (among other transcription factors) are involved in determining floral symmetry, floral meristem and floral organ identity on the other hand are defined by combinations of MADS-domain transcription factors (Schwarz-Sommer *et al.* 1990, Sommer *et al.* 1990, Yanofsky *et al.* 1990, Theissen *et al.* 2000). MADS-domain proteins operate as homo- or heterodimers (Troebner *et al.* 1992, Davies *et al.* 1996b) or also in higher-order complexes (Egea-Cortines *et al.* 1999, Honma and Goto 2001, Theissen and Saedler 2001) and modifications in these combinations often result in

changes of floral organ identity. To explain the interplay of gene functions the ABC model has been proposed (Schwarz-Sommer *et al.* 1990, Yanofsky *et al.* 1990, Coen and Meyerowitz 1991, Weigel and Meyerowitz 1994). It states that A, B, and C functions are each acting in two neighboring whorls of organs, where their individual and mixed activities determine the fate of organ primordia: A in whorls one and two, B in whorls two and three, and C in whorls three and four. Accordingly, A function alone determines the identity of sepals, but A and B functions act together to control petal identity, B and C functions together are responsible for the identity of stamens, and the C function alone specifies carpel identity. Further, A and C functions act as antagonists, and if one is lacking, the other spreads out to take over the whole flower. Later this model has been extended to the ABCDE model: Additional MADS-box genes have been identified that are required for determination of ovule identity (D-function; Colombo *et al.* 1995) and others that are important for the establishment of sepal, petal, stamen and carpel identity (E-function; Pelaz *et al.* 2000). E-function genes in *Arabidopsis thaliana* are the *SEPALLATA* genes *SEPALLATA1* (*SEP1*) / *AGAMOUS-LIKE 2* (*AGL2*), *SEPALLATA2* (*SEP2*) / *AGAMOUS-LIKE 4* (*AGL4*), *SEPALLATA3* (*SEP3*) / *AGAMOUS-LIKE 9* (*AGL9*) and *SEPALLATA4* (*SEP4*) / *AGAMOUS-LIKE 3* (*AGL3*).

It has been hypothesized that the arrangement of MADS-domain proteins in oligomeric complexes defines which kind of floral organ develops and that changes in composition of the protein complex can result in homeotic substitutions (Theißen *et al.* 2002). Hence, variation in the arrangement of these protein complexes could offer the starting material for the evolution of floral morphological novelties.

1.4 Sepal identity changes

Changes of floral organ identity can be generated by ectopic expression of combinations of *A. majus* MADS-box genes in transgenic *Nicotiana tabaccum* plants: expression of B-function genes in whorl one of tobacco flowers leads to transformation of sepals into petals (Davies *et al.* 1996a).

There are also examples in nature showing that heterotopic expression of transcription factors might be involved in the evolution of morphological novelties. For instance, due to heterotopic expression of B-function genes in the Liliaceae, the perianth is made

of two whorls of petalaloid organs (named tepals), instead of one whorl of petals and one with sepals (Kanno *et al.* 2003).

In several plant species, mutants have been described in which first whorl organs are transformed into foliose (leaf-like) structures.

In the *tunicate* mutant of *Zea mays*, in which the kernels are covered by glumes instead of being naked, *ZMM19*, a MADS-box transcription factor, is ectopically expressed (He *et al.* 2004). In wild type maize, this gene is only expressed in leaves, stem and husks, while in the *tunicate* mutant high levels of transcript are also found in male and female inflorescences. The mutation underlying this expression change involved gene duplication and alterations in the promoter region of *ZMM19*. The gene belongs to the *STMADS11*-superclade of MADS-box genes (He *et al.* 2004, He and Saedler 2005). It has been suggested that members of this clade induce a higher proliferative potential of vegetative tissues (Kim *et al.* 2002). Ectopic expression of *STMADS11*-superclade members *ZMM19, SHORT VEGETATIVE PHASE (SVP),* and *AGAMOUS-LIKE 24 (AGL24)* in transgenic *A. thaliana* transforms sepals into leaf-like organs (He *et al.* 2004).

In *Solanum lycopersicum* (tomato), the gene *LeMADS-MC* affects sepal development and inflorescence determinacy (Vrebalov *et al.* 2002). Its mutant features enlarged sepals. However, in contrast to *ZMM19, LeMADS-MC* is a member of the *SQUAMOSA (SQUA)*-clade of MADS-box genes (Huijser *et al.* 1992). Mutations in the *APETALA1* (*AP1*) gene, which also belongs to the *SQUA*-clade, lead to the conversion of sepals of the first whorl into leaf-like organs (Mandel *et al.* 1992). Apart from its function in determining sepal identity, the A-function gene is also required for normal petal development and for specifying floral meristem identity. Together with LEAFY (LFY), AP1 represses *AGL24* expression in floral meristems of *A. thaliana* to promote floral development (Yu *et al.* 2004).

In summary, ectopic expression of members of the *STMADS11*-superclade of MADS-box genes leads to enlarged leaf-like sepals, and mutations in the *SQUA*-clade MADS-box genes have similar effects. Results from *A. thaliana* suggest that *SQUA*-clade genes can act as negative regulators of *STMADS11*-superclade genes (Yu *et al.* 2004, Liu *et al.* 2007) preventing their expression in floral tissues. Thus the effect observed in the *lemads-mc* mutant may also be a result of ectopic expression of *STMADS11*-

superclade genes, highlighting the potential importance of this gene family in the evolution of first whirl organ morphology.

1.5 A morphological novelty in Solanaceae

In several genera of the family Solanaceae morphological changes of sepals in the first whorl have occurred. Though the majority of the genera have small sepals, several show more elaborate morphologies, like large tubular calyces (e.g. *Brugsmansia, Nicotiana* and *Datura*) or inflated lantern-like calyces, which encapsulate the entire mature fruit (Fig. 1.5.1). In *Physalis*, this trait has been termed "Chinese lantern" or "Inflated-Calyx-Syndrome" (ICS; He *et al.* 2004). So far, species featuring inflated calyces were discovered in at least nine of the 96 genera of the family Solanaceae (D'Arcy 1991), including *Cuatreasia, Exodeconus, Margaranthus, Nicandra, Physalis, Physaliastrum, Physochlaina, Przewalskia*, and *Withania*. Initially, phylogenetic reconstructions suggested multiple origins for this trait in the Solanaceae (Hu and Saedler 2007).

Figure 1.5.1: Calyx development in *Physalis floridana* (Picture: Britta Grosardt)

From left to right: buds, flowers, ICS with developing fruit; scale bar = 1 cm

1.6 Molecular mechanisms underlying ICS in *Physalis*

The molecular mechanisms underlying ICS formation have been studied mainly in the genera *Physalis* and *Withania*. He and Saedler (2005) discovered that the recruitment

of the MADS-domain transcription factor MPF2 from vegetative into a floral context via heterotopic expression enabled the novel morphological trait ICS in *Physalis floridana*.

MPF2 from *P. floridana* and *STMADS16,* its ortholog in *Solanum tuberosum* (potato), are members of the *STMADS11*-superclade of MADS-box genes (He and Saedler 2005). While the expression of *STMADS16* from *S. tuberosum* has been reported to be solely restricted to vegetative tissues (Garcia-Maroto *et al.* 2000, He and Saedler 2005), its ortholog *MPF2* is expressed in vegetative as well as in floral tissues of *P. floridana* (He and Saedler 2005). Transgenic *Physalis* plants, in which *MPF2* expression was knocked down via RNA interference (RNAi), showed dramatically reduced ICS formation, small leaves, and were partially male sterile (He and Saedler 2005, He *et al.* 2007). The latter indicates that *MPF2* function is not only a prerequisite for ICS formation, but is also involved in leaf development and male fertility. Further, it was discovered that pollination or a signal, such as hormones, released after fertilization is required for ICS development. In fact, hormones like gibberellins and cytokinins can trigger ICS formation on depistillated or emasculated flower buds, even prior to pollination in *Physalis* (He and Saedler 2007). *S. tuberosum* does not feature ICS. Yet an ICS-like trait could be generated in transgenic *S. tuberosum* plants by ectopic expression of *STMADS16* and treatment of the developing flower buds with the above-mentioned hormones (He and Saedler 2005). Overexpression of *STMADS16* or *MPF2* in *A. thaliana* does only lead to an increased sepal size, if it is combined with a cytokinin treatment (He and Saedler 2007). According to these studies cytokinin facilitates the transport of MPF2 into the nucleus, where it promotes calyx cell division, whereas gibberellin mediates cell elongation, which is necessary for enlargement of the calyx (He and Saedler 2007).

In summary, at least two factors play an important role in the generation of ICS:

1. Expression of *MPF2*-like genes in sepals

2. Hormone signals generated by fertilization or the developing fruit

Further, three *MPF2*-related functions have been discovered in *Physalis*:

1. Leaf development, 2. ICS formation, 3. Male fertility / normal pollen development

Whether the selective advantage of heterotopic expression in floral organs of *Physalis* was based on the formation of ICS itself and/or on the involvement of *MPF2* in male

fertility is still under discussion and potential "hitch-hiking" effects need to be further investigated. In *Przewalskia tangutica* the ICS seems to decrease the specific weight of the fruits indicating that the trait could facilitate wind dispersal and therefore be of adaptive advantage (Knapp 2002).

1.7 Molecular mechanisms underlying ICS in *Withania*

Species of the genus *Withania*, which also belongs to the Physaleae, possess also ICS-like structures. In contrast, the closely related *Tubocapsicum* species feature only rudimentary calyces. A recent study revealed that two classes of *MPF2*-like genes, termed *MPF2*-like-A and *MPF2*-like-B genes, exist in the tetraploid *Withania*, whereas only one class, the *MPF2*-like-B genes, could be isolated from *Tubocapsicum* species (Khan 2009, Khan *et al.* 2009). The most prominent structural differences between the two classes were a three amino acid deletion in the C-domain and an eight amino acid extension at the C-terminal end of the MPF2-like-A proteins. Only overexpression of *MPF2*-like-A in transgenic *A. thaliana* plants led to elongated sepals. Ectopically expressed *MPF2*-like-B genes did not affect sepal size, indicating that the two proteins are functionally divergent. Also their native expression patterns differ: In *Withania*, the *MPF2*-like-A transcripts were detectable in vegetative and also in floral tissues, where they most likely play a role in ICS formation, whereas *MPF2*-like-B is solely expressed in vegetative tissues. By contrast, in *Tubocapsicum MPF2*-like-B transcripts were found in vegetative and floral tissues. It has been assumed that only MPF2-like-A proteins can trigger sepal growth, while MPF2-like-B proteins cannot, explaining the absence of ICS in *Tubocapsicum*. Accordingly, also the type of *MPF2*-like gene expressed in sepals seems to play a role in ICS formation.

1.8 Loss and gain of ICS

As mentioned previously, several genera in the family of Solanaceae contain species featuring ICS, whereas others do not. Regardless of the ancestral state, it would be interesting to know the changes that are responsible for the loss or gain of this trait. Current knowledge indicates that the following changes in the pathway underlying ICS formation could be involved in gain or loss events:

1. Changes of the spatial or temporal release of hormone signals

2. Changes in MPF2-like proteins leading to altered functional properties, in particular gain / loss of the ability to promote sepal growth

3. Changes of the spatial and temporal expression pattern of *MPF2*-like genes in floral organs by alterations in *cis*-regulatory elements or *trans*-acting factors

1.9 Transcriptional regulation of *MPF2*-like gene expression

This study is primarily concerned with the putative role of *cis*-regulatory elements in differential expression of *MPF2*-like genes in sepals, since this had been suggested by previous findings: In contrast to the highly conserved coding sequences of *MPF2* and *STMADS16*, the putative promoter regions of these genes share only 42% sequence similarity (He and Saedler 2005). The low sequence similarity in the upstream regions might account for the divergent expression patterns in *P. floridana* and *S. tuberosum*. Further, two MEF2- or N10-type CArG-boxes, binding sites for MADS-domain transcription factors, are present in the first 1.3 kb upstream region of *STMADS16*. In contrast, the putative promoter region of *MPF2* does not contain these motifs. Expression of MADS-box genes is often controlled by other MADS-domain proteins (Yu, et al. 2004). Hence another MADS-domain protein might act as a negative regulator by binding to the CArG-boxes in the *STMADS16* promoter and preventing its expression in floral tissues of potato. The previously mentioned SQUA-like proteins (see chapter 1.4) are candidates for such a *trans*-regulator in potato.

Although the differences in the upstream putative promoter region are a prominent feature, additional *cis*-elements controlling expression of *MPF2*-like genes could be also located in introns of the genes. In fact, such intronic elements are known for other MADS-box transcription factor genes like *PLENA* (*PLE*; Davies *et al.* 1999) and *AGAMOUS* (*AG*; Hong *et al.* 2003).

1.10 *MPF2*-like gene expression in Physaleae

In this study, several species from the Physalinae and Iochrominae, two closely related Physaleae subtribes (Olmstead *et al.* 1999, Hu and Saedler 2007), were investigated.

The Physalinae include the genera *Physalis* and *Margaranthus,* which feature ICS, and the genus *Witheringia* (without *Witheringia coccoloboides*, according to Hu and Saedler 2007), which contains non-ICS species. Most *Physalis* species are diploid (e.g. *P. floridana*), although some are tetraploid, e.g. *Physalis minima* and *Physalis peruviana*. The majority of *Physalis* species are native to Mexico, even though *Physalis alkekengi* probably has its origin in the Old World (Codex Aniciae Julianae before 512).

The Iochrominae clade consists of 34 species and six traditionally accepted genera: *Acnistus, Dunalia, Eriolarynx, Iochroma, Saracha* and *Vassobia* (Smith and Baum 2007). Most of the Andean species occur between Columbia and Peru, where the shrubs and small trees grow in scrub or cloud forest between 2200 and 2900 m (Smith *et al.* 2008). Their flowers, which are pollinated by hummingbirds, show an immense diversity (Smith and Baum 2006) and according to published chromosome counts they are probably all diploid (Hunziker 2001 and references therein). The three species used in this study (see Fig. 1.10.1) featured only small calyces, although Iochrominae with larger inflated calyces do exist (e.g. *Iochroma cornifolium*).

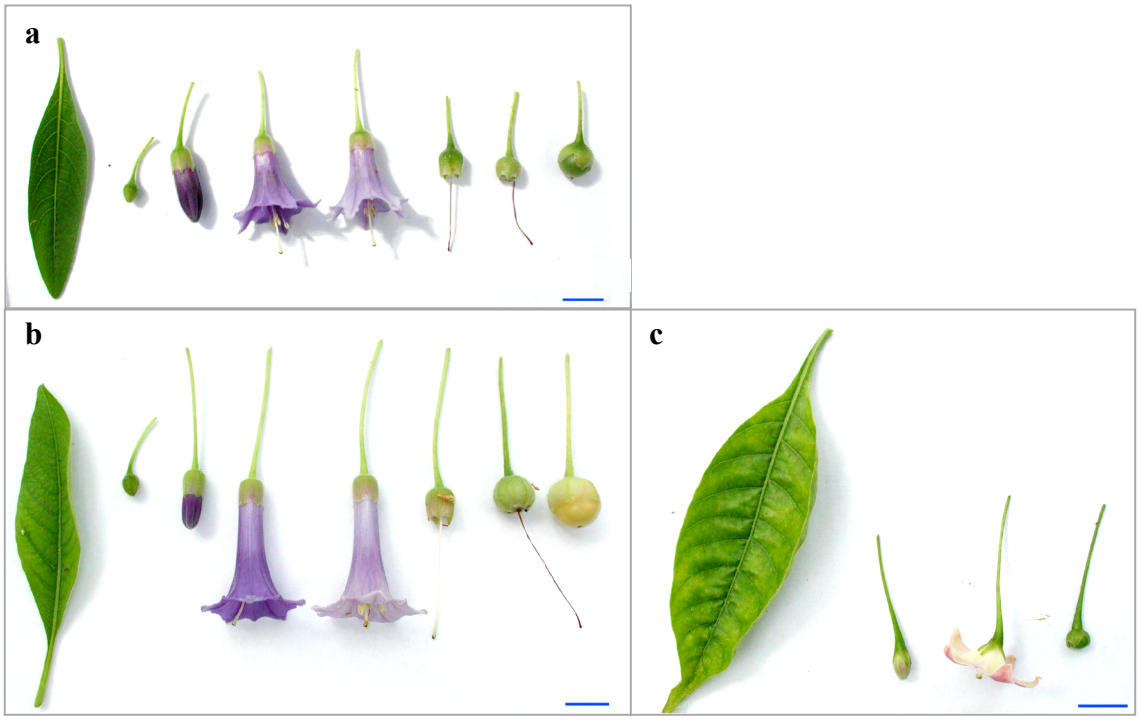


Figure 1.10.1 : Leaves, buds, flowers and fruits from three Iochrominae species

a) *Dunalia fasciculata,* b) *Iochroma australe* and c) *Vassobia breviflora*; scale bar = 1cm

MPF2-like gene expression in calyx tissues has been extensively studied in the Physaleae (see Fig. 1.10.2; Hu and Saedler 2007). All tested species from the genera *Withania* and *Physalis*, which feature ICS, expressed *MPF2*-like genes in their calyces. Further, two of the Iochrominae species, *I. australe* and *D. fasciculata*, which do not form an ICS, also lack *MPF2*-like gene expression in sepals. Interestingly, several species exhibit *MPF2*-like gene expression in sepals (*Capsicum baccatum, Lycianthes biflora, Tubocapsicum anomalum, Witheringia solanacea*, *V. breviflora*), although they produce only tiny calyces that do not react to externally applied hormones.

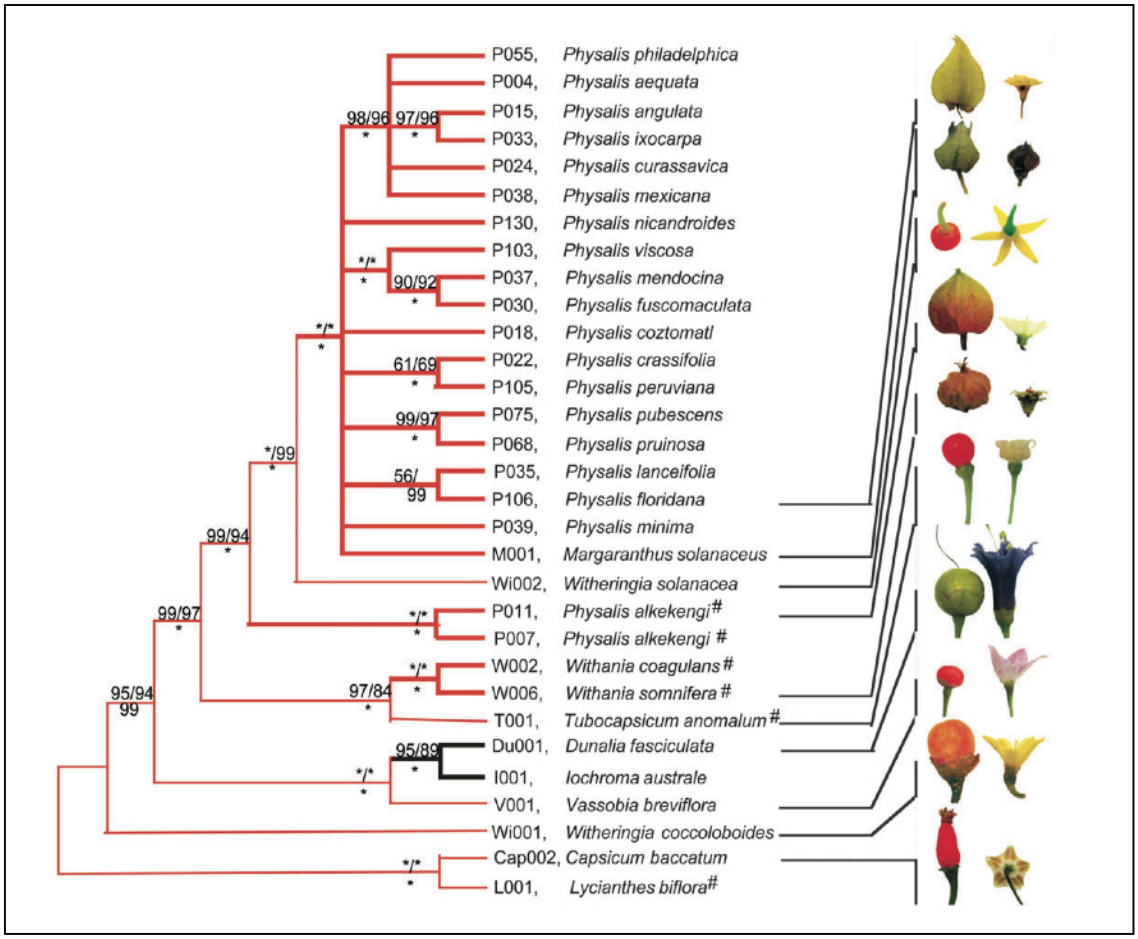


Figure 1.10.2: Phylogeny of the Physaleae (adopted from Hu and Saedler, 2007)

"A strict consensus MP tree with *Capsicum* and *Lycianthes* as outgroups is shown, based on concatenation of the 2 chloroplast sequences *atpB* and *matK* and cDNA sequences of the nuclear *MPF1*- and *MPF2*-like genes. (…) Branches leading to species that express the *MPF2*-like gene in the calyx are shown in red; bold red indicates species displaying ICS. Bold black branches denote species that show no *MPF2*- like gene expression in the calyx and do not develop ICS. The number sign (#) indicates Old World origin; all other species are of New World origin. Calyx phenotypes at flower and fruit stages are shown on the right. The vertical scale bars = 1 cm." (Hu and Saedler, 2007)

These findings, together with data mentioned in the two previous chapters, show that *MPF2*-like gene expression in sepals is necessary, but obviously in many species not sufficient to induce formation of an ICS-like structure. Despite the lack of expression in sepals observed in two of the Iochrominae, it has been suggested that heterotopic expression of *MPF2*-like genes seems to be a plesiomorphic character within the Physaleae and the Capsiceae and that secondary mutations in the ICS pathway might have taken PLACE (Hu and Saedler 2007).

The variability in *MPF2*-like gene expression observed in the Iochrominae was very interesting for this study, since it provided an example for a putative secondary loss of *MPF2*-like gene expression in the calyx, where the mechanisms underlying heterotopic expression can be studied. The nature of these changes, for instance whether alterations in *cis*-regulatory elements play a role, is not known.

1.11 Goals of this thesis

The major goal of this thesis is therefore to provide insights concerning the role of *cis*-regulatory elements in differential transcriptional regulation of *MPF2*-like MADS-box transcription factors in Physaleae. The identification of important regulatory regions and further the detection of changes in these, leading to loss or gain of floral expression (especially in sepals), would improve current knowledge about the mechanisms underlying ICS. For this purpose, upstream and intronic sequences of *MPF2*-like genes were isolated and analyzed *in silico*. In addition to that, transgenic *Arabidopsis* harboring promoter::*GUS* fusion constructs were produced to allow conclusions regarding functionality.

An additional investigation was the isolation of cDNAs and Southern blot analysis of *MPF2*-like genes from Physaleae. Assumptions concerning their ability to trigger calyx growth, based on present phylogenetic data, will be discussed attempting to give possible explanations for the absence of ICS in *V. breviflora*.

2. Materials and Methods

2.1 Chemicals, antibiotics, kits, enzymes, primers and radioisotopes

Chemicals

The chemical products of analytical quality were purchased from the following companies: AppliChem (Darmstadt), Biomol GmbH (Hamburg), Bio-Rad (Munich), Biozym (Hessisch Oldendorf), Clontech (Heidelberg), Merck (Darmstadt), Promega (Mannheim), Roche (Mannheim), Roth (Karlsruhe), Serva (Heidelberg) and Sigma-Aldrich (Munich).

Antibiotics

All antibiotics were ordered from Duchefa (Haarlem, The Netherlands).

Kits

The kits were drawn from Macherey-Nagel (Düren), peqlab (Erlangen), Roche (Mannheim) and Qiagen (Hilden).

Enzymes

The enzymes were ordered from New England Biolabs (Schwalmbach), Fermentas (St. Leon-Rot), Roche (Mannheim), Takara Bio Inc. (Otsu, Japan), Ampliqon (Hamburg), Finnzymes (Espoo, Finland) and Invitrogen (Karlsruhe). All enzymatic reactions were performed according to the manufacturer's instruction manuals.

Primer synthesis

All primers were synthesized by Invitrogen (Karlsruhe).

Radioisotopes

The radioisotope [α-^{32}P]-dCTP (110 TBq/mmol) was purchased from Hartmann Analytic (Braunschweig).

2.2 Buffers, media and solutions

The buffers, media and solutions used in the experiments were prepared according to Sambrook and Russell (2001) unless otherwise noted below.

Partial GUS histochemical buffer
100 mM $NaPO_4$ (pH 7.0)
0.5 mM $K_3Fe(CN)_6$
0.5 mM $K_4Fe(CN)_6$
10 mM EDTA
0.1% Triton X-100 (v/v)
Store at 4 °C, dark

GUS histochemical buffer (10 ml)

X-Gluc stock in DMF (50 mg/ml)	120 µl
Partial GUS buffer	8 ml
Methanol	2 ml

Prepare fresh

CTAB buffer (for extraction of genomic DNA)
100 mM Tris pH 8.0,
1.4 M NaCl
20 mM EDTA
2% CTAB (w/v)
1% PVP (w/v)

YEB Medium
0.5% Beef extract (w/v)
0.1% Yeast extract (w/v)
0.5% Peptone (w/v)
0.5% Sucrose (w/v)
Adjust pH 7.5
Add 1 M $MgSO_4$ (10 ml/l) after autoclaving
For solid medium add Agar (15 g/l) prior autoclaving

Infiltration Medium (for *Arabidopsis* transformation)
1/2 x Murashige & Skoog medium including Gamborg B5 vitamins
1/2 x Gamborg's vitamin solution
5% Sucrose (w/v)
0.044 µM 6-benzylaminopurine (stock: 1 mg/ml in DMSO)
0.005% Surfactant SILWET L-77 (v/v)

MS medium (for transient transformation of Solanaceae)
3% Sucrose (w/v)
1 x Murashige & Skoog medium including Gamborg B5 vitamins
0.45% MES (w/v)
Adjust to pH 5.7

2.3 Online tools and scientific software

Table 2.3.1: Online tools for literature search and sequence analyses

Tool	Application	URL	Reference
Primer3	Primer design	http://frodo.wi.mit.edu/primer3/input.htm	(Rozen and Skaletsky 2000)
PlantGDB: GeneSeqer	Exon / intron structure	http://www.plantgdb.org/	(Usuka *et al.* 2000)
NCBI BLAST	Sequence information,	http://blast.ncbi.nlm.nih.gov/	(Altschul *et al.* 1990)
PubMed	Literature search	http://www.ncbi.nlm.nih.gov/	-
Credo: DIALIGN	Global alignment tool, detection of conserved motifs	http://bioinfo.mpiz-koeln.mpg.de/credo_test/credo.htm	(Morgenstern 1999)
DIALIGN and Chaos	Alignments of genomic sequences	http://dialign.gobics.de/	(Brudno *et al.* 2004)
Mulan and multiTF	Local multiple sequence alignments, transcription factor binding sites	http://mulan.dcode.org/	(Ovcharenko *et al.* 2005)
PLACE	*cis*-regulatory elements database	http://www.dna.affrc.go.jp/PLACE/	(Higo *et al.* 1999)

Table 2.3.2: Scientific software for sequence and image analysis

Software	Application	Source/Reference
MacVector 9.0	sequence editing, alignment	Accelrys Inc. (Cambridge, UK)
AssemblyLIGN	sequence editing, alignment	Accelrys Inc.
Image Quant	quantification	Molecular Dynamics (Krefeld)
PAUP4.0b10	phylogenetic reconstruction	(Swofford 2002)

2.4 Bacteria and plasmids

Escherichia coli strains:

- DH5α (Genotype: fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17)
- DH10ß (Genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara,leu)7697 araD139 galU galK nupG rpsL λ-)

Agrobacterium tumefaciens strains:

- GV3101 pMP90 (Marker genes: Rifampicin, Gentamycin)
- LBA4404 pAL4404 (Marker genes: Rifampicin, Spectinomycin/Streptomycin)

Table 2.4.1 Plasmids used for transformation

Plasmid	Application and resistance	Reference/Source
pGEM-T Easy	Cloning of PCR products Ampicillin	Promega (Mannheim)
pGPTV-bar	Plant transformation vector Kanamycin, Basta	(Becker *et al.* 1992)
p27IGUS	Plant transformation vector Kanamycin, Spectinomycin/Streptomycin	(Shang *et al.* 2007)

2.5 Plant materials

All Solanaceae accessions (see Supplement, Table 8.1.6) were obtained from seed banks worldwide and seeds were stored in the seed bank at the Max-Planck-Institute for Plant Breeding Research (MPIZ).
Columbia ecotype (Col-0) plants were used for the generation of transgenic *A. thaliana*.

2.6 Published sequences

Published sequences, containing *MPF2*-like partial and full-length mRNAs or genomic loci, were downloaded from NCBI (see Supplement, Tables 8.1.1, 8.1.2 and 8.1.3).

2.7 Plant growth conditions

All Solanaceae plants were grown in greenhouses on soil under long day conditions (16 h light, 8 h dark) at 20 to 26 °C by MPIZ gardeners.

Further, all *A. thaliana* Col-0 plants were cultivated on soil in greenhouses at 22 °C. The wild-type was first kept under standard short day conditions (8 h light, 16 h dark) and later transferred to long day greenhouse conditions (16 h light, 8 h dark), while transgenic plants were always kept under long day conditions.

2.8 Isolation of genomic DNA from plant tissues

Genomic DNA for Southern blotting was extracted from frozen leaves with the DNeasy Plant Mini Kit (Qiagen).

Total genomic DNA for all other purposes including 'genome walking' was isolated using a modified CTAB protocol based on the original method by Doyle and Doyle (1987): Young leaves were grounded in liquid nitrogen. Then 1 g plant material was incubated in 15 ml of CTAB extraction buffer at 65 °C for 20 min, mixed with 15 ml of chloroform and centrifuged at 5000 rpm for 10 min. The upper phase was transferred to a new tube and the nucleic acids were precipitated by addition of 30 ml isopropanol. After precipitation at -20 °C overnight the nucleic acids were centrifuged at 5000 rpm for 30 min. Afterwards the pellet was resolved in 1 ml TE supplemented with 20 µg/ml DNase-free RNase (Qiagen) and incubated for 30 min at 37 °C to remove RNA contaminations. An equal volume of phenol-chloroform-isoamyl alcohol (24:24:1) was added and after centrifugation at 5000 rpm for 5 min the upper phase was transferred into a new tube. In the next step 1 Vol of chloroform was added and after centrifugation at 5000 rpm for 5 min the aqueous phase was pipetted into a new tube. Then the DNA was precipitated with 100 µl 3 M sodium acetate (pH 5.2) and 700 µl isopropanol. After centrifugation for 1 h at 5000 rpm and 4 °C the supernatant was removed and the pellet washed with 70% ethanol twice and finally resolved in TE buffer.

2.9 Total RNA isolation from plant tissue

To gain RNA for RT-PCR, samples of young leaves and floral organs at different stages were collected separately by using forceps and razorblades, quickly frozen in liquid nitrogen, and stored at –80 °C for nucleic acids extraction. Stamens from open flowers were harvested into 0.3 M mannitol buffer to collect mature pollen grains, as described by Honys and Twell (2003).

Biomol reagent (Biomol GmbH) was used to isolate ribonucleic acids from young leaves for 5' and 3'RACE. To obtain total RNA for RT-PCR the RNeasy Mini Kit (Qiagen) was applied according to the manufacturer's instructions. DNA contaminations were removed by treatment with RNase-free DNase (Roche).

2.10 Isolation of plasmid DNA from bacteria

Plasmid DNA from *E. coli* was isolated with the peqGOLD Plasmid Miniprep Kit I (peqlab) in case of small (<10 kb), high copy number plasmids and the Miniprep Plasmid Purification Kit (Qiagen) was applied for the isolation of large (> 10 kb) constructs.

2.11 Purification of PCR products

All PCR fragments for cloning and sequencing were purified using the NucleoSpin Extract II kit (Macherey-Nagel) according to the manufacturer's instructions. This was either done directly or after fractioning the products on 1.2% agarose gels.

2.12 Standard PCR reaction

All PCR reactions were performed in a T3000 thermocycler (Biometra). For standard reactions (e.g. colony PCR) Ampliqon Taq DNA Polymerase (Ampliqon) was the enzyme of choice.

However, for high accuracy PCR reactions either the Expand High-Fidelity PCR System (Roche) or the Phusion High-Fidelity DNA Polymerase (Finnzymes) were used, while LA Taq DNA Polymerase (Takara) was chosen for amplification of larger fragments.

All PCR reactions were performed as described for each enzyme in the manufacturers instruction manuals. A standard PCR reaction protocol for Ampliqon Taq DNA Polymerase and the Roche Expand High-Fidelity PCR System is given:

Standard PCR reaction

Reagent	Final concentration
dNTPs	200 µM (of each dNTP)
Primer 1	200 nM
Primer 2	200 nM
10x PCR Puffer	1x
DNA template	variable
Distilled H_2O	---- (ad up to 50 µl)
DNA Polymerase	1-5 units
Total volume	50 µl

Annealing Temperature:	55-65 °C
Elongation Time:	1 min/kb
Number of cycles:	30-35

2.13 Isolation of cDNAs by 5' and 3'RACE

cDNAs were isolated by RACE (rapid amplification of cDNA ends) using the 5'/3'Race Kit, 2nd generation, for rapid amplification of cDNA ends (Roche). All PCR reactions were performed with the Expand High-Fidelity PCR System (Roche). For 5'RACE either universal primers or (in the case of *Capsicum* and *Physalis* species) gene-specific primers were used. All primers were located in the highly conserved MADS box. The universal primers P1 and P2 were designed based on the MADS-boxes of *MPF2* and *STMADS16* and kindly provided by Chaoying He. For 3'RACE primers in the 5'UTR and the MADS domain were applied. Further details on primers can be found in the supplement (see Table 8.1.7).

2.14 Isolation of upstream sequences by RAGE

The genomic sequence upstream of the MADS-box (containing the 5'UTR and the putative promoter sequence) was isolated by RAGE (rapid amplification of genomic DNA ends) according to the Universal Genome Walker kit (Clontech). Genomic DNA was completely digested with DraI, EcoRV, ScaI, HpaI, HaeIII, PvuII or StuI (NEB),

respectively, and the blunt-ended DNA fragments were ligated to adaptors using T4 DNA ligase (Roche). The first PCR was carried out using an adaptor primer and a gene-specific primer, while the nested PCR was done with a nested adaptor primer and a nested gene-specific primer. All PCR reactions were performed with the Expand High-Fidelity PCR System (Roche) polymerase. The adaptor primers are described in the manual (Clontech), while the gene-specific primers are listed in the supplement (see Table 8.1.7). After each genome walking step new specific primers were designed. After assembly of the partial upstream sequences with the programs AssemblyLIGN (Accelrys Inc.) and MacVector (MacVector), the complete sequence was verified by PCR amplification of a fragment spanning the entire sequence.

2.15 Isolation of intronic sequences by long template PCR

Further, the sequences of first introns were also isolated by PCR using LA Taq DNA Polymerase (Takara). Therefore, the region ranging from the 5' end of the isolated upstream sequence to the start of the second exon (I-region) was amplified using gene-specific primers located in the upstream sequence and in the I-region. Detailed information on the primers used for amplification is listed in the supplement section (see Table 8.1.7).

2.16 Sequencing

All DNA sequencing reactions were carried out by ADIS (Automatic DNA Isolation and Sequencing), the core facility for DNA-related technology at the Max-Planck-Institute for Plant Breeding Research, Cologne.

2.17 Southern blot analysis

Southern blot analysis was performed in cooperation with two group-members. The individual contributions were as follows: The DNA templates for radioactive oligo labeling were provided by Ramzan Khan. My own contribution included the preparation of DNA samples (DNA isolation and digestion) and membranes (blotting and stripping) and the hybridization with the probe for *MPF2*-like-A genes, while the hybridization with the probe for *MPF2*-like-B genes was performed by Chaoying He.

Blotting

Highly pure genomic DNA, isolated with the DNeasy Plant Mini Kit (Qiagen), was used for Southern blotting. The DNA samples (8 - 10 µg) from *W. coagulans* (W002), *W. somnifera* (W006 and W007), *T. anomalum* (T001), *V. breviflora* (V001) and *I. australe* (I001) were digested for 16 h at 37 °C in 200 µl volumes with the restriction enzymes BamHI, EcoRI and KpnI (New England Biolabs), respectively. Afterwards the samples were precipitated by adding 20 µl 3 M sodium acetate and 200 µl isopropanol. After centrifugation for 45 min at 15 g, pellets were washed with 70% (v/v) ethanol and dissolved in 50 µl TE pH 8.0. The fragments were separated overnight on a 0.7% agarose gel. After depurination of the DNA in 0.25 M HCL and denaturation in 0.4 M NaOH the DNA was transferred to a positively charged Hybond N(+) nylon membrane (Amersham) by downward alkaline blotting (Koetsier et al. 1993). After 5 h the membrane was neutralized for 1 min in 5x SSPE and 2x SSPE, before the DNA was immobilized by UV crosslinking in a UV Stratalinker 2400 (Stratagene).

Radioactive labeling and hybrization

Two PCR fragments of approximately 300 bp (containing the C-terminus and 3'UTR of W006 *MPF2*-like-A and *MPF2*-like-B cDNAs, respectively) were amplified from plasmids and provided by Ramzan Khan (for primers see Supplement, Table 8.1.7). The fragments were radioactively labeled with [$\propto$-^{32}P]-dCTP by random oligonucleotide-primed synthesis: 200 ng DNA template were diluted in 35 µl H_20. After denaturation at 95 °C for 10 min the mixture was quickly chilled on ice. Then 5 µl 10x oligo buffer, 3 µl BSA (2 mg/ml, NEB), 2 µl Klenow enzyme (2 u/µl, Roche) and 5 µl [$\propto$-32P]-dCTP (110 TBq/mmol) were added. After 2 h the radioactive labeled product was purified using the NucleoSpin Extract II kit (Macherey-Nagel).

Hybridizations were performed as described previously by He et al. (2002) with minor modifications. Hybridization buffer was supplemented with denatured and fragmented herring sperm DNA before use. After prehybridization for 3 h at 68 °C in hybridization buffer, hybridization was done in the same buffer containing the probe for 16 h at 68 °C. The filters were washed with 2x SSC, 0.1% SDS (68 °C, 20 min), then with 1xSSC, 0.1% SDS (68 °C, 20 min) and finally with 0.5x SSC, 0.l% SDS (68 °C, 5-20 min). The filters were exposed to a Storage Phosphor Screen (Molecular

Dynamics), and signals were read with a Typhoon 8600 Phosphor Imager (Amersham).

Stripping

Probes were stripped by pouring two times boiling 0.1% SDS on the membrane and shaking it until cooling down to room temperature.

2.18 RT-PCR analyses

Reverse transcription polymerase chain reaction (RT-PCR) was performed to study expression patterns of *MPF2*-like genes in mature pollen of several solanaceous species and further vegetative and floral tissues (at different developmental stages) of *I. australe* and *V. breviflora*. Gene-specific primers for the amplification of *MPF2*-like genes were specifically designed from *MPF2*-like cDNAs. Primer specificity was confirmed by sequencing. The constitutively expressed 18S rRNA was used as a control. Shuping Xing provided the 18S rRNA primers (for primer sequences see Supplement, Table 8.1.7). First-strand cDNA synthesis was carried out using Superscript II (Invitrogen) with 2 μg of total RNA in a 20 μl volume. The following PCR conditions were used: *MPF2*-like: 94 °C for 30 s, 59 °C for 30 s, 72 °C for 30 s, 35 cycles; 18S rRNA: 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, 15 cycles. The reactions were analyzed by agarose gel electrophoresis and documented with a Typhoon 8600 Phosphor Imager (Amersham, Sunnyvale, USA). Pictures were analyzed with Image Quant (Molecular Dynamics, Sunnyvale, USA). Two technical replicates were performed.

2.19 Bioinformatic and phylogenetic analyses

The phylogenetic reconstructions were done using the methods Maximum Parsimony (cDNA and protein sequences) and Maximum likelihood (upstream sequences) from the PAUP4.0b10 software package. The settings for the calculation of Maximum Parsimony trees can be found in Hu and Saedler (2007), while the following settings were used for the calculation of Maximum likelihood trees: Number of trees held at each step during stepwise addition = 2, Branch-Swapping algorithm = tree-bisection reconnection (TBR), the ML settings correspond to the HKY85 model.

The isolated upstream sequences were further analyzed using the programs listed in Tables 2.3.1 and 2.3.2. Putative transcription initiation sites were deduced from the sequences of the longest cDNA fragments obtained by 5'RACE. This was done with the help of the online program Geneseqer. DIALIGN and Mulan were applied to find conserved regions in promoter and first intron sequences. Putative *cis*-regulatory elements in the promoters were predicted using PLACE and multiTF.

2.20 *MPF2*-like promoter::*GUS* constructs

For functional analysis of the isolated sequences *GUS* fusion constructs were made.
Although the included upstream sequence proportions represented always only the putative promoter regions of the genes, the constructs will be termed promoter::*GUS* fusion constructs for simplicity.
Four different types of promoter::*GUS* constructs were made (see Fig. 2.20.1): In type 1, termed short promoter constructs (*pMPF2*-like_sh::*GUS*) only part of the isolated upstream sequence is inserted in front of the bacterial β-glucuronidase gene (*uidA*), while in the type 2 constructs, called extended promoter constructs (*pMPF2*-like_ex::*GUS*) almost the entire isolated upstream sequence is included. In type 3, the short promoter + first intron constructs (*pMPF2*-like_sh-I::*GUS*), and in type 4, the extended promoter + first intron constructs (*pMPF2*-like_ex-I::*GUS*), the short or the extended upstream region plus the MADS-box and the first intron in the coding region are inserted before the *GUS* reporter gene.
The upstream and intronic sequences were amplified with Phusion High-Fidelity DNA Polymerase (Finnzymes) using primers containing artificial restriction sites (*D. fasciculata, I. australe, V. breviflora, P. alkekengi, W. coccoloboides*: HindIII; *P. lanceifolia*: XmaI; *W. solanacea*: ArvII, which created compatible ends for XbaI digested vector). The primers are listed in the supplement section (see Table 8.1.7). After digestion the fragments were cloned into the pGPTV-bar vector, which contained the *uidA* gene. All constructs contained the 5' UTR region of the gene. In promoter constructs without intron, the ATG codon of the MADS-box was mutated by PCR into TTG and the 3’ end of the MADS-box was fused to the reporter gene. For construction of *GUS* fusion constructs containing the promoter, the MADS-box and the first intron, the whole region was amplified and subcloned into pGPTV-bar. In these constructs the

ATG was not mutated, but the 3' end of second exon was fused in frame to the reporter gene.

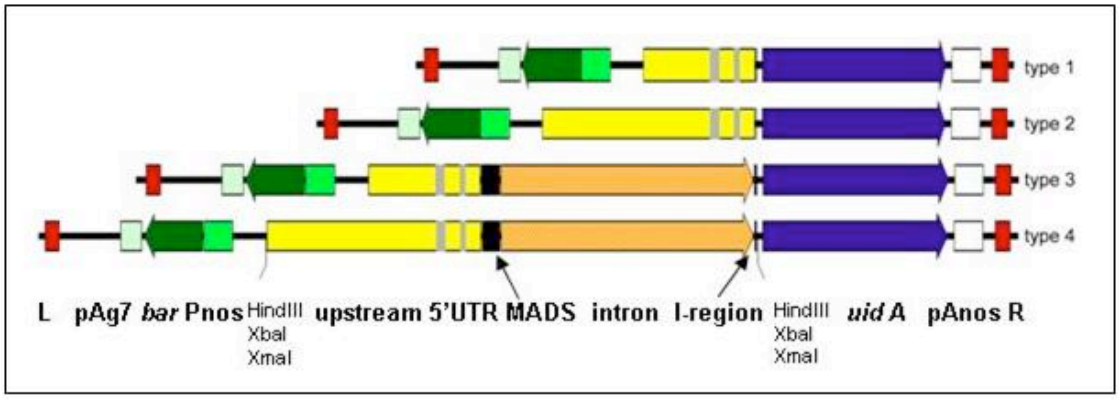


Figure 2.20.1: Schematic diagrams of the four types of *GUS* fusion constructs used for transformation of *A. thaliana* (proportions based on *VbrM201_1*)

The putative regulatory regions were inserted into the plant binary vector pGPTV-bar (Becker et al. 1992) via the restriction sites HindIII, XbaI or XmaI. The upstream regions (containing the putative promoter regions) are indicated by yellow boxes and the exons in the 5'UTR are marked as grey boxes. The MADS-box and the start of the I-region are shown as black boxes, while the intron between them is represented by an orange arrow. The GUS reporter gene *uidA* is indicated by a blue arrow. The Basta selection cassette, containing the *bar* gene, is shown in green color. The T-DNA borders are represented by red boxes. Arrows indicate direction of transcription; R, right T-DNA border; L, left T-DNA border; pAg7, agropine synthase polyadenylation signal; Pnos, promoter of the nopaline synthase gene; pAnos, nopaline synthase polyadenylation signal

1) short promoter (0.9 – 1.8 kb) = *pMPF2*-like_sh::*GUS*
2) extended promoter (1.3 to 3.6 kb) = *pMPF2*-like _ex::*GUS*
3) short promoter + MADS-box + first intron = *pMPF2*-like_sh-I::*GUS*
4) extended promoter + MADS-box + first intron = *pMPF2*-like_ex-I::*GUS*

To keep the Iochrominae expression constructs as similar as possible, the same universal primers were used for all three species - except two types of reverse primers in the MADS domain. Constructs from different Iochrominae species often differed dramatically in size due to sequence differences (e.g. large indels, see Fig. 3.11.1) between these species.

All constructs are listed in Table 2.20.2.

Table 2.20.2: promoter::*GUS* fusion constructs for analysis in *A. thaliana**

Species	Gene/Allele**	Type	Promoter construct	Upstream (in bp)***	Intron ****
I. australe	*IauM201_2*	1	*pIauM201_2*_sh::*GUS*	1800	-
	IauM201_2	2	*pIauM201_2*_ex::*GUS*	2833	-
	IauM201_1	1	*pIauM201_1*_sh::*GUS*	1103	-
	IauM201_1	2	*pIauM201_1*_ex::*GUS*	2121	-
	IauM201_1	3	*pIauM201_1*_sh-I::*GUS*	1103	+
	IauM201_1	4	*pIauM201_1*_ex-I::*GUS*	2121	+
D. fasciculata	*DfaM201_2*	1	*pDfaM201_2*_sh::*GUS*	1796	-
	DfaM201_2	2	*pDfaM201_2*_ex::*GUS*	2831	-
	DfaM201_1	1	*pDfaM201_1*_sh::*GUS*	1106	-
	DfaM201_1	2	*pDfaM201_1*_ex::*GUS*	3675	-
	DfaM201_1	3	*pDfaM201_1*_sh-I::*GUS*	1106	+
	DfaM201_1	4	*pDfaM201_1*_ex-I::*GUS*	3675	+
V. breviflora	*VbrM201_1*	1	*pVbrM201_1*_sh::*GUS*	1022	-
	VbrM201_1	2	*pVbrM201_1*_ex::*GUS*	2100	-
	VbrM201_1	3	*pVbrM201_1*_sh-I::*GUS*	1022	+
	VbrM201_1	4	*pVbrM201_1*_ex-I::*GUS*	2100	+
	VbrM201_2	1	*pVbrM201_2*_sh::*GUS*	951	-
	VbrM201_2	2	*pVbrM201_2*_ex::*GUS*	1996	-
	VbrM201_2	3	*pVbrM201_2*_sh-I::*GUS*	951	+
	VbrM201_2	4	*pVbrM201_2*_ex-I::*GUS*	1996	+
P. alkekengi	*PalM211_1*	1	*pPalM211_1*_sh::*GUS*	1124	-
	PalM211_1	2	*pPalM211_1*_ex::*GUS*	2586	-
	PalM211_1	3	*pPalM211_1*_sh-I::*GUS*	1124	+
	PalM211_1	4	*pPalM211_1*_ex-I::*GUS*	2586	+
W. solanacea	*WIsoM202_1*	2	*pWIsoM202_1*_ex::*GUS*	2916	-
	WIsoM202_1	4	*pWIsoM202_1*_ex-I::*GUS*	2916	+
P. lanceifolia	*PlaM235_1*	2	*pPlaM235_1*_ex::*GUS*	2219	-
	PlaM235_1	3	*pPlaM235_1*_sh-I::*GUS*	1079	+
W. coccoloboides	*WIcoM201_1*	2	*pWIcoM201_1*_ex::*GUS*	1282	-

* plasmid backbone (all constructs): pGPTV-bar
** for nomenclature see chapter 3.1
*** amount of 5'upstream sequence inserted in front of the *uidA* gene
**** MADS-box and first intron included (+ = yes, - = no)

2.21 Generation of transgenic *Arabidopsis thaliana*

Transgenic *A. thaliana* Col-0 plants were generated using *A. tumefaciens* strain GV3101 and a "Simplified *Arabidopsis* Transformation Protocol" from Clough and Bent (Clough and Bent 1998) with minor modifications. The plants were grown under short day conditions for six to eight weeks, before they were transferred to long day greenhouses. After two additional weeks, the tip of the emerging bolt was cut off to induce the growth of secondary inflorescences. After 10 days the plants were transformed by dipping the inflorescences into infiltration medium with agrobacteria (harboring the construct of interest). After seed germination, transgenic plants were screened by spraying three times with 0.15% Basta (Bayer CropScience, Monheim). Basta is an herbicide, which contains 200 g/L glufosinate-ammonium as an active ingredient.

2.22 *Agrobacterium* infiltration of Solanaceae tissues

Floral tissues and leaves were infiltrated with *A. tumefaciens* strain GV3101 harboring the p27IGUS vector. The cells were cultured in 10 ml YEB medium with the appropriate antibiotics overnight, pelleted and resuspended in MS medium to a concentration of OD600 = 1.0. The suspension was supplemented with 100 µM acetosyringone and cultured for 4 h. After adding Silwet L-77 (0,001%), 10 ml cultures were filled into Falcon tubes. Flowers, lanterns and leaves were injured with tips and scalpels, transferred into the *A. tumefaciens* cultures and agro-infiltrated using a vacuum chamber. Then the tissues were cultured on MS medium plates at 25 °C under artificial light (16 h photoperiod) for 48 h before examining reporter gene activity.

2.23 Reporter gene assay

The tissues (*Arabidopsis*: rosette and cauline leaves, stem, inflorescences and siliques, Solanaceae: buds, flowers, lanterns, leaves) were histochemically assayed for GUS activity. The samples were incubated for 40 h at 37 °C in GUS histochemical buffer containing X-Gluc. Afterwards the buffer was exchanged by 70% (v/v) ethanol to remove the chlorophyll and preserve the samples. The GUS-positive plant tissues were examined with a binocular and photographed with a digital camera (Leica). GUS-

stained tissues and plants shown in this thesis represent the typical results of at least three independent lines for each construct. Plants carrying a *pSVP*::*GUS* construct (kindly provided by Chaoying He), which show amongst other tissues strong GUS expression in leaves, were used as a positive control.

3. Results

PartI: Structural characterization of *MPF2*-like genes

The major goal of this thesis was to elucidate the role of *cis*-regulatory elements in transcriptional regulation of *MPF2*-like MADS-box transcription factors in Physaleae. However, when this project was initiated, only short *MPF2*-like upstream sequences and loci from the quite distantly related solanaceous species *S. tuberosum* and *P. floridana* were available for structural and functional comparison (He and Saedler 2005). Therefore *MPF2*-like genes from Physaleae species and their upstream regions and first introns were isolated to enable identification of important regulatory regions and further *cis*-elements in these.

3.1 Allelic variation of *MPF2*-like genes of several Physaleae

In addition to *MPF2*-like upstream and intronic sequences (see Table 3.7.1) the cDNAs of *MPF2*-like genes from the studied species were collected (see Table 3.1.1). This was done (i) to ensure that the putative regulatory sequences isolated for promoter analysis belong to a transcribed *MPF2*-like gene, (ii) to predict the putative start of transcription and (iii) to possibly gain information about how many and which types of *MPF2*-like genes might be present in these species.

Published *MPF2*-like cDNA sequences like *MPF2, MPP3, MSM2* and *STMADS16* were downloaded from NCBI, while most others needed to be isolated by 5'/3'RACE and RT-PCR. 3'RACE was performed to isolate *MPF2*-like cDNAs from *D. fasciculata, W. coccoloboides, W. solanacea, M. solanaceus* and *P. alkekengi*. At the same time, Jinyong Hu independently isolated *MPF2*-like cDNA sequences from several Physaleae (Hu and Saedler 2007). Prior to publication he provided several *MPF2*-like cDNAs: the partial coding sequences of *C. baccatum, I. australe, P. acutifolia, P. crassifolia, P. fuscomaculata, P. lanceifolia, P. minima* and *P. philadelphica* and the complete coding sequence of *V. breviflora.* 5'RACE was performed to obtain the 5'UTR and the MADS-box of the *MPF2*-like genes from the above-mentioned species - except for *I. australe* and *P. lanceifolia*, as their 5'UTRs were provided by Britta Grosardt.

In total, 18 *MPF2*-like cDNAs from 12 Physaleae and one *Capsicum* species were isolated or extended (see Table 3.1.1). Each sequence was found in at least two independent clones. In case of uncertainties at a specific nucleotide position ambiguity codes were used. All sequences have been submitted to the EMBL database.

Table 3.1.1: Isolated or extended *MPF2*-like cDNAs

Species	Gene/Allele	Accession number	Full-length (bp)	5'UTR (bp)	CDS (bp)	3'UTR (bp)	Protein (aa)
Physalis crassifolia	*PcrM222_1**	FN356432	(827)	113	(714)	-	(>238)
Physalis fuscomaculata	*PfuM230_1**	FN356433	(848)	134	(714)	-	(>238)
Physalis lanceifolia	*PlaM235_1**	FN356434	(899)	185	(714)	-	(>238)
Physalis minima	*PmiM239_1**	FN356435	(901)	187	(714)	-	(>238)
	*PmiM239_2**	FN356436	(936)	222	(714)	-	(>238)
Physalis philadelphica	*PphM255_1**	FN356437	(873)	137	(736)	-	(>245)
Margaranthus solanaceus	*MsoM201_1 **	FN356430	(1053)	170	750	(133)	249
Witheringia solanacea	*WIsoM202_1*	FN356442	1275	189	750	336	249
Physalis alkekengi	*PalM211_1*	FN356431	1295	247	750	298	249
Dunalia fasciculata	*DfaM201_1*	FN356426	1223	169	747	307	248
	DfaM201_2	FN356427	1299	244	747	308	248
Iochroma australe	*IauM201_1*	FN356428	(912)	137	747	(28)	248
	*IauM201_2**	FN356429	(912)	137	747	(28)	248
Vassobia breviflora	*VbrM201_1*	FN356438	(986)	162	747	(77)	248
	VbrM201_2	FN356439	(907)	160	747	-	248
Witheringia coccoloboides	*WIcoM201_1*	FN356440	1144	202	747	195	248
	WIcoM201_2	FN356441	1144	202	747	195	248
Capsicum baccatum	*CbaM202_1 **	FN356425	1297	244	747	306	248

* = same allele as previously published by Hu and Saedler (2007); () = incomplete

The following nomenclature was used for the isolated *MPF2*-like genes: (the first - or in the case of *Witheringia* the first two letters - of the genus in capital letters) + (the first two letters of the species name in small letters) + (M for MADS-box gene) + (2 for *MPF2*-like) + (the last two figures of the cultivar number) + (_) + (number of the allele).

Interestingly, for several species more than one *MPF2*-like cDNA sequence was found. All these isolated sequences contained a highly conserved MADS-box, a short I-region, a K-box and a C-terminal region. Although the latter was not completely isolated from the American *Physalis* species, it seemed that among the isolated cDNAs no pseudogenes were present, since no premature stop codons were found in the first 714 bp of their coding region. In those cases, where the full-length coding region was isolated, the *MPF2*-like genes encoded predicted polypeptides of 248 amino acids (Iochrominae, *W. coccoloboides* and *C. baccatum*) or 249 amino acids (Physalinae, not including *W. coccoloboides*).

In the following paragraphs the results from intraspecific comparisons of the isolated *MPF2*-like cDNA sequences from the studied species are given – also with respect to the dataset of *MPF2*-like sequences published by Hu and Saedler in 2007. This is done to give an overview about the intraspecies variation and to summarize the gain of sequence data due to this study.

The previously published partial coding sequences of *MPF2*-like genes from several American *Physalis* species, now renamed *PcrM222_1* (*P. crassifolia*), *PfuM230_1* (*P. fuscomaculata*), *PlaM235_1* (*P. lanceifolia*), *PmiM239_1* (*P. minima*), *PmiM239_2* (*P. minima*) and *PphM255_1* (*P. philadelphica*) were extended in this study by adding the 5'UTR and the complete MADS-box. However, the C-terminal end of the coding sequences is still incomplete. Based on full-length Physalinae *MPF2*-like cDNAs a coding region of 750 bp is assumed, hence approximately 36 bp are lacking. In contrast to the other *Physalis* species, in *P. minima* two different kinds of *MPF2*-like sequences coexist, which is compatible with the tetraploid nature of this species. The two sequences, *PmiM239_1* and *PmiM239_2,* showed a homology of 97% in the first 714 bp of the coding region and the first 238 amino acids of the predicted polypeptides were 96% identical.

Hu and Saedler (2007) published the partial coding sequence of an *MPF2*-like cDNA from *M. solanaceus*. Now the 5'UTR, the full coding sequence and part of the 3'UTR

of this gene, termed *MsoM201_1*, have been isolated. The full sequence exhibits a length of 1053 bp.

A 1275 bp long full-length cDNA was isolated from *W. solanacea* (also belonging to *Physalinae*) and termed *WIsoM202_1*. The nucleotide sequence of this allele (as well as the translated protein sequence) and the previously isolated cDNA sequence from this species shared 98% sequence identity in the coding region.

I isolated a 1295 bp long full-length *MPF2*-like cDNA sequence from *P. alkekengi*, termed *PalM211_1*, while the previously isolated cDNA (Hu and Saedler 2007) comprised only part of the coding sequence and the 3'UTR. So far only this one allele has been found for this diploid Old World *Physalis* species.

Two full-length cDNA sequences were isolated from *D. fasciculata*: *DfaM201_1* (1223 bp) and *DfaM201_2* (1299 bp). Both alleles shared 98% sequence identity in the coding region at the nucleotide level and also at the protein level. Their last 612 bp (or 203 amino acids) of the coding region were 98% and 99% identical with the previously isolated partial cDNA sequence from this species, suggesting that the three cDNAs could be different alleles.

Two cDNA sequences from *I. australe* (*IauM201_1* and *IauM201_2*) were isolated. They were both 912 bp long and contained the 5'UTR, the full coding sequence and part of the 3'UTR. As the coding regions of the nucleotide sequences are 99% identical and the protein sequences are even 100% identical, the two sequences are most likely allelic. The overlapping regions of *IauM201_2* and the published partial coding sequence of a *MPF2*-like gene from *I. australe* (Hu and Saedler 2007) were completely identical, indicating that they derive from the same allele.

Two new cDNA sequences were found in *V. australe*: *VbrM201_1* and *VbrM201_2*, which were 986 bp and 907 bp long and contained the 5'UTR and the full coding sequence. Sequence comparisons with the previously published cDNA sequence revealed that the coding regions of all three sequences shared 99% sequence identity at the nucleotide level and 98% to 99% at the protein level. Hence also these sequences might be allelic. However, in the cDNA from Hu and Saedler (2007) a duplication of 19 amino acids is present in its C-terminus – this was not included in the sequence identity calculations.

From *W. coccoloboides* two different full-length putative allelic cDNA sequences were extracted. *WIcoM201_1* and *WIcoM201_2* were both 1144 bp in size. The codogenic parts of their nucleotide sequences and their protein sequences were 99% identical.

The coding region of the previously published cDNA shared 97% sequence identity with *WIcoM201_1* and 99% with *WIcoM201_2* at the nucleotide level and the translated protein sequence 98% with both. Further, the published sequence showed a duplication of 19 amino acids in the C-domain, which was not included in the sequence identity calculations. This still comparatively high sequence conservation again hints towards the existence of several different alleles in this species.

So far only the partial coding sequence and the 3'UTR of a *MPF2*-like gene from *C. baccatum* were published. In this study the cDNA sequence was extended to gain the full coding sequence and the 3'UTR. The complete transcript was 1297 bp long and was termed *CbaM202_1*.

In summary, only in one of the American *Physalis* species (*P. minima*), but in all three Iochrominae and both *Witheringia*, more than one *MPF2*-like cDNA sequence was found so far. However, all these sequences exhibited a high degree of homology showing only a few nucleotide differences, which e.g. in the case of *V. breviflora* does not even affect the translated protein sequence. Therefore they were considered to be alleles and not treated as different genes. As several of these di- or tetraploid species are wild, not inbred plants, a certain amount of heterozygosity can be expected. Other explanations would be that the different sequences originated by recent duplications or that the mutations occurred during PCR amplification. Since high fidelity enzymes were used for amplification PCR-derived mutations are rather unlikely, but possible.

3.2 Sequence comparison of MPF2-like proteins among Physaleae

In this section the results from interspecies comparisons of *MPF2*-like cDNA sequences are described. An alignment of MPF2-like proteins from Physaleae, *Capsicum* and *Solanum* is shown in Fig. 3.2.1.

For further comparison *MPF2*-like cDNA sequences from one A-gene (*WSA206* from *W. somnifera*) and two B-genes (*WSB206* from *W. somnifera, TAB201* from *T. anomalum*) were kindly provided by Ramzan Khan, while several other Solanaceae sequences (He and Saedler 2005, Hu and Saedler 2007) were downloaded from NCBI.

Multiple sequence comparisons revealed that the predicted polypeptides of all isolated *MPF2*-like sequences exhibit the typical MIKC structure: a MADS domain, followed by an intervening (I-) domain, a K-domain and a C-terminal domain (Fig. 3.2.1, domain borders after Kim *et al.* (2002)).

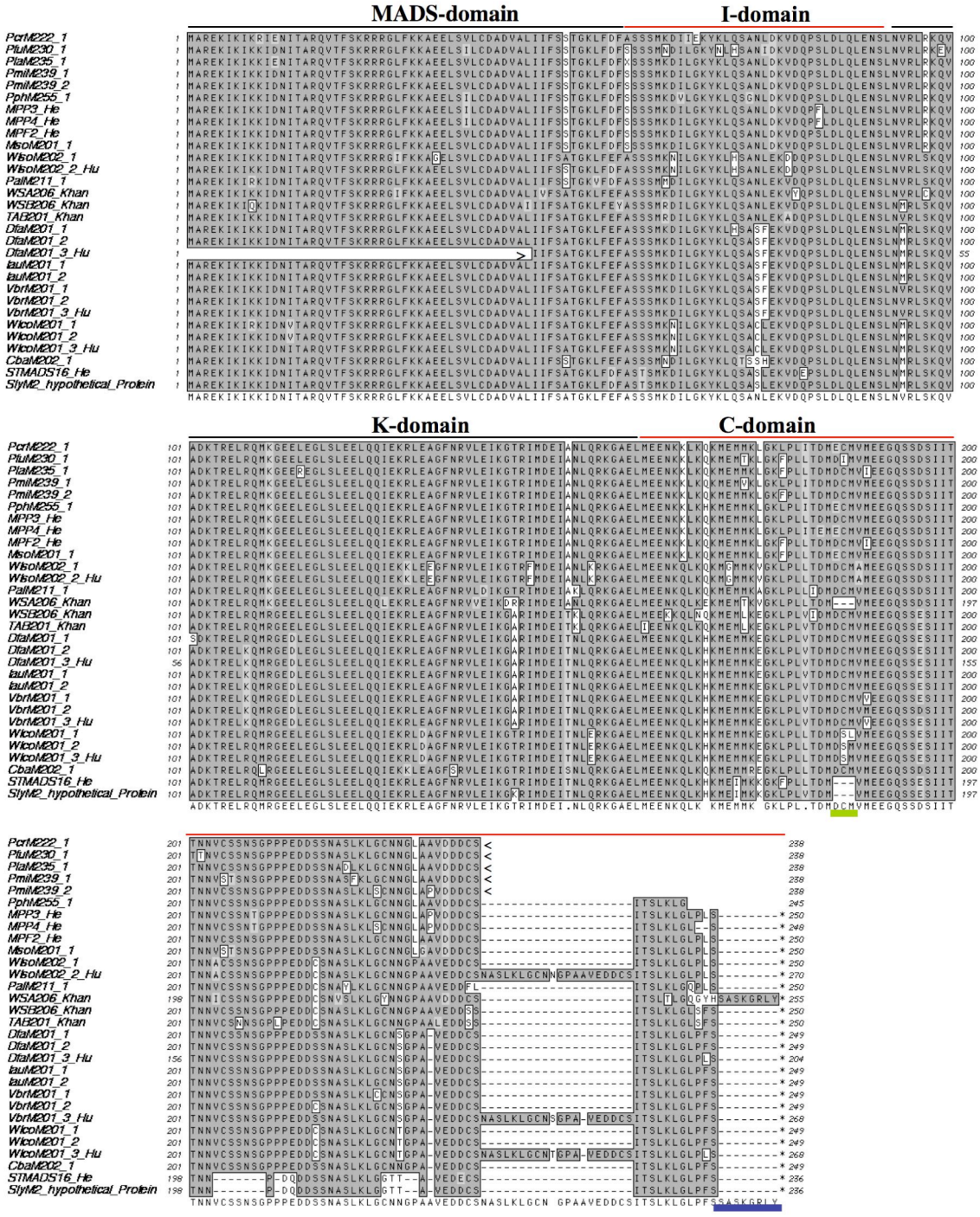


Figure 3.2.1: Alignment and sequence comparison of MPF2-like MADS-domain proteins from Physaleae, *Capsicum* and *Solanum*

> indicates that the sequence is incomplete at its N-terminal end, while < shows that the sequence is incomplete at its C-terminal end. Identities are depicted in a dark shade and similarities in lighter grey. Gaps introduced to optimize the alignment are indicated by dashes. The MADS, I, K and C domains (borders based on Kim *et al.* 2002) are marked by red and black lines. The 3 amino acid deletion and the 8 amino acid extension in WSA206 are indicated by green and blue lines. Accession numbers of sequences from this study can be found in Table 3.1.1. Sequences from other sources are indicated by "He" (He and Saedler 2005), "Hu" (Hu and Saedler 2007) and "Khan" (Khan 2009, Khan *et al.* 2009). The accession numbers can be found in supplementary Tables 8.1.1 and 8.1.2.

In the preceding section (see chapter 3.1) it was already mentioned that the coding sequences of the isolated *MPF2*-like cDNAs differ in length: those from *C. baccatum, W. coccoloboides* and the Iochrominae are 747 bp long (proteins: 248 amino acids), while the sequences from the Physalinae (not including *W. coccoloboides*) possess 1 amino acid more (Alanine or Proline) in the C-domain (see Fig. 3.2.1, alignment position 233) - like *MPF2, MPP3* and *WSB206.* The MPF2-like proteins from different subgroups e.g. from the American *Physalis* or the Iochrominae had each certain characteristics in their amino acid sequence. However, when the protein sequences from all species are taken into account, it is not possible to clearly correlate differences in MPF2-like protein sequences and the abundance of ICS.

When WSA206, the MPF2-like-A sequence of *W. somnifera*, is compared to the other Physaleae proteins, the two most obvious differences are observed in the C-domain (Khan 2009, Khan *et al.* 2009) – although there were amino acid exchanges in all four domains of the protein. WSA206 had a 3 amino acid deletion identical to the one featured by STMADS16 and MSM2. In addition to that it had an extension of 8 amino acids at its carboxy terminal end. These two structural features, which were signatures of the *Withania* MPF2-like-A proteins, were not observed in any of the other isolated Physaleae proteins.

In the next step the amount of sequence identity between MPF2, WSA206, WSB206 and complete MPF2-like protein sequences from Physaleae and *C. baccatum* were calculated (see Table 3.2.2). As the multiple sequences found in some species were considered allelic, only one sequence from each species was included. Overall it can be postulated that MPF2-like proteins in Physaleae are highly conserved: so far the sequence identity found ranged from 84% (WSA206 vs. Iochrominae proteins) up to 97% (MPF2 vs. MsoM201_1). The latter is similar to the degree of sequence identity observed in different alleles from one species, fitting to the high relatedness of *M. solanaceus* and the American *Physalis*. While all Physalinae proteins, as well as CbaM202_1, were most similar to MPF2, the Iochrominae and *W. coccoloboides* MPF2-like proteins displayed the highest sequence identity values, when they were compared to WSB206. However, these differences in degree of sequence identity were very subtle.

Table 3.2.2: Pairwise identities of *MPF2*-like cDNA and protein sequences

Species		Pairwise identities in % (nucleotide / protein)		
		MPF2	WSA206	WSB206
P. floridana	MPF2	**100/100**	90/86	92/87
M. solanaceus	**MsoM201_1**	**98/97**	90/85	92/86
W. solanacea	**WIsoM202_1**	**93/90**	91/86	92/87
P. alkekengi	**PalM211_1**	**95/91**	90/85	93/90
D. fasciculata	**DfaM201_1**	92/89	90/84	**94/91**
I. australe	**IauM201_1**	92/89	90/84	**93/92**
V. breviflora	**VbrM201_1**	91/89	89/84	**93/91**
W. somnifera	WSA206	90/86	**100/100**	90/84
W. somnifera	WSB206	92/87	90/84	**100/100**
W. coccoloboides	**WIcoM201_1**	**92**/88	89/84	**92/90**
C. baccatum	**CbaM202_1**	**92/90**	89/83	**92**/89
S. tuberosum	STMADS16	**86/84**	84/79	**86**/83

Accession numbers of sequences from this study (given in bold) can be found in Table 3.1.1. Accession numbers of sequences from other sources (He and Saedler 2005, Hu and Saedler 2007, Khan 2009, Khan *et al.* 2009) can be found in supplementary Tables 8.1.1 and 8.1.2.

3.3 Gene tree of *MPF2*-like genes in Physaleae

To understand the evolutionary relationships between the *MPF2*-like genes from different Physaleae species, phylogenetic reconstructions were conducted in cooperation with Thomas Münster.

Two different datasets were used: (i) the *MPF2*-like cDNA sequences of Physaleae species and (ii) their translated protein sequences. The datasets contained *MPF2*-like sequences from this study and sequences from other studies (He and Saedler 2005, Hu and Saedler 2007, Khan 2009, Khan *et al.* 2009). *STMADS16* and *MSM2* were used as outgroup. As *MPF2*-like genes from Physaleae are extremely conserved, most of the coding region, except a short part of the C-terminal region, was used for the phylogenetic reconstructions to have enough informative sites. Therefore all sequences, which did not contain the complete MADS-box, I- and K-region, were excluded. The ClustalW alignments were done using the Macvector software (Settings: alignment speed = slow, open gap penalty = 5.0, extend gap penalty = 1.0, delay

divergent = 40%, transitions = weighted) and manually adjusted afterwards. Further, the duplications in the C-termini of *WIsoM202_2, VbrM201_3* and *WicoM201_3* were excluded from the analyses. Both alignments can be found in the supplement section (see Fig. 8.2.1 and 8.2.2).

Fig. 3.3.1 shows a Maximum Parsimony tree based on positions 1 to 714 of a cDNA alignment of the coding regions of *MPF2*-like MADS-box genes. In this nucleotide alignment (see Supplement, Fig. 8.2.1) the third codon positions have not been masked to prevent loosing putative informative sites. The tree topology suggests that the *MPF2*-like genes from Physaleae can be subdivided into five groups: *MPF2*-like Physalinae genes, *MPF2*-like-A genes, *MPF2*-like-B genes, *MPF2*-like *W. coccoloboides* genes and *MPF2*-like Iochrominae genes. The *MPF2*-like Iochrominae genes were a highly supported subclade, which seems to be located basal to the other Physaleae genes. Further, the *MPF2*-like *W. coccoloboides* genes were clearly separated from the *MPF2*-like Physalinae, the *MPF2*-like-A and the *MPF2*-like-B genes. Further, high bootstrap values supported a branching that places the *MPF2*-B-genes in a basal position to the sister groups *MPF2*-like Physalinae genes and *MPF2*-like-A genes. The fact that *WSA206* was placed closer to the *MPF2*-like Physalinae genes than *WSB206* stands in contrast to the calculated sequence identities for the whole coding regions, which suggests that *WSB206* has a slightly higher degree of homology to the *MPF2*-like Physalinae genes. Within the Physalinae group, genes from the American *Physalis* species clustered together, while *PalM211_1* from the also geographically distant Old World *Physalis* species *P. alkekengi* clustered together with the *W. solanacea* sequences. With the exception of the positioning of the *MPF2*-like *W. coccoloboides* genes, the phylogeny of the Physaleae *MPF2*-like genes matched the Physaleae phylogeny proposed by Hu and Saedler (2007).

Further, the different putative allelic cDNAs from *V. breviflora, W. coccoloboides* and *W. solanacea* clustered together in each case, supporting the assumption that they are alleles or very recent duplications.

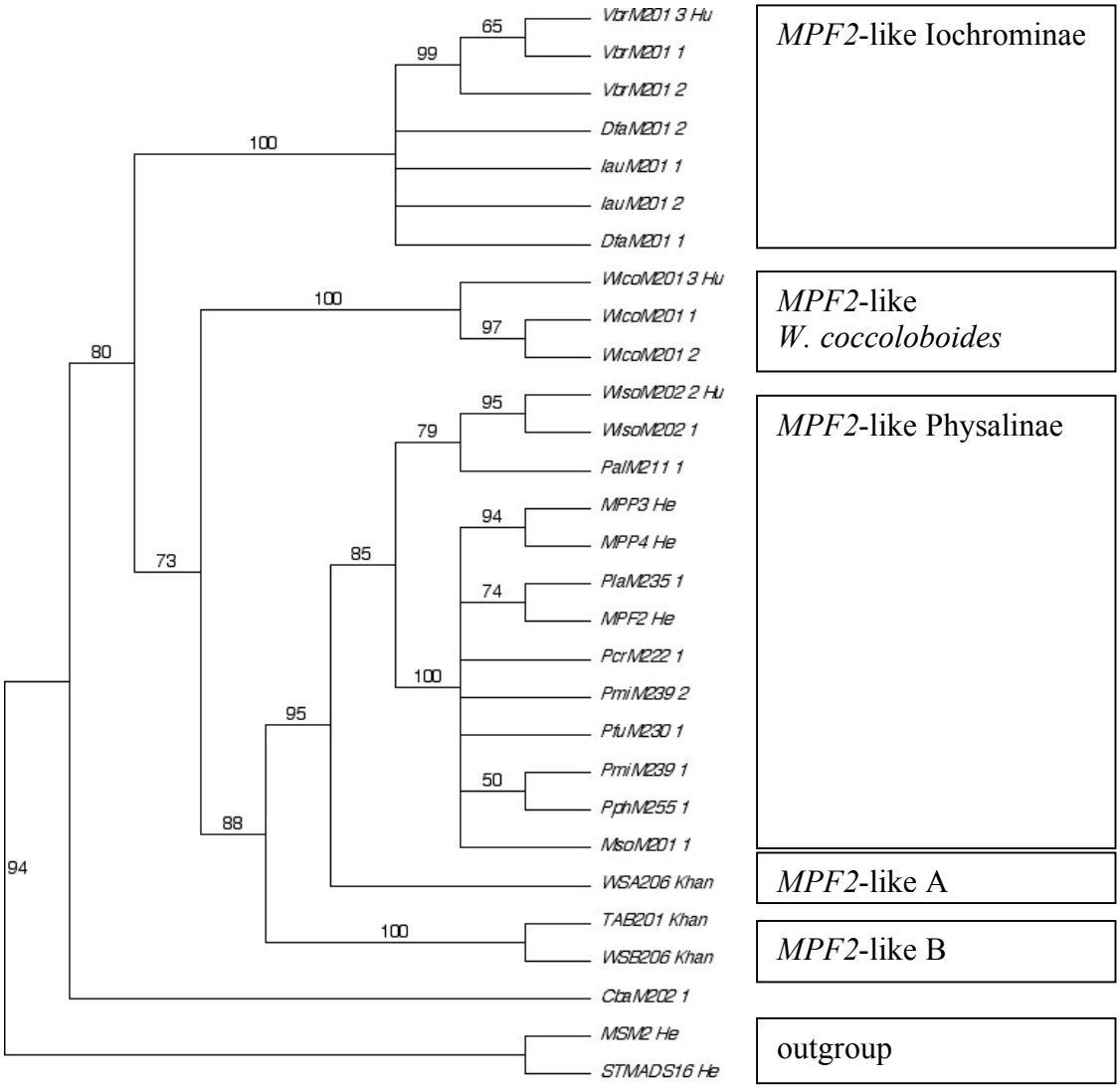


Figure 3.3.1: Phylogenetic tree of *MPF2*-like cDNA sequences from Physaleae and *Capsicum*

Maximum Parsimony tree with *STMADS16* (*S. tuberosum*) and *MSM2* (*S. macrocarpon*) as outgroup. Bootstrap values are shown. Accession numbers of sequences from this study can be found in Table 3.1.1. Sequences from other sources are indicated by "He" (He and Saedler 2005), "Hu" (Hu and Saedler 2007), "Khan" (Khan 2009, Khan *et al.* 2009). The accession numbers can be found in the supplementary section (see Tables 8.1.1 and 8.1.2).

Fig. 3.3.2 shows a Maximum Parsimony tree, which was based on positions 1 to 238 of the translated protein alignment depicted in supplementary Fig. 8.2.2. Here, the resolution was decreased in comparison to the tree based on *MPF2*-like nucleotide sequences (see Fig. 3.3.1). The Physaleae MPF2-like proteins fell into four subgroups: MPF2-like Physalinae and MPF2-like-A proteins, MPF2-like-B proteins, MPF2-like *W. coccoloboides* proteins and MPF2-like Iochrominae proteins. Three of these were supported by high bootstrap values: all proteins from the Iochrominae clustered

together, as did the MPF2-like-B proteins and also the *W. coccoloboides* proteins. Further, a weakly supported branch contained CbaM202_1 from *Capsicum*, the MPF2-like-A protein WSA206 and the Physalinae proteins. Within the latter the MPF2-like proteins from the American *Physalis* species clustered together, as expected.

Again the tree topology supported the assumption that the different sequences from *W. solanacea, V. breviflora* and *W. coccoloboides* could be alleles or very recent duplications.

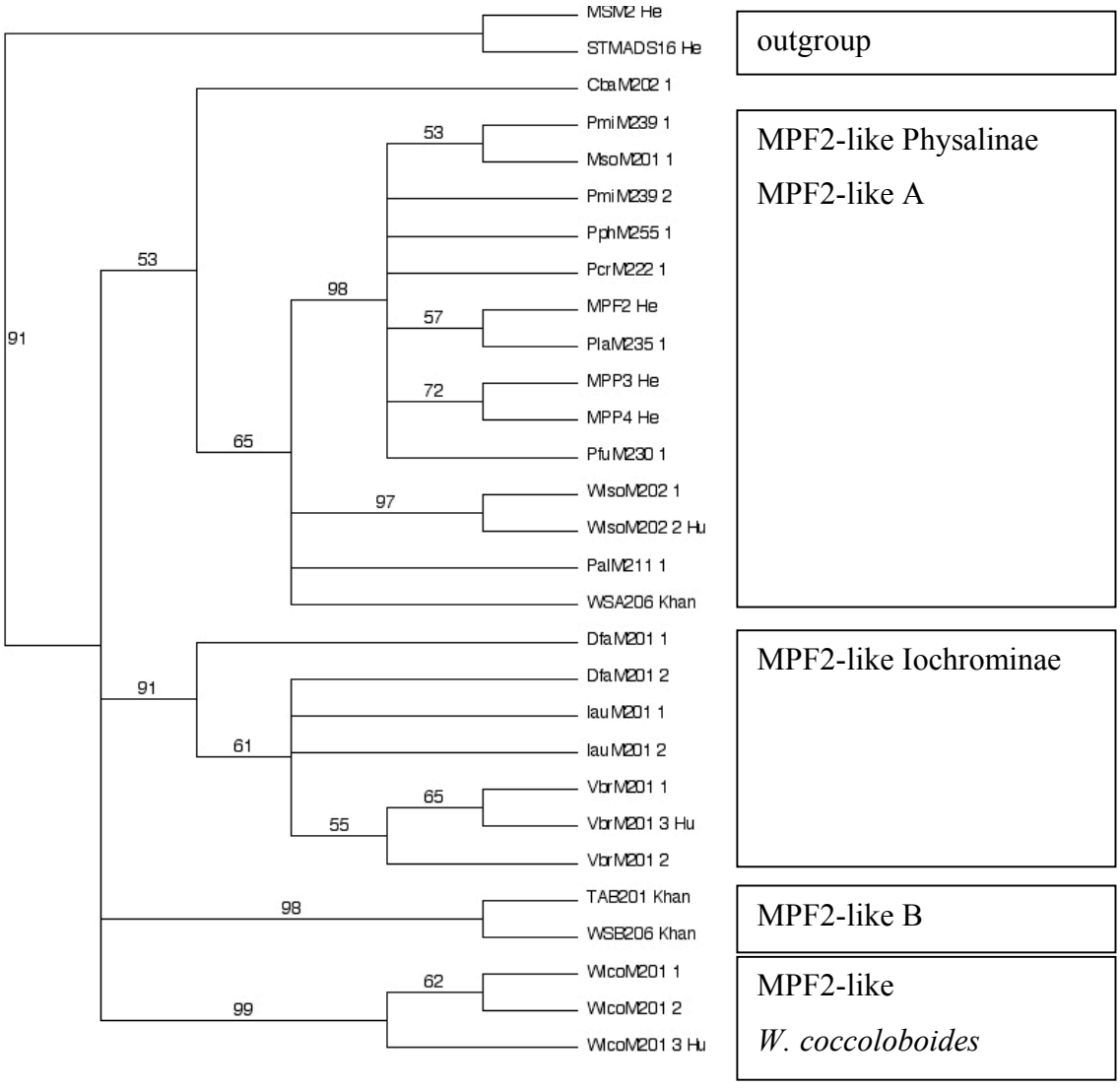


Figure 3.3.2: Phylogenetic tree of *MPF2*-like protein sequences from Physaleae and *Capsicum*

Maximum Parsimony tree with STMADS16 from *S. tuberosum* and MSM2 from *S. macrocarpon* as outgroup. Bootstrap values are shown. Accession numbers of sequences from this study can be found in Table 3.1.1. Sequences from other sources are indicated by "He" (He and Saedler 2005), "Hu" (Hu and Saedler 2007) and "Khan" (Khan 2009, Khan *et al.* 2009). The accession numbers can be found in supplementary Tables 8.1.1 and 8.1.2.

3.4 Searching for *MPF2*-like-A genes in Iochrominae

As already mentioned in the introduction, two structurally and functionally different types of *MPF2*-like genes were found in *Withania* by Ramzan Khan, which were then termed *MPF2*-like-A and *MPF2*-like-B genes. However, in species from the closely related genus *Tubocapsicum* he could detect only *MPF2*-like-B genes. In the Physaleae species studied here also only one type of *MPF2*-like cDNAs could be isolated, even though always several *MPF2*-like cDNA clones were sequenced. To verify the assumption that, while in *Withania* two types of *MPF2*-like genes are present, the A-type is missing in several other Physaleae genera, Southern blotting experiments were performed (see Fig. 3.4.1).

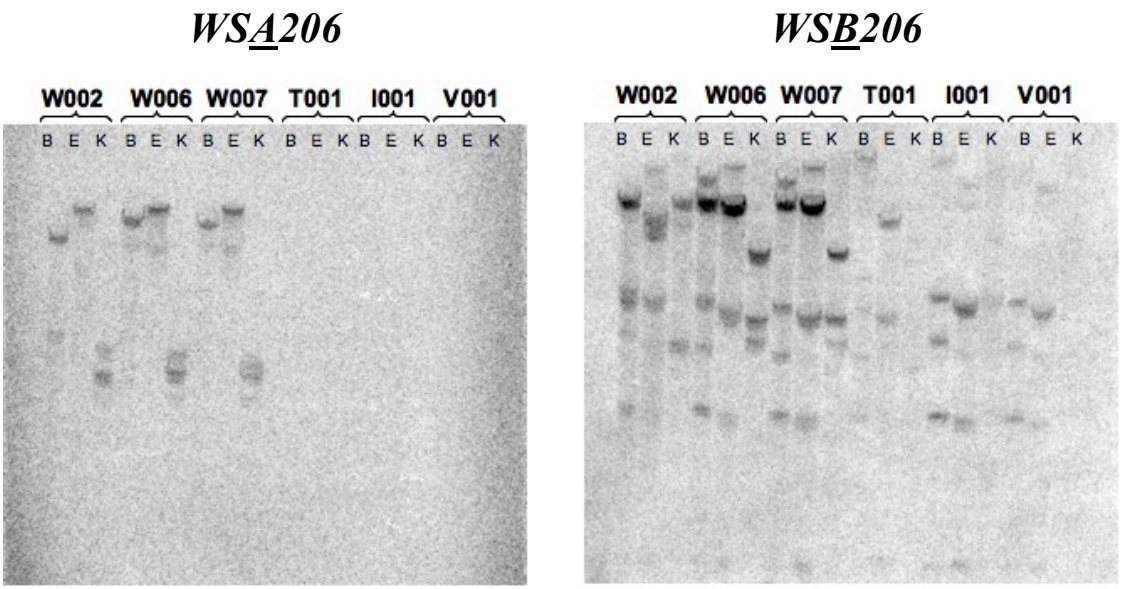


Figure 3.4.1: Genomic Southern hybridization for detection of *MPF2*-like-A and *MPF2*-like-B genes

10 µg genomic DNA of *W. coagulans* (W002), *W. somnifera* (W006, W007), *T. anomalum* (T001), *I. australe* (I001) and *V. breviflora* (V001) were digested with BamHI (B), EcoRI (E) and KpnI (K), separated on a 0,7% agarose gel and blotted to a nylon membrane. The hybridization with gene-specific probes, which contained the C-terminal part of *WSA206* and *WSB206*, and the washing were done at 68 °C with high stringency.

When a gene-specific probe for the *W. somnifera MPF2*-like-A gene *WSA206* was used, hybridization signals could be detected for all three *Withania* accessions tested, while no signal was obtained for *T. anomalum, I. australe* and *V. breviflora*. These results indicate that this gene is absent in the genomes of *Tubocapsicum* and both

Iochrominae species. By contrast, hybridization with a probe specific for the *MPF2*-like-B gene *WSB206* led to multiple signals in all examined species. The high number of signals in *Withania* can be explained by the polyploid genomes of these species. However, the signals observed in *T. anomalum, I. australe* and *V. breviflora* were obviously weaker than in *Withania*, which was very likely due to the heterologous probe.

3.5 Database search of *MPF2*-like genes in tomato

Information about the genomic location of an *MPF2*-like gene in tomato (*S. lycopersicum)* and the location of neighboring genes was thought to be helpful for this project, since knowledge about the upstream flanking gene might have facilitated upstream sequence isolation by PCR, instead of genome walking. Although the size of the intergenic region does not necessarily define the size of the promoter, this information should still have been interesting and could have been used at least for initial decisions about the amount of upstream sequence to isolate for promoter analysis in Physaleae.

To identify the *MPF2*-like locus from *S. lycopersicum,* megablast searches (settings: database = nucleotide collection, organism = *Solanum lycopersicum*) were performed regularly at the NCBI Blast site using the published full-length *STMADS16* mRNA (Accession no.: AY643733.1) as query. These searches yielded a putative *MPF2*-like locus in the clone LE_HBa-272I8 (Accession no.: CU468638.8), which was located on chromosome 4.

In order to annotate this locus, I tried to detect putative *MPF2*-like cDNAs from the tomato via megablast search (settings: database = expressed sequence tags (ests), organism = *Solanum lycopersicum*) using the published full-length *STMADS16* mRNA (AY643733.1) as query. Subsequently the tomato ESTs found were blasted again against EST entries from *S. tuberosum*. Hereby five partial putative *MPF2*-like mRNAs (DB685265.1, AI482813.1, AW929235.1, AI482810.1 and AI773429.1) and one putative partial *MPF1*-like mRNA (DB696642.1) could be identified. A hypothetical full-length *MPF2*-like mRNA sequence from *S. lycopersicum* was inferred from alignments of the *MPF2*-like tomato ESTs, the genomic tomato *MPF2*-like locus and the *STMADS16* cDNA. The hypothetical gene was termed *SlyM2* (for sequence see Supplement, Table 8.1.5). Further, the hypothetical exon-intron structure

(see Fig. 3.5.1) was determined in the program geneseqer. Apparently, the exon-intron structure is very similar to *MPF2* and *STMADS16* and also here the first intron in the coding region is the largest intron.

The genomic region around the *SlyM2* locus (positions 63991 to 89720 = 10kb upstream until 10kb downstream of the coding region) was blasted against the *S. lycopersicum*, *Solanum* and plant EST databases to find the neighboring genes. All searches yielded similar results. At position 80919 (approximately 1119 bp downstream of the predicted stop codon of the hypothetical *SlyM2* gene) a gene belonging to the DUF246 family seemed to be located (start of its ORF). This gene had the same orientation as *SlyM2*. The downstream gene was followed by a putative MscS (Small-conductance mechanosensitive channel) family gene. Several ESTs were detected at 3.5 kb upstream of the locus, but yielded higher identities to published genomic fragments from other tomato chromosomes. However, at position 68118 (around 5873 bp downstream of the coding region of *SlyM2*) an oppositely oriented DnaJ/Hsp40 (heat shock protein 40) gene might start (start of its ORF). In summary, the intergenic region upstream of the *SlyM2* locus appears to be much larger (nearly 6 kb) than the downstream intergenic region (approximately 1 kb).

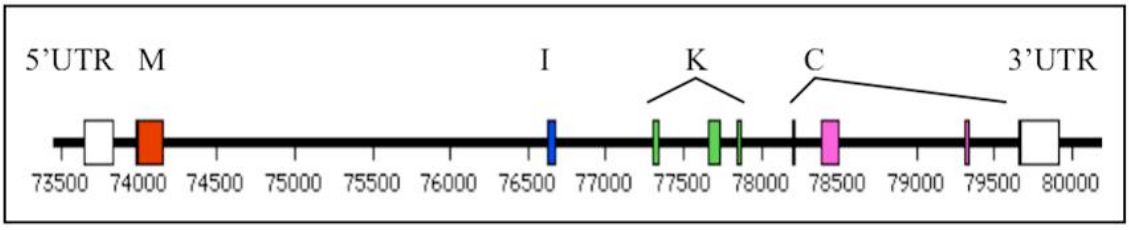


Figure 3.5.1: Hypothetical *S. lycopersicum MPF2*-like gene *SlyM2*: Position of the locus in clone LE_HBa-272I8 (chromosome IV) and predicted exon-intron structure

Exons are depicted as boxes (white boxes: 5' and 3'UTR, red: MADS-box, blue: I-region, green: K-region, pink: C-terminal region). Coding region: 73991-79720; 5'UTR: 73646-73990; 3'UTR: 79721-79931; Exon 1: 73646-73857 (212 n); Exon 2: 73990-74172 (183 n); Exon 3: 76622-76700 (79 n); Exon 4: 77301-77362 (62 n); Exon 5: 77662-77761 (100 n); Exon 6: 77847-77888 (42 n); Exon 7: 78200-78241 (42 n); Exon 8: 78391-78524 (134 n); Exon 9: 79306-79356 (51 n); Exon 10: 79648-79931 (284 n)

Unfortunately this tomato clone was published when the isolation of *MPF2*-like upstream sequences from the different Physaleae had already been completed. Hence

information about the upstream gene could not be used for isolation of upstream sequences in the plant species of this study. Further, it was decided not to include the upstream sequence of *SlyM2* in the *in silico* promoter analysis (see chapter 3.7), since only a hypothetical *SlyM2* mRNA was available. However, the *SlyM2* locus was included in the bioinformatics analysis of whole *MPF2*-like genomic loci due to the lack of other *MPF2*-like loci from species, which show no *MPF2*-like expression in sepals - like *S. lycopersicum* (verified by Northern Blot experiments - Chaoying He, unpublished data).

3.6 Mulan analysis of *MPF2*-like loci revealed conserved regions in introns

Cis-regulatory elements may be located in the 5' upstream region, in exons (Zheng et al. 2001), in introns and in the 3' UTR. *Cis*-elements have been found in introns of many other MADS-box genes (Davies *et al.* 1999, Hong *et al.* 2003, Liu *et al.* 2007). As this was also the case for *AGL24* (Liu *et al.* 2007), the putative orthologue of *MPF2* in *A. thaliana*, it seemed to be likely that introns of *MPF2*-like genes might also contain *cis*-elements. In an attempt to narrow down the number of putative important intronic regulatory regions, the available sequences of *MPF2*-like loci from Solanaceae were analyzed with the program Mulan (Multiple sequence Local Alignment and conservation visualization tool) to detect conserved regions in introns. At that time the sequences of four *MPF2*-like loci were available: two from species showing expression in sepals (*MPF2* and *WSA206*) and two from species lacking expression in sepals (*STMADS16* and *SlyM2*). Overall conserved regions might contain *cis*-elements involved in vegetative expression, while other regions, which are only conserved between species with a similar *MPF2*-like gene expression pattern, could be involved in regulating floral expression or expression in sepals.

The results are displayed in Figure 3.6.1. This graphical output shows regions, which are conserved between a base sequences (*MPF2* or *STMADS16*) and another sequence based on the "stacked-pairwise approach" (Ovcharenko *et al.* 2005). The first intron, which is also the largest one, contained two overall conserved blocks. Further overall conserved blocks in introns 3 and 5 were only identified with *STMADS16* as base

sequence. Conserved regions correlating with *MPF2*-like gene expression in sepals were also detected: The two genes from *Solanum* species exhibited conserved regions in their introns 1, 7 and 8 and in the 3'UTR, while the Physaleae species shared conserved blocks in introns 1, 5 and 8. As the first intron seemed to contain many conserved and divergent regions, it was decided to include this intronic region in the further analyses.

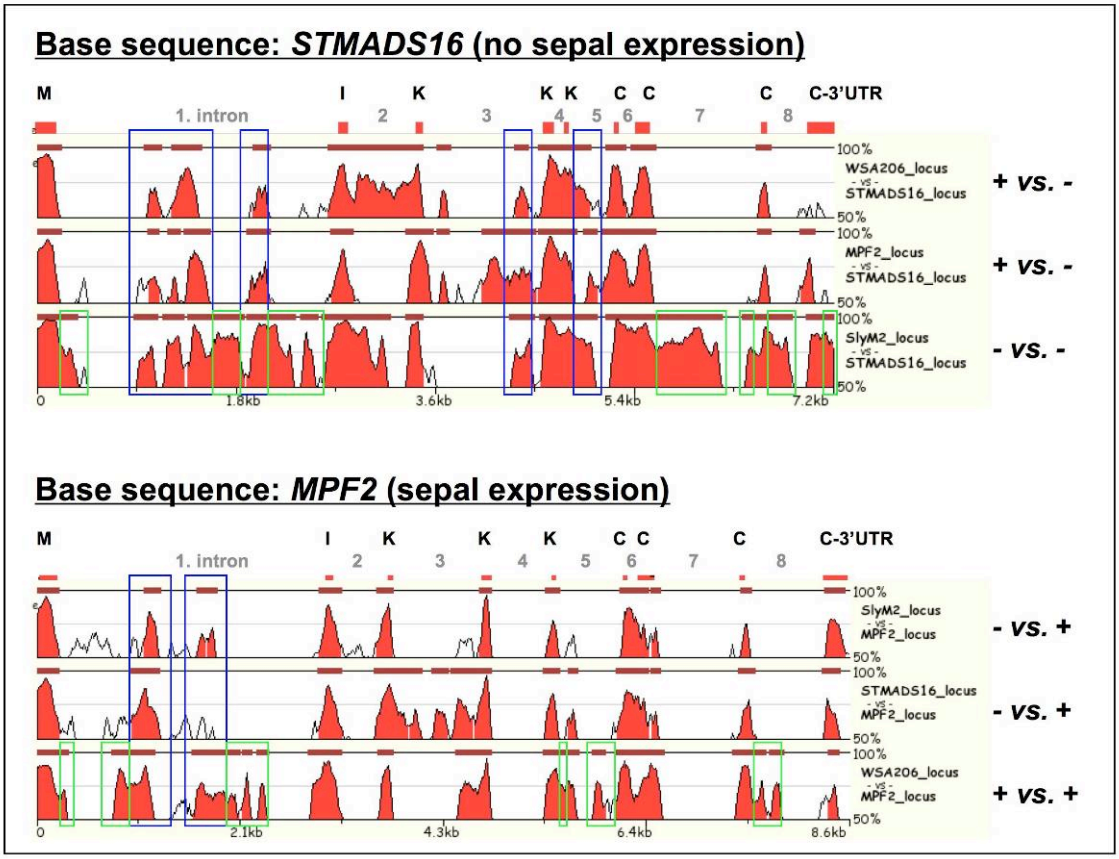


Fig 3.6.1: Conserved and divergent regions in *MPF2*-like loci

The conservation profiles between the genomic loci of *STMADS16, SlyM2, MPF2* and *WSA206* are graphically visualized. Apart from the setting "smooth graph", the default settings (ECR (Evolutionarily Conserved Regions) length = at least 100 bases, ECR similarity = at least 70%, Bottom cut-off = 50%, ECR present in = 1 species) were used. The lowest line indicates 50% and the upper one 100% conservation. The exon-intron structure of *MPF2* and *STMADS16* is shown above the diagrams: exons are indicated as light red boxes with the domain abbreviation (M, I, K or C) above, while the intronic regions are numbered from 1 to 8. Blue boxes indicate overall conserved regions, while green ones show regions, which are conserved between species with a similar *MPF2*-like gene expression pattern in sepals. For accession numbers see Supplement, Table 8.1.3.

3.7 Allelic variation in upstream sequences of *MPF2*-like genes

For the promoter analysis the upstream sequences of several Physaleae and also of *C. baccatum* were isolated using genome-walking. Further, the published short upstream sequences of *MPF2* and *STMADS16* were extended using the same technique. The upstream sequences of *P. floridana* and *P. lanceifolia* were provided by Britta Grosardt. All isolated upstream sequences are anchored in the highly conserved MADS-box. The first nucleotide of the MADS-box (ATG) was defined as position +1. Further, the first intron in the coding region, located between the MADS-box and the I-region, was isolated from *P. lanceifolia, P. philadelphica, P. alkekengi, W. solanacea, V. breviflora, I. australe* and *D. fasciculata.* The already published first introns from *P. floridana* and *S. tuberosum* were also available for further analyses (He and Saedler 2005). In total 24 *MPF2*-like upstream sequences from 17 species were collected: 23 were isolated or extended and the upstream sequence from *S. lycopersicum* was detected in the published genomic clone LE_HBa-272I8. However, for the subsequent bioinformatic analyses several sequences were excluded from the dataset: First, sequences from species with unclear gene expression pattern in sepals (*P. acutifolia,* see supplementary Table 8.1.4) and second, the upstream sequences *pPphM255_2* (*P. philadelphica,* see Table 3.7.1) and *pSlyM2*, as so far no matching or only hypothetical cDNAs were available for these sequences. Hence, the collected sequence dataset was reduced to 20 *MPF2*-like upstream sequences (0.9 to 3.6 kb) from 15 Solanaceae and 10 first intron sequences from nine different species, including three, which do not express *MPF2*-like genes in their calyces: *S. tuberosum, D. fasciculata* and *I. australe* (Hu and Saedler 2007). The sequences are listed in Table 3.7.1 (for nomenclature see chapter 3.1).

Two different *MPF2*-like upstream sequences were found in the American *Physalis* species *P. minima* and *P. philadelphica*, in all three Iochrominae species and in *W. coccoloboides*. These sequences usually differed mainly by several indels, which ranged from one to several hundred bp in size, while the rest of the sequence was very conserved. In the case of *I. australe* for example the following homology was found after removal of all gaps: *IauM201_1 vs. IauM201_2* = 2047 identities in 2108 bp aligned length (97%). Further, I tried to amplify both *I. australe* sequences from genomic DNA samples deriving from two different individual plants with different sets of primers: One of the plants contained both sequences, while from the other one only

one sequence (*pIauM201_2*) could be amplified (data not shown), suggesting that one plant might be heterozygous and one homozygous.

For now, as the multiple *MPF2*-like cDNAs found in these six species were considered to be alleles of the same gene, their matching upstream sequences were also treated as alleles for the reasons mentioned in chapter 3.1.

Table 3.7.1: *MPF2*-like 5'upstream and intronic sequences from Solanaceae

No.	Species	Promoter of gene/allele	Accession no.	Isolated upstream (bp)	Complete 1. intron (bp)	cDNA
1	*P. crassifolia*	*pPcrM222_1*	FN356451	1718	-	+
2	*P. fuscomaculata*	*pPfuM230_1*	FN356452	1614	-	+
3	*P. lanceifolia*	*pPlaM235_1*	FN356453	2357	2372	+
4	*P. minima*	*pPmiM239_1*	FN356454	1728	-	+
5		*pPmiM239_2*	FN356455	1702	-	+
6	*P. philadelphica*	*pPphM255_1*	FN356457	2734	2360	+
-		*pPphM255_2*	FN356458	2273	1841	-
7	*P. floridana*	*pMPF2*	FN356460	2465	(2860) *	(+)*
8	*M. solanaceus*	*pMsoM201_1*	FN356447	2003	-	+
9	*W. solanacea*	*pWIsoM202_1*	FN356465	3062	3231	+
10	*P. alkekengi*	*pPalM211_1*	FN356450	2603	3854	+
11	*D. fasciculata*	*pDfaM201_1*	FN356443	3675	2829	+
12		*pDfaM201_2*	FN356444	2880	-	+
13	*I. australe*	*pIauM201_1*	FN356445	2187	2906	+
14		*pIauM201_2*	FN356446	2833	-	+
15	*V. breviflora*	*pVbrM201_1*	FN356461	2100	2463	+
16		*pVbrM201_2*	FN356462	1996	2459	+
17	*W. coccoloboides*	*pWIcoM201_1*	FN356463	1290	-	+
18		*pWIcoM201_2*	FN356464	1305	-	+
19	*C. baccatum*	*pCbaM202_1*	FN296152	912	-	+
20	*S. tuberosum*	*pSTMADS16*	FN356456	2587	(2536) *	(+)*

Published cDNAs and intronic sequences from He and Saedler (2005) are shown in brackets and are marked with a "*", for accession numbers see supplementary Tables 8.1.2 and 8.1.3. Species, which do not express *MPF2*-like genes in their calyx (He and Saedler 2005, Hu and Saedler 2007), are marked in red. The amount of upstream sequence isolated is given in bp. A "+" in the cDNA column shows that a matching cDNA is available.

3.8 Phylogeny based on *MPF2*-like upstream sequences

Evolutionary relationships among genes are usually inferred from their protein sequences. However, in the case of *MPF2*-like genes from Physaleae the protein sequences are highly conserved, which prevents a deeper resolution in the American *Physalis* clade. Therefore, a phylogenetic tree based on non-coding upstream sequences was calculated in cooperation with Thomas Münster.

The dataset contained 20 *MPF2*-like upstream sequences (sequences no. 1 to 20, Table 3.7.1) from 13 Physaleae, one *Capsicum* and one *Solanum* species. The latter, the upstream sequence of *STMADS16* from *S. tuberosum*, was used as an out-group. Multiple alignments of the upstream sequences were calculated by CHAOS and DIALIGN (Brudno *et al.* 2004). Afterwards the alignment was cut to the size of the shortest upstream sequence (912 bp from *C. baccatum*) and further manually adjusted. The final alignment can be found in the supplementary Table 8.2.3.

Figure 3.8.1 shows a Maximum Parsimony tree, which was based on partial upstream sequences of *MPF2*-like MADS-box genes. The tree topology suggests that the upstream sequences of *MPF2*-like genes from Physaleae can be subdivided into three groups: 1. *MPF2*-like Physalinae upstream sequences and 2. *MPF2*-like *W. coccoloboides* upstream sequences and 3. the more basal *MPF2*-like Iochrominae upstream sequences.

Also in the phylogeny based on upstream non-coding sequences the putative allelic sequences found in *W. solanacea, V. breviflora* and *W. coccoloboides* clustered together in each case, which supported the assumption that they might be alleles.

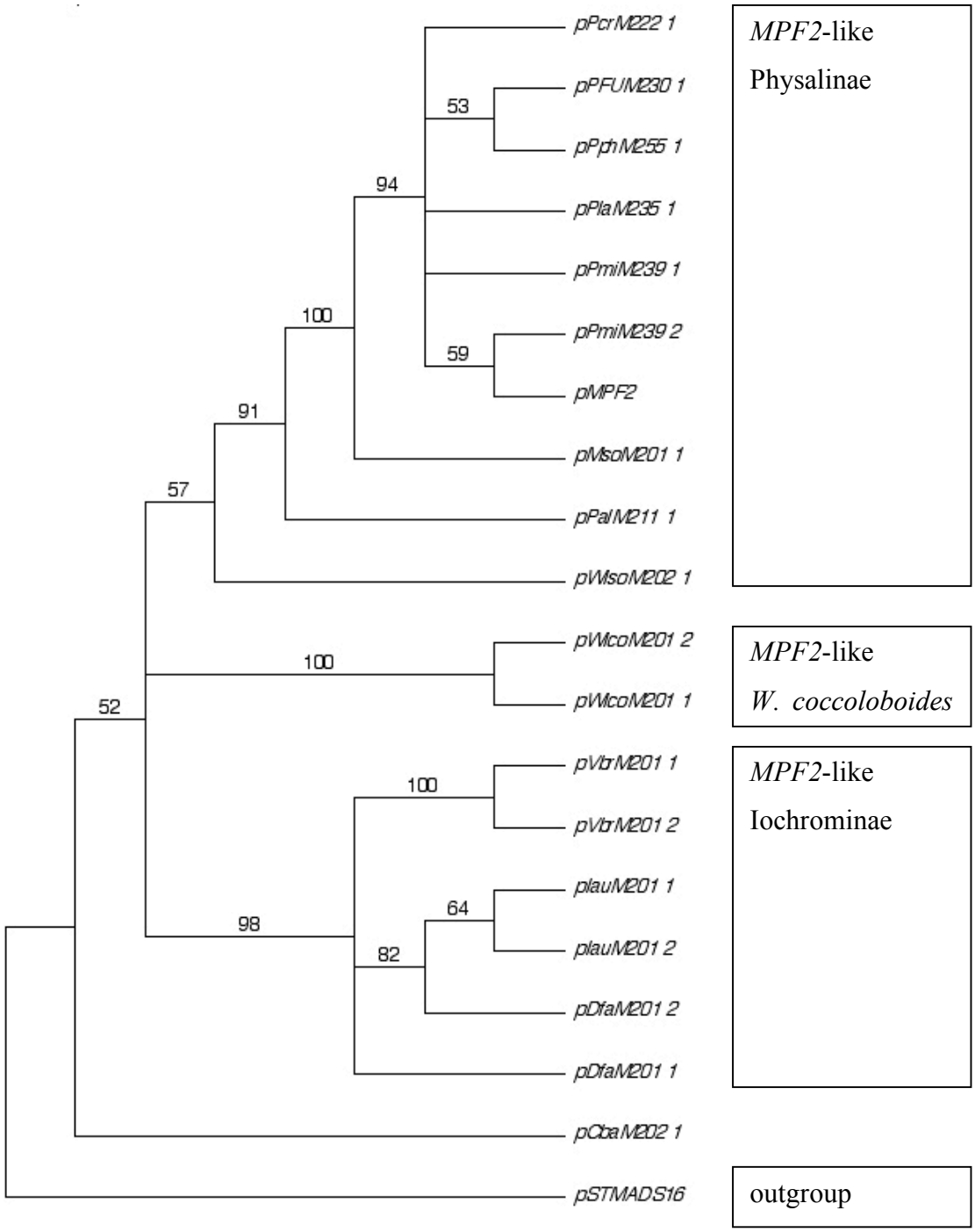


Figure 3.8.1: Phylogenetic tree of *MPF2*-like upstream sequences from Physaleae

Maximum Parsimony tree of *MPF2*-like upstream sequences from Physaleae and *Capsicum*, with the upstream sequence of *STMADS16* from *S. tuberosum* as out-group. Bootstrap values are shown. For accession numbers see Table 3.7.1.

3.9 Transcription initiates between – 690 bp and – 300 bp

The putative transcription initiation sites were inferred from the longest cDNAs obtained by 5'RACE (see Table 3.9.1 and Fig. 3.10.1). The published *MPF2* and *STMADS16* sequences (He and Saedler 2005) were taken for comparison. Transcription appears to initiate between position – 690 bp and – 300 bp and one or two exons were present in the 5'UTR.

The 5'UTR was included in the promoter analyses, since results from another study suggested that this region might also exhibit promoter activity (Khan 2009).

Table 3.9.1: Putative transcription start and 5'UTR

Species	Gene/allele	Putative transcription start	No. of exons in 5'UTR
P. crassifolia	*PcrM222_1*	- 472	1
P. fuscomaculata	*PfuM230_1*	- 384	2
P. lanceifolia	*PlaM235_1*	- 473	2
P. minima	*PmiM239_1*	- 475	2
	PmiM239_2	- 308	1
P. philadelphica	*PphM255_1*	- 358	2
P. floridana	*MPF2*	(- 470)*	(2)*
M. solanaceus	*MsoM201_1*	- 300	1
W. solanacea	*WIsoM202_1*	- 401	2
P. alkekengi	*PalM211_1*	- 533	2
D. fasciculata	*DfaM201_1*	- 456	2
	DfaM201_2	- 531	2
I. australe	*IauM201_1*	- 426	2
	*IauM201_2**	- 425	2
V. breviflora	*VbrM201_1*	- 447	2
	VbrM201_2	- 357	2
W. coccoloboides	*WIcoM201_1*	- 462	2
	WIcoM201_2	- 462	2
C. baccatum	*CbaM202_1**	- 690	2
S. tuberosum	*STMADS16*	(- 469)*	(2)*

Data inferred from published sequences from He and Saedler (2005) are marked with a "*", for accession numbers see supplementary Tables 8.1.2 and 8.1.3. Species, which do not express *MPF2*-like genes in sepals (He and Saedler 2005, Hu and Saedler 2007), are marked in red.

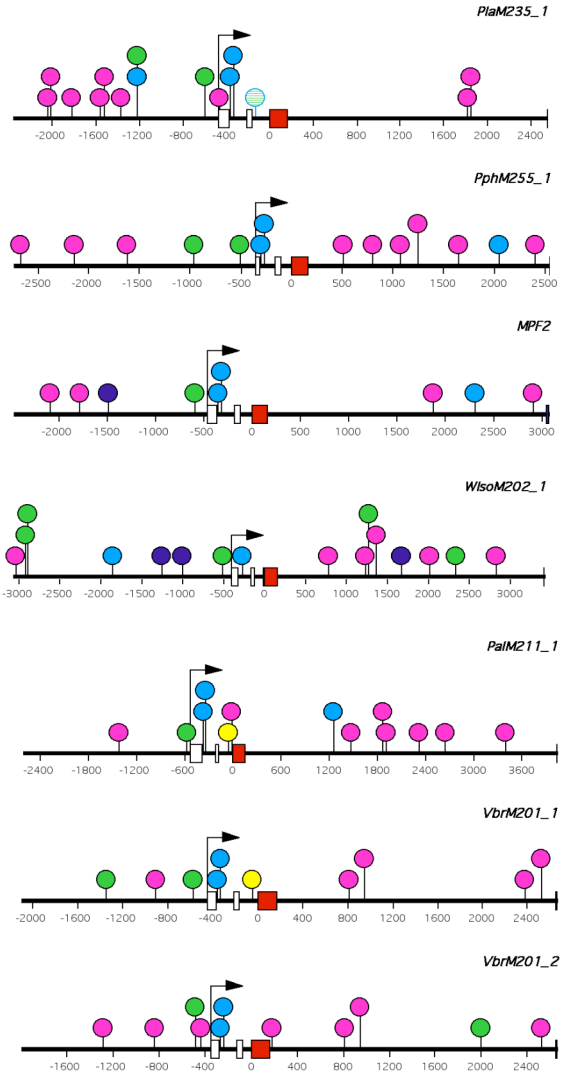
3.10 All *MPF2*-like upstream sequences contain CArG-motifs

It has been hypothesized that the absence of strictly defined MEF2- / N10-type CArG-boxes (CTA(A/T)$_4$TAG; Pollock and Treisman 1991, Shore and Sharrocks 1995) in the upstream sequence of *MPF2* from *P. floridana* could be causal for heterotopic expression of this gene in floral tissues of this species (He and Saedler 2005). Therefore, those ten upstream sequences, for which also the first introns were available, were screened manually for CArG-consensus motifs using the computer program MacVector. However, published data on plant CArG-motifs came almost exclusively from *Arabidopsis* and *Antirrhinum*, which are only distantly related to our species. Thus very wide definitions of CArG-motifs were used for this initial search, since it could not be expected that these motifs showed 100% identity across species borders. The chosen CArG-box motifs were:

1. MEF2- or N10-type (relaxed): C(A/T)$_8$G,
2. MEF2- or N10-type (strict): CTA(A/T)$_4$TAG (Pollock and Treisman 1991, Shore and Sharrocks 1995),
3. the serum response element (SRE)-type: CC(A/T)$_6$GG (Pollock and Treisman 1991, Riechmann *et al.* 1996)
4. an intermediate CArG motif: C(A/T)$_7$GG (Tang and Perry 2003)
5. an intermediate CArG motif: CC(A/T)$_7$G (Tang and Perry 2003)

Not surprisingly, CArG-motifs were found in all screened upstream sequences and first introns (see Fig. 3.10.1). Especially “relaxed” N10-type CArG-boxes occurred quite frequently. However, the MEF2- or N10-type CArG-motifs, in strict sense, were only found in four upstream sequences (*pPalM211_1, pVbrM201_1, pDfaM201_1* and *pSTMADS16*) and in none of the first introns. In the upstream sequence of *STMADS16* three of these motifs were detected, while in the other three sequences only one strict N10-type CArG-box was found right before the ATG. No obvious correlation between the *MPF2*-like gene expression pattern in sepals and the occurrence of strict N10-type CArG-boxes (or any of the other searched CArG-motifs) could be detected in Physaleae. Nevertheless, an interesting observation made was the similar CArG-motif distribution around the putative transcription initiation sites: the first intron in the 5’UTR is flanked by one CC(A/T)$_7$G motif upstream and two C(A/T)$_7$GG motifs downstream (except for *pWIsoM202_1*, where the second C(A/T)$_7$GG downstream is missing, which is caused by only one nucleotide difference).

Species with expression in sepals:



Species without expression in sepals:

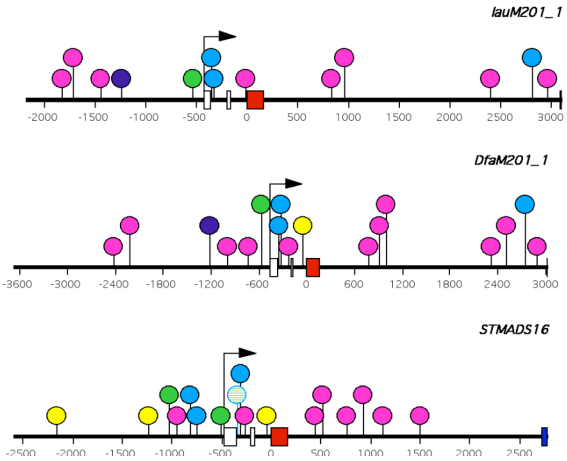


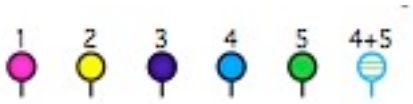
Figure 3.10.1: CArG-motifs in *MPF2*-like upstream and intronic sequences

The partial *MPF2*-like loci (upstream sequence and first intron) are indicated by black lines. The exons are marked as boxes (5'UTR: white, MADS-box: red, the I-region: dark blue or black). The putative start of transcription was drawn from the longest 5'RACE product and is indicated by an arrow.

The CArG-box motifs are depicted as pins:

1. $C(A/T)_8G$,
2. $CTA(A/T)_4TAG$
3. $CC(A/T)_6GG$
4. $C(A/T)_7GG$
5. $CC(A/T)_7G$

4.+5. $CC(A/T)_7GG$



3.11 Conserved regions and motifs

The isolated *MPF2*-like upstream and intronic sequences were analyzed with the programs DIALIGN, Mulan and multiTF to identify (i) conserved regions and putative *cis*-regulatory elements in these, which might be involved in the transcriptional regulation of vegetative *MPF2*-like gene expression and further (ii) divergent regions and the putative *cis*-regulatory elements located there, as they might be responsible for differences in gene expression in sepals.

DIALIGN and PLACE analysis of upstream sequences

In the credo DIALIGN analysis output (see Fig. 3.11.1) conserved motifs, which were found in two or more upstream sequences, were depicted as colored boxes. Therefore, larger conserved regions containing several conserved motifs could be identified easily. The original output (see Fig. 3.11.1) and a schematic drawing of the conserved regions are shown (see Fig. 3.11.2)

After analyzing all 20 *MPF2*-like upstream sequences with DIALIGN, it became clear, that when different upstream sequences with a similar absolute length were compared, the "informative length" at the motif level was extremely variable. In some of the *MPF2*-like upstream sequences the motifs appeared to be much more compactly arranged than in others (e.g. *V. breviflora vs. D. fasciculata*). Even sequences from the same or closely related species often differed dramatically in size due to large indels (e.g. *pDfaM201_1* and *pDfaM201_2* from *D. fasciculata*). Based on the motifs found at the 5'upstream end of putative promoter sequences, it was concluded that the sequences, which were isolated from the American *Physalis* and *W. coccoloboides*, were the "longest" ones (see Fig. 3.11.1: green and light blue circled regions).

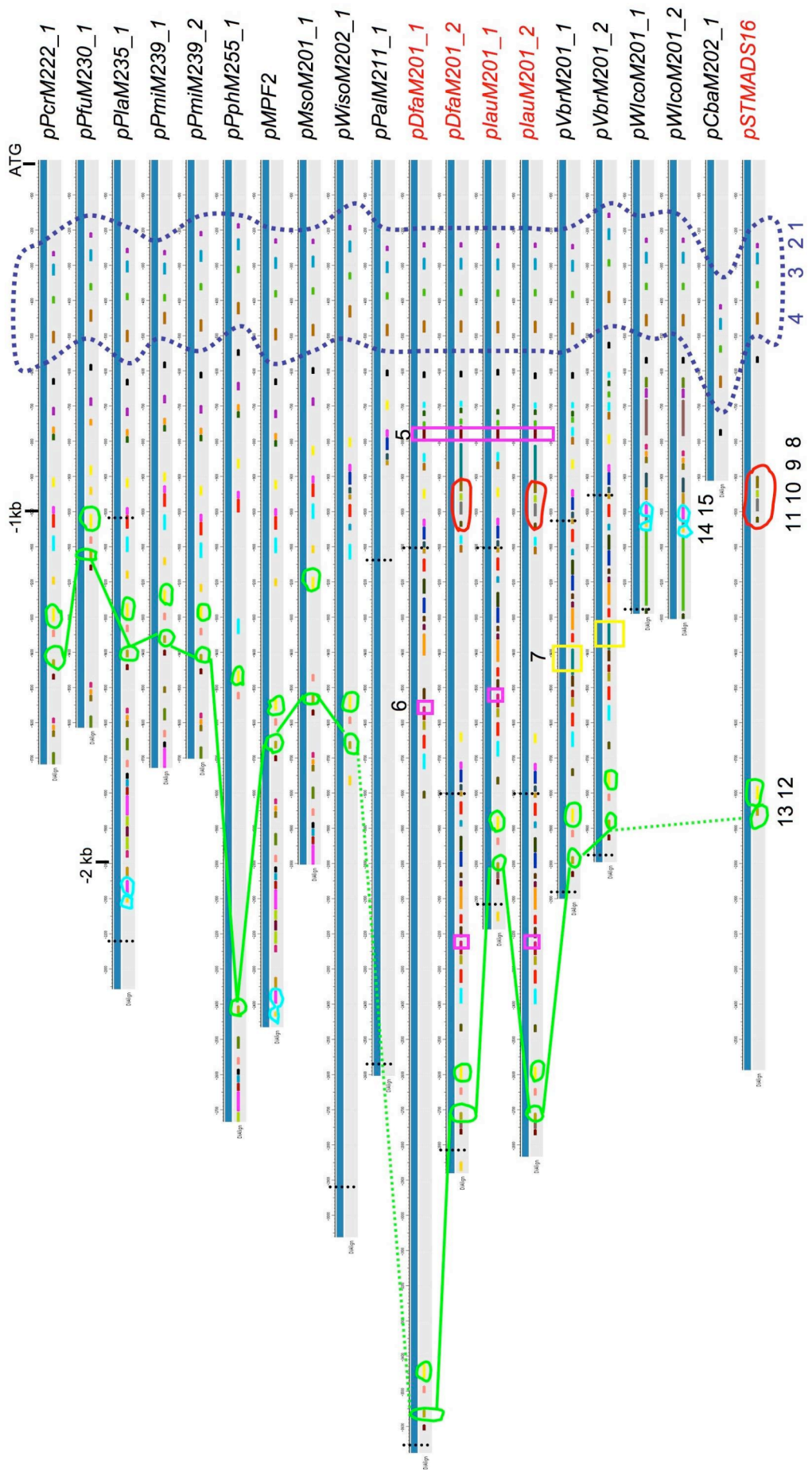

pPcrM222_1
pPfuM230_1
pPlaM235_1
pPmiM239_1
pPmiM239_2
pPphM255_1
pMPF2
pMsoM201_1
pWisoM202_1
pPalM211_1
pDfaM201_1
pDfaM201_2
plauM201_1
plauM201_2
pVbrM201_1
pVbrM201_2
pWlcoM201_1
pWlcoM201_2
pCbaM202_1
pSTMADS16
ATG
-1kb
-2 kb
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 3.11.1: Conserved and divergent motifs in *MPF2*-like upstream sequences

A credo DIALIGN analysis output (default settings + Phylogenetic footprinting I + motif size = 16) is shown. Sequences from species, which show no expression of *MPF2*-like genes in sepals, are written in red. The overall conserved motifs 1, 2, 3 and 4 are surrounded by a dark blue dashed line, while motifs only conserved in *I. australe* and *D. fasciculata* sequences are circled in pink (motifs 5 and 6). Four motifs at the upstream end, which are conserved in several species, are marked in green (motifs 12 and 13) and light blue (motifs 14 and 15) for better orientation. Another motif-rich region conserved between two Iochrominae sequences and *pSTMADS16* is marked in red (motifs 9, 10 and 11). Fragments taken for promoter::*GUS* constructs are indicated by vertical dashed black lines.

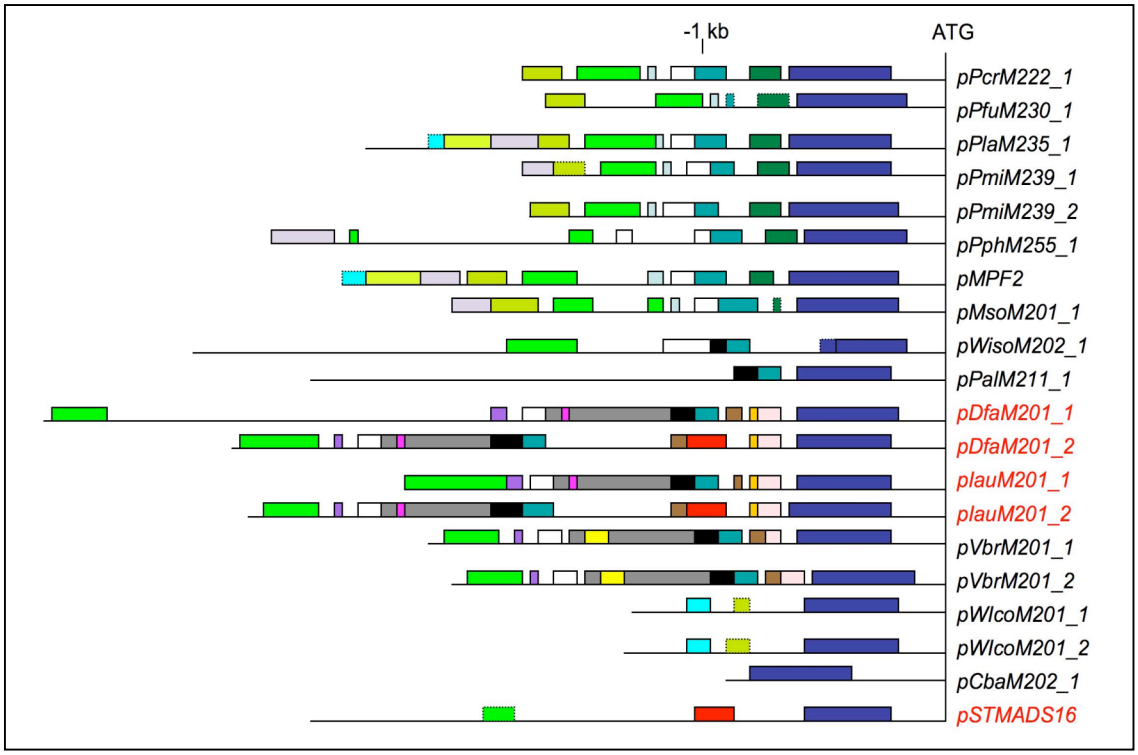


Figure 3.11.2: Conserved motif-rich regions in *MPF2*-like upstream sequences

The schematic drawing is based on the DIALIGN analysis output in Fig. 3.11.1. Motif-rich regions are depicted as boxes. Sequences from species, which show no expression of *MPF2*-like genes in sepals, are written in red. An overall conserved region is marked in dark blue, while regions containing motifs only conserved in *I. australe* and *D. fasciculata* are depicted as dark pink and orange boxes.

The credo DIALIGN analysis indicated that an over 300 bp long block (in dark blue, see Fig. 3.11.1 and Fig. 3.11.2), which was located between positions – 150 and – 650 bp upstream of the MADS-box and comprised basically the 5'UTR regions, was conserved in all 20 *MPF2*-like upstream sequences. This block consisted of four

motifs, which were conserved in all species (see Fig. 3.11.1). Their consensus sequences are shown in Fig. 3.11.3.

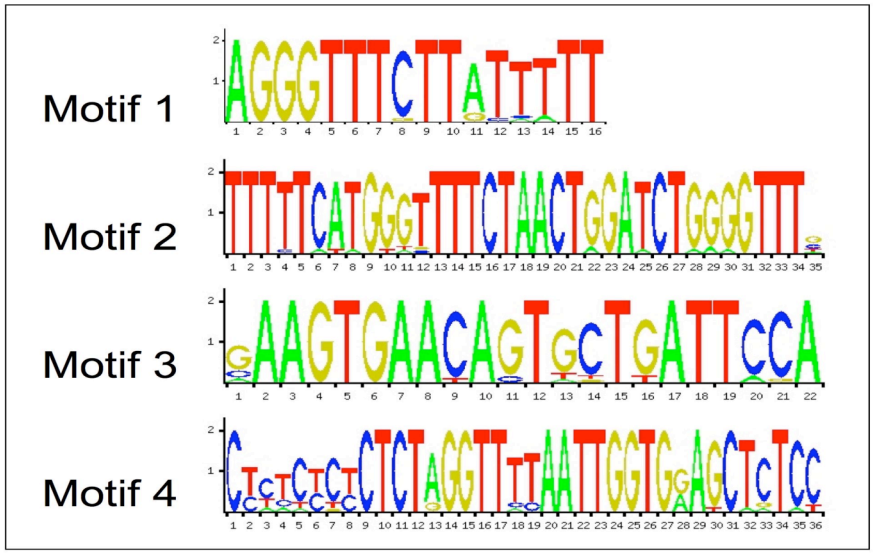


Figure 3.11.3: Overall conserved motifs in *MPF2*-like upstream sequences

The motifs were identified by credo DIALIGN analysis. The default settings + Phylogenetic footprinting = I and minimum motif size = 16 were used. Motif positions in upstream sequences of *MPF2*, *VbrM201_1* and *IauM201_1* are given: Motif 1: *MPF2*: -238 to -223, *VbrM201_1*: -254 to -239, *IauM201_1*: -251 to -236; Motif 2: *MPF2*: -310 to –276, *VbrM201_1*: -318 to -284, *IauM201_1*: -314 to -280; Motif 3: *MPF2:* -391 to –370, *VbrM201_1*: -393 to -372, *IauM201_1*: -389 to -368; Motif 4: *MPF2:* -513 to –479, *VbrM201_1*: -497 to -462, *IauM201_1*: -493 to 458

The motifs found were screened in the PLACE database, which contains a large collection of published *cis*-regulatory elements. Usually PLACE searches yield a high number of short motifs, which occur in high frequency in genomic sequences. Therefore in the entire results part of this thesis always only a selection of the motifs found by PLACE is mentioned. These are all motifs found in the + strand, which consist of five or more base pairs and are overall conserved.
The PLACE search indicated that motif 2 featured a binding site ("TAACTG") for AtMYB2, which is an *Arabidopsis* MYB homolog (Urao *et al.* 1993).

Several other motif-rich regions, which were conserved in several species, were identified (see Fig. 3.11.1). However, the conserved region before the ATG was the only one, which was overall conserved in all analyzed sequences, suggesting that this

region might be the core promoter, while the elements responsible for "fine-tuning" differential gene expression in floral tissues might be located elsewhere.

When all 20 sequences were included in the analysis, no motifs were found to be strictly present or absent in correlation to the presence or absence of *MPF2*-like gene expression in sepals of the source species. However, as the loss or gain of expression does not necessarily have to be caused by changes in the same element(s) in all Physaleae and Solanaceae, the motif distribution in the different subgroups was examined separately.

Physalinae

All American *Physalis* species shared very similar *MPF2*-like upstream sequences, reflecting their close relationship. By contrast, *pPalM211_1*, the upstream sequence from the geographically isolated Old World species *P. alkekengi*, seemed to be very different from the other *Physalis* sequences: only approximately 850 bp upstream of the ATG shared similarities with other *Physalis* sequences. Further upstream than –850 bp no similarities between *pPalM211_1* and any other sequences from our dataset were found. This finding suggested that the core promoter of *MPF2*-like genes in *Physalis* might comprise only this short region. However, the presence of large indels in many upstream sequences offered an alternative explanation: that a large insertion could have occurred in the upstream sequence of the *P. alkekengi*. Unfortunately, no motifs, which are conserved in all Physalinae, but not present in upstream sequences from *D. fasciculata* and *I. australe*, were detected.

Iochrominae

Also the Iochrominae possessed very similar *MPF2*-like upstream sequences. The most prominent differences between the sequences from different Iochrominae were several large indels (up to 1.5 kb in size in *D. fasciculata*), which were mainly responsible for the differences in length among the isolated upstream sequences. In the sequences from *D. fasciculata* and *I. australe* two conserved motifs (motif 5 and 6, see Fig. 3.11.1 and Fig. 3.11.4) were detected, which were not present in any other sequences analyzed and might be responsible for their loss of gene expression in sepals. Further, the sequences from *V. breviflora*, which shows *MPF2*-like gene

expression in sepals, contained both one motif (motif 7, see Fig. 3.11.1 and Fig. 3.11.4), which was not present in the other Iochrominae sequences.

The motifs found were screened in the PLACE database for published *cis*-regulatory elements and two interesting sites were detected: Motif 5 featured e.g. the elements "AGAAA", one of two co-dependent regulatory elements responsible for pollen specific activation of the *S. lycopersicum* gene *lat52* (Filichkin *et al.* 2004). Motif 6 contained the sequence "GGATA", the core motif of the transcriptional activator MybSt1 (a potato MYB homolog) binding site (Baranowskij *et al.* 1994).

Interestingly, the only *Solanum* sequence, *pSTMADS16*, possessed in addition to the overall conserved block proximal to the MADS-box also two other regions, which showed similarity to regions in *MPF2*-like upstream sequences from Physaleae: around –1 kb (motifs 8 to 11, red circle, see Fig. 3.11.1) and –1.8 kb (motifs 12 and 13, light green circle, see Fig. 3.11.1). The motifs found in the region around -1 kb occurred only in sequences from species showing no gene expression in sepals: *pSTMADS*, *pDfaM201_2* and *pIauM201_2*. However, these motifs were not present in *pDfaM201_1* and *pIauM201_1*, the other alleles from *D. fasciculata* and *I. australe*.

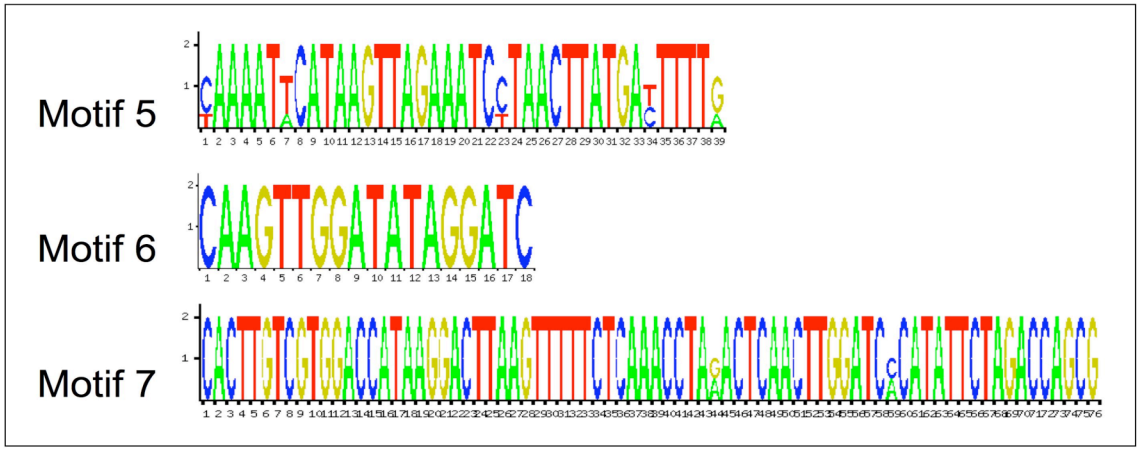


Figure 3.11.4: Divergent motifs in *MPF2*-like upstream sequences from Iochrominae

The motifs were identified by credo DIALIGN analysis. The default settings + Phylogenetic footprinting = I and minimum motif size = 16 were used. Motifs 5 and 6 were detected in sequences from *I. australe* and *D. fasciculata*, while motif 7 only occurred in *V. breviflora*. Motif positions in upstream sequences of *VbrM201_1* and *IauM201_1* are given: Motif 5: *IauM201_1*: -797 to -759; Motif 6: *IauM201_1*: -1534 to -1517; Motif 7: *VbrM201_1*: -1464 to -1389

Mulan and multiTF analysis of upstream sequences

Mulan generates local multiple DNA sequence alignments and the multiTF program can identify transcription factor binding sites evolutionarily conserved across multiple species.

Also the Mulan analysis confirmed that *pPalM211_1*, the *MPF2*-like upstream sequence from *P. alkekengi*, showed only sequence similarity to other sequences in the dataset until approximately -850 bp upstream of the MADS-box (see Fig. 3.11.5).

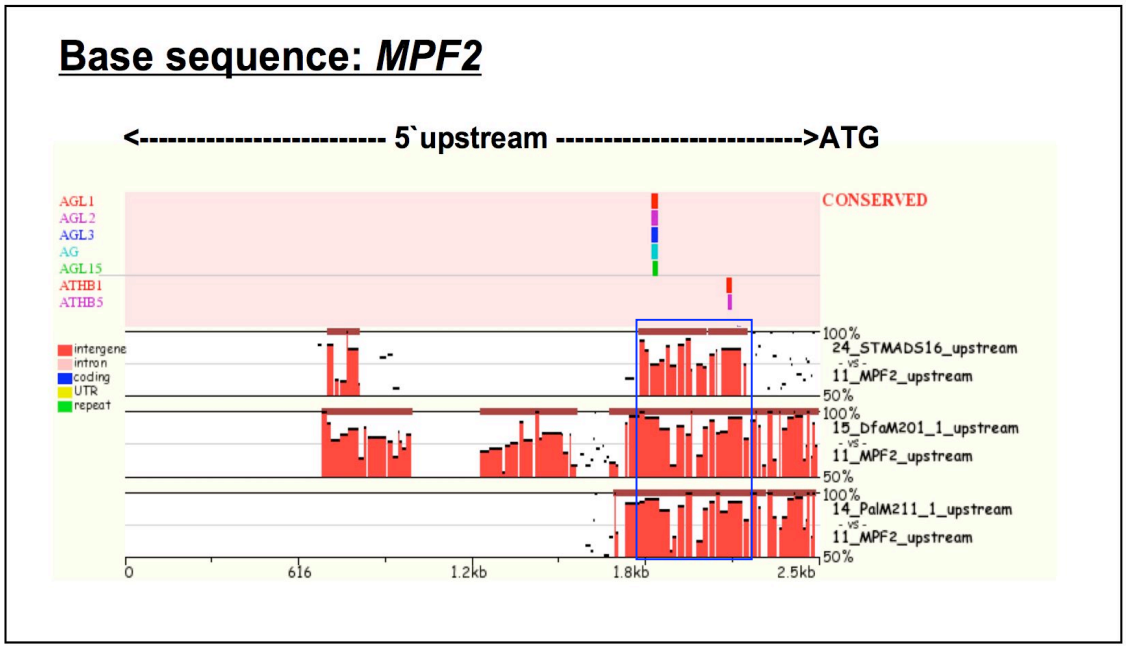


Figure 3.11.5: Conserved regions and transcription factor binding sites in *MPF2*-like upstream sequences

The conservation profiles between *pMPF2* and *pPalM211_1*, *pDfaM201_1* and *pSTMADS16* are graphically visualized. The default settings (ECR length = at least 100 bases; ECR similarity = at least 70%, Bottom cut-off = 50%, ECR present in = 1 species) were used. The lowest line indicates 50% and the upper one 100% conservation. A blue box indicates an overall conserved region. Small colored boxes indicate the locations of overall conserved transcription factor binding motifs, which were identified by the program multiTF. For accession numbers see Table 3.7.1.

With the help of the program multiTF several sites containing published transcription factor binding motifs were detected in this region. Two transcription factor binding sites were conserved in all 20 upstream sequences (see Fig. 3.11.5 and Fig.3.11.6): a CArG-box, which was referred to be a binding sites for the *Arabidopsis* MADS-domain proteins AGAMOUS-LIKE 1 (AGL1) / SHATTERPROOF 1 (SHP1), AGL2,

AGL3, AGAMOUS-LIKE 15 (AGL15) and AG and further a binding site for ATHB1 and ATHB5 transcription factors (see Fig. 3.11.6), which are homeodomain-leucine zipper class I (HD-Zip I) proteins. The 9 bp long binding site for ATHB-like transcription factors was located within the previously mentioned second CArG-like motif (C(A/T)$_7$GG) downstream of the first exon in the 5'UTR (see chapter 3.10, Fig. 3.10.1).

The analysis of a subset containing only upstream sequences from Physaleae, yielded no additional "Physaleae specific" conserved transcription factor binding sites, since the additional motif (AGP1), which was found, was only absent in *pCbaM202_1*, but not in *pSTMADS16*.

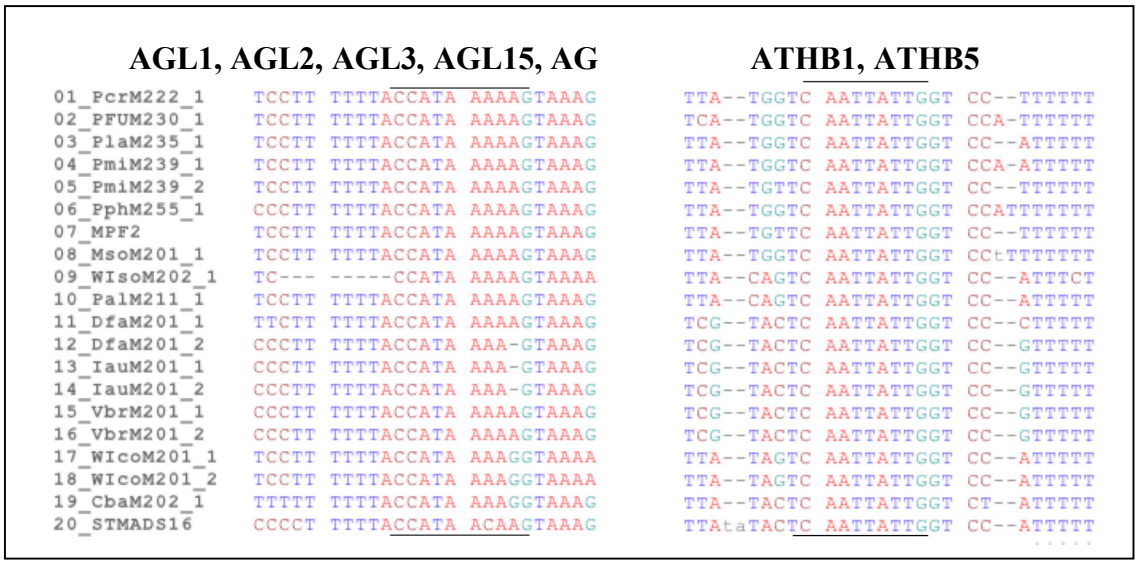


Figure 3.11.6: Overall conserved transcription factor binding sites

The graph shows the section of the alignment, where the conserved transcription factor binding sites are located. The CArG-core "CCATAA(A/C)A(A/G/_)G" and the core of the ATHB binding motif "CAATTATTG" are indicated by black lines. Motif positions in upstream sequences of *MPF2*, *VbrM201_1* and *IauM201_1* are given: CArG-box: *MPF2*: -590 to 581, *VbrM201_1*: -573 to -564, *IauM201_1*: -564 to 556; ATHB1/5: *MPF2*: -310 to –276, *VbrM201_1*: -332 to 324, *IauM201_1*: -314 to –280.

For the detection of differences in the distribution of transcription factor binding sites, which could be correlated with expression differences, the dataset was split into different subsets, which were then analyzed with Mulan and multiTF. Conserved binding sites found were rechecked manually in a DIALIGN alignment to ensure that these sites only occurred exclusively in this subset.

The binding sites, which were found to be conserved in sequences from species exhibiting *MPF2*-like gene expression in sepals (different subsets: 1. all, 2. Physaleae without *Capsicum*, 3. only Physalinae), were also present in upstream sequences from *S. tuberosum*, *D. fasciculata* and/or *I. australe*. The "putative promoter" sequences from the latter three species featured six conserved transcription factor binding sites, but all of these motifs were also found in sequences from Physaleae species, which express *MPF2*-like genes in their sepals. Therefore it seemed not to be very likely that these conserved sites could be relevant for differential *MPF2*-like gene expression in sepals.

However, the four upstream sequences from *D. fasciculata* and *I. australe* contained two conserved transcription factor binding sites, which were not detected at these positions in other upstream sequences: an AUXIN RESPONSE FACTOR (ARF) binding site and a binding site for AGL1, AGL2, AGL3 and AG (see Fig. 3.11.7). In the sequences from *V. breviflora* these binding sites showed only one nucleotide difference, in all other sequences they were completely absent.

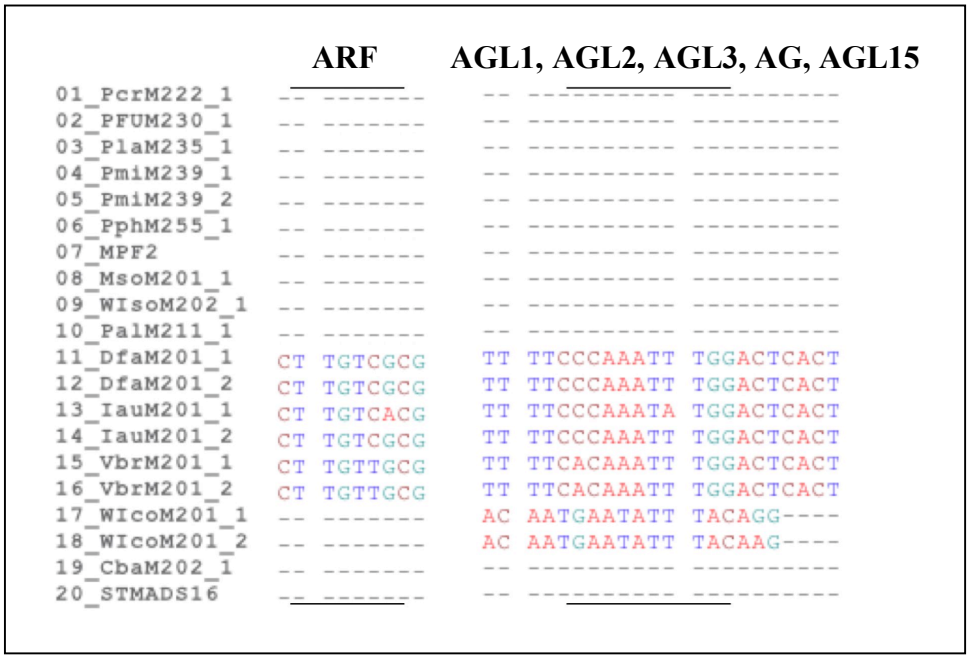


Figure 3.11.7: Divergent transcription factor binding sites in Iochrominae

The graph shows the section of the alignment, where the transcription factor binding sites, which are only conserved in *D. fasciculata* and *I. australe*, are located. The CArG-core CCAAAT(A/T)TGG and the core of the ARF binding motif are indicated by black lines. The sites were found to be conserved in sequences from *I. australe* and *D. fasciculata*: ARF: *IauM201_1*: -1374 to -1366; CArG-box: *IauM201_1*: -1247 to -1238

DIALIGN and PLACE analysis of first intron sequences

The 10 first intron sequences of *MPF2*-like genes from nine Physaleae species and *S. tuberosum* were analyzed with DIALIGN. The original graphical visualization, containing all conserved motifs found (see Fig. 3.11.8), and a simplified scheme indicating conserved regions (see Fig. 3.11.9) are shown on the following pages.

As expected the sequences from the American *Physalis* species seemed to be very conserved, as well as the sequences from the Iochrominae, while the intronic sequence from the distantly related potato showed only few homologous regions to introns from Physaleae species.

The analysis revealed seven overall conserved motifs (see Fig. 3.11.8: motifs 1 to 7, circled in dark purple), which were present in all ten intron sequences. Their consensus sequences are shown in Fig. 3.11.10. These elements could be involved in the transcriptional regulation of vegetative expression of *MPF2*-like genes.

Further, the analysis revealed also motifs, whose presence or absence could be correlated with the presence and absence of *MPF2*-like transcripts in sepals of the source species. Although no motifs were found, which were only conserved in all Physaleae species, which feature expression of *MPF2*-like genes in sepals, five motifs were found to be conserved in the first introns from the Physalinae subset (without *W. coccoloboides*). Hence, these motifs could be responsible for heterotopic *MPF2*-like gene expression in this clade (see Fig. 3.11.8: motifs 11 to 15, marked in orange). The motif consensus sequences are depicted in Fig. 3.11.11.

Furthermore, three motifs in a region at the downstream end of the first introns were only conserved between *I. australe* and *D. fasciculata* (see Fig. 3.11.8, motifs 8 to 10, in pink). Therefore these motifs were also considered to be interesting in terms of differential gene expression. Their consensus sequences are shown in Fig. 3.11.12.

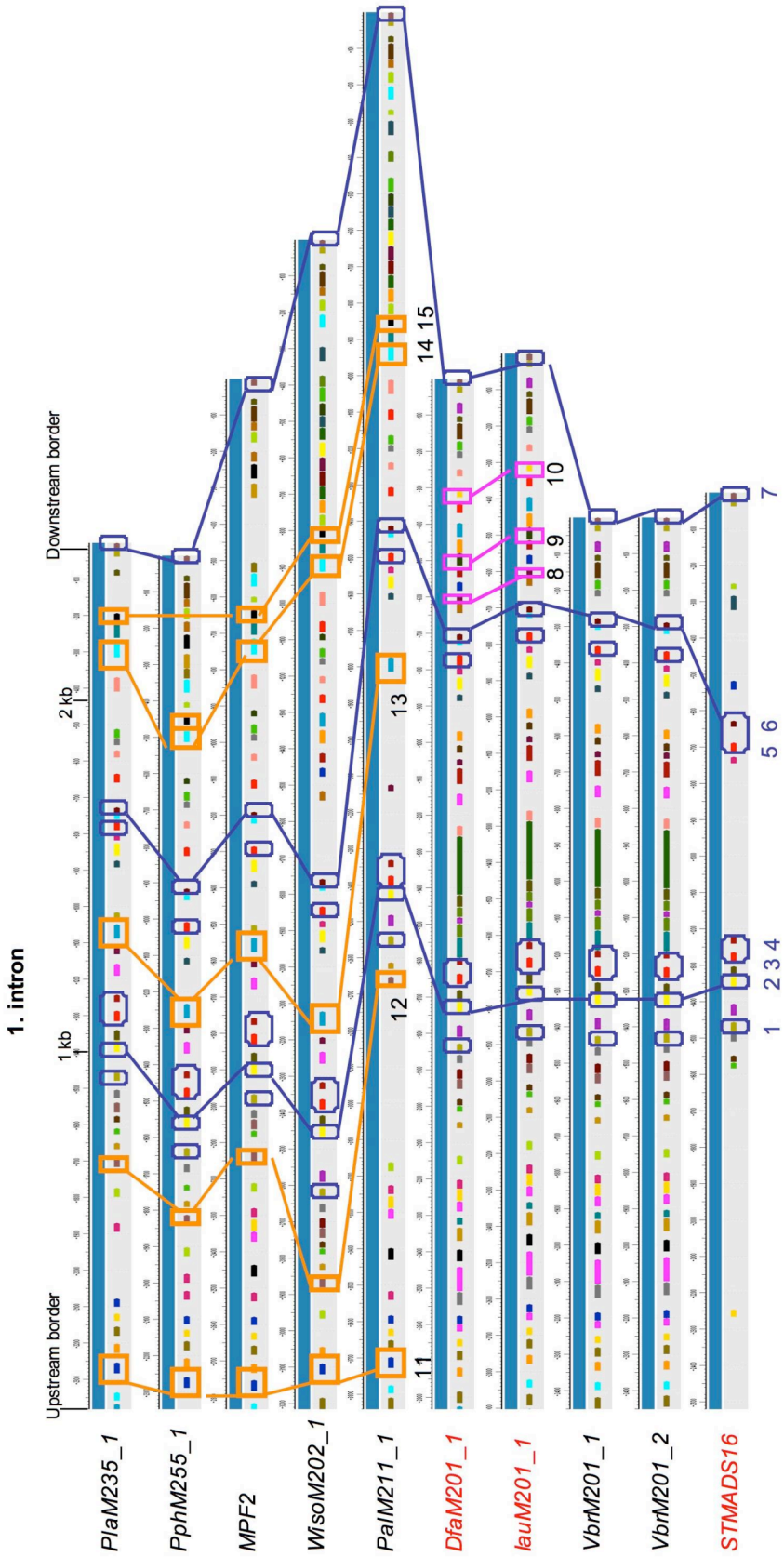

1. intron
Upstream border
1 kb
2 kb
Downstream border
PlaM235_1
PphM255_1
MPF2
WisoM202_1
PalM211_1
DfaM201_1
lauM201_1
VbrM201_1
VbrM201_2
STMADS16
1 2 3 4
5 6
7
8 9
10
11
12
13
14 15

Figure 3.11.8: Conserved and divergent motifs in the first intron of *MPF2*-like genes

A credo DIALIGN analysis output (default settings + Phylogenetic footprinting I + motif size = 16) is shown. Sequences from species, which show no expression of *MPF2*-like genes in sepals, are written in red. The overall conserved motifs 1 to 7 are marked in dark blue, while motifs only conserved in *I. australe* and *D. fasciculata* sequences are circled in pink (motifs 7 to 9). Motifs conserved in Physalinae are shown in orange (motif 11 to 15).

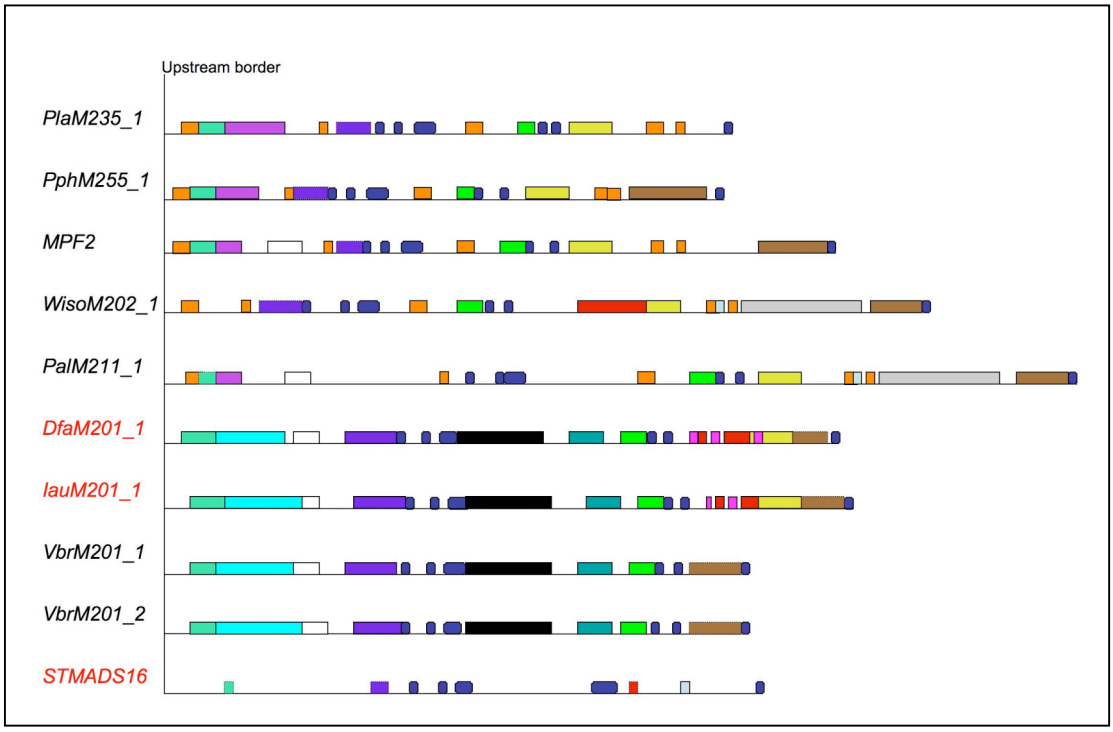


Figure 3.11.9: Conserved motif-rich regions in first introns of *MPF2*-like genes

The schematic drawing is based on the DIALIGN analysis output in Fig. 3.11.8. Motif-rich regions are depicted as boxes.

For the overall conserved motifs 1 to 7 (see Fig. 3.11.10) searches in the PLACE database were performed to see whether these motifs contain the consensus sequences of known *cis*-regulatory elements. However, in none of the overall conserved motifs a known binding site, which was longer than 4 bp, could be detected.

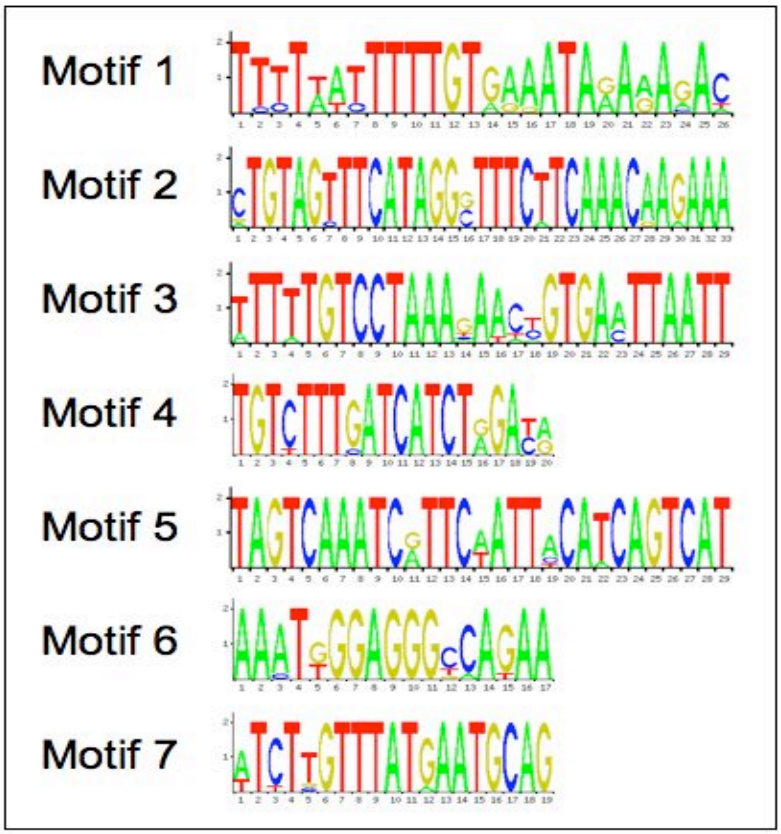


Figure 3.11.10: Overall conserved motifs in first introns of *MPF2*-like genes

The motifs were identified by credo DIALIGN analysis. The default settings + Phylogenetic footprinting I and minimum motif size = 16 were used. Motif positions in first introns of *MPF2*, *VbrM201_1* and *IauM201_1* are given: Motif 1: *MPF2*: 1051 to 1076, *VbrM201_1*: 1196 to 1221, *IauM201_1*: 1211 to1236; Motif 2: *MPF2*: 1130 to 1162, *VbrM201_1*: 1308 to 1340, *IauM201_1*: 1323 to 1355; Motif 3: *MPF2:* 1214 to 1242, *VbrM201_1*: 1385 to 1413, *IauM201_1*: 1400 to 1428; Motif 4: *MPF2:* 1267 to 1286, *VbrM201_1*: 1437 to 1456, *IauM201_1*: 1452 to 1471; Motif 5: *MPF2:* 1729 to 1757, *VbrM201_1*: 2264 to 2292, *IauM201_1*: 2294 to 2322, Motif 6: *MPF2:* 1836 to 1852, *VbrM201_1*: 2353 to 2369, *IauM201_1*: 2377 to 2393, Motif 7: *MPF2:* 3024 to 3042, *VbrM201_1*: 2627 to 2645, *IauM201_1*: 3070 to 3088

Three motifs (motifs 8 to 10, see Fig. 3.11.11) were found to be specific for introns of *I. australe* and *D. fasciculata*. The search in the PLACE database yielded the following binding sites: motif 8 contained the sequence “TTATCC”, a "sugar-repressive element (SRE)" (Tatematsu *et al.* 2005), motif 9 featured the motif “CACCTG”, which is a recognition site for the *Arabidopsis* transcription factor RAV1 (Kagaya *et al.* 1999) and motif 10 exhibits a binding site (“TACTATT”) for the nuclear factor SP8BF (Ishiguro and Nakamura 1992).

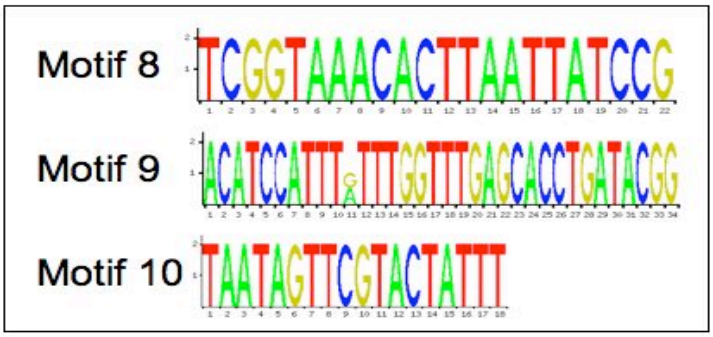


Figure 3.11.11: Motifs conserved only in first introns of *MPF2*-like genes from *D. fasciculata* and *I. australe*

Motifs identified by credo DIALIGN analysis. The motifs were identified by credo DIALIGN analysis. The default settings + Phylogenetic footprinting I and minimum motif size = 16 were used. Motif positions in the first intron of *IauM201_1* are given: Motif 1: *IauM201_1*: 2477 to 2498; Motif 2: *IauM201_1*: 2578 to 2611; Motif 3: *IauM201_1*: 2765 to 2782

Also the motifs, which were conserved in the introns from Physalinae species, were screened in the PLACE database: Motif 12 contained the consensus "GANTTNC", which is a binding site of the MYB transcription factor LCR1 (Yoshioka *et al.* 2004).

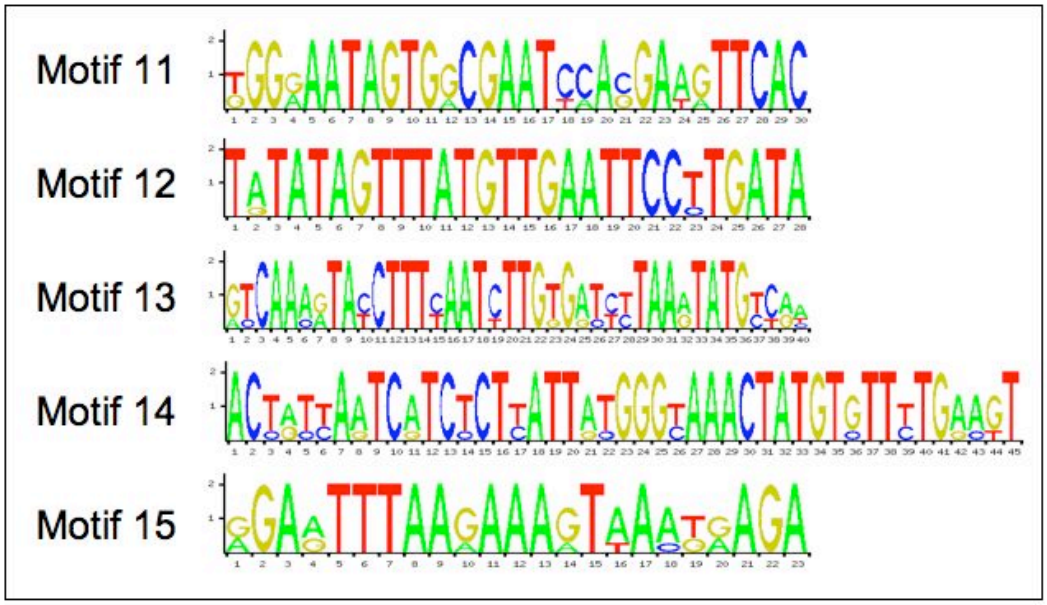


Figure 3.11.12: Motifs conserved in first introns of *MPF2*-like Physalinae genes

Motifs identified by credo DIALIGN analysis. The motifs were identified by credo DIALIGN analysis. The default settings + Phylogenetic footprinting I and minimum motif size = 16 were used. Motif positions in the first intron of *MPF2* are given: Motif 11: *MPF2*: 261 to 290; Motif 12: *MPF2*: 890 to 917; Motif 13: *MPF2*: 1467 to 1506, Motif 14: *MPF2*: 2285 to 2329; Motif 15: *MPF2*: 2385-2407

Mulan and multiTF analysis of first intron sequences

Like the upstream sequences also the MADS-box and first intron sequences of *MPF2*-like genes were analyzed with Mulan and multiTF to identify transcription factor binding sites evolutionarily conserved across multiple species.

The Mulan analysis showed that the intron contains several overall conserved regions and not surprisingly that the MADS-box is conserved across species. All MADS-boxes contained a BZIP910 motif, which is a binding site for a bZIP transcription factor from *A. majus* (Martinez-Garcia *et al.* 1998). In the intron no overall conserved transcription factor binding sites were detected, when the intron from *S. tuberosum* was included, but all Physaleae sequences contained a conserved binding site for the INDETERMINATE protein, ID1 (see Fig. 3.11.13). However, in *S. tuberosum* this motif showed only one nucleotide difference.

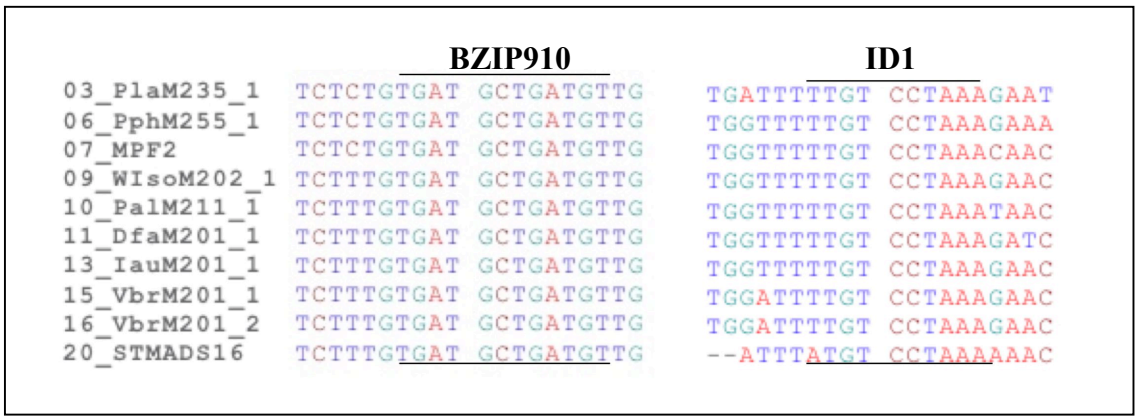


Figure 3.11.13: Conserved transcription factor binding sites

The graph shows the section of the alignment, where the overall conserved transcription factor binding site BZIP910 and the ID1 binding site, which is conserved in Physaleae, are located. The positions of the motifs are indicated in the alignment by black lines. Motif positions in *MPF2*, *VbrM201_1* and *IauM201_1* are given: BZIP910: *MPF2*: 117 to 128, *VbrM201_1*: 117 to 128, *IauM201_1*: 117 to 128; ID1: *MPF2*: -1217 to 1226, *VbrM201_1*: 1388 to 1397, *IauM201_1*: 1403 to 1412.

No binding sites, which were only conserved in Physalinae or Physaleae species, which show heterotopic expression of *MPF2*-like genes in sepals, but were absent in other analyzed species, could be detected.

However, the Mulan analysis of the first introns also gave a very interesting result: the intronic sequences from *I. australe* and *D. fasciculata* shared an approximately 400 bp

long conserved region, which was completely absent in both intronic sequences from *V. breviflora* and only little homology to sequences from other species was found (see Fig. 3.11.14). Therefore this region appeared to be a strong candidate for a regulatory regions suppressing *MPF2*-like gene expression in sepals of *I. australe* and *D. fasciculata.*

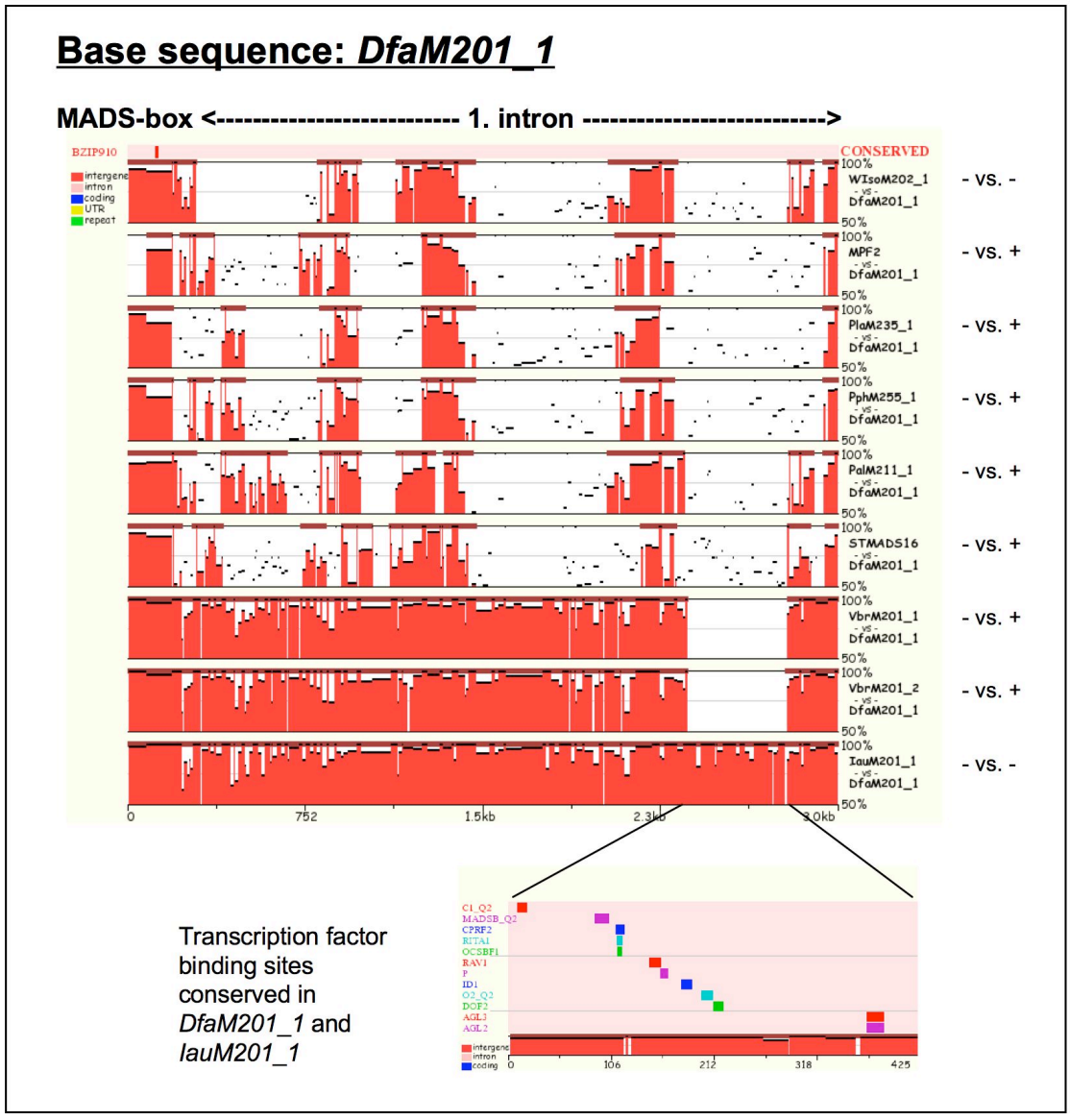


Figure 3.11.14: Conserved regions and transcription factor binding sites in the MADS-box and the first intron of *MPF2*-like genes

The conservation profiles between the first intron from *DfaM201_1* and nine other first introns from *MPF2*-like genes are graphically visualized. The default settings (ECR length = at least 100 bases, ECR similarity = at least 70%, Bottom cut-off = 50%, ECR present in = 1 species) were used. The lowest line indicates 50% and the upper one 100% conservation. A blue box indicates an overall conserved region. For accession numbers see Tables 3.7.1 and 8.1.3. The native gene expression pattern in sepals is indicated by + (= expressed) and – (= not expressed).

The program multiTF identified several conserved transcription factor binding sites in this region. Seven of these were located on the + strand: C1 (ccAACtatcct), MADSB (taATATAAATGTATa), CPRF2 (aaCACGT(A/G)t(g/t)), RITA1 (acACGT(g/a)), OCSBF1 (CACGT), RAV1 (agCACCTGatac), ID1 (TTTCTCGTTAT) and AGL3 (ttatCTAAATAAGGgtaa). The sequences of these motifs are shown in Fig. 3.11.15.

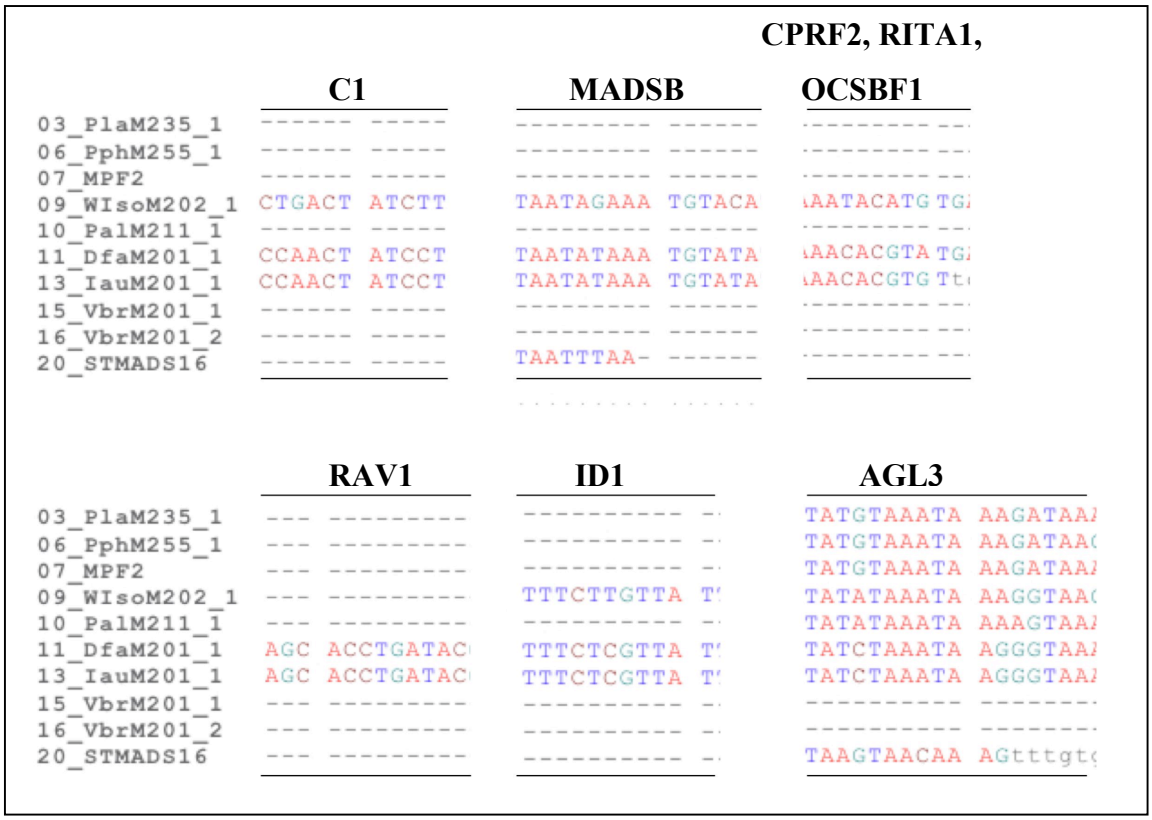


Fig. 3.11.15: Divergent transcription factor binding sites in Iochrominae

The graph shows the section of the alignment, where the transcription factor binding sites, which are only conserved in *D. fasciculata* and *I. australe*, are located. The positions of the motifs are indicated in the alignment by black lines. Motif positions in the first intron of *IauM201_1* are given: C1, position in *IauM201_1*: 2465 to 2475; MADSB, position in *IauM201_1*: 2546 to 2560; CPRF2 (including RITA1 and OCSBF1), position in *IauM201_1*: 2567 to 2576; RAV1, position in *IauM201_1*: 2598 to 2609; ID1, position in *IauM201_1*: 2631 to 2641; AGL3, position in *IauM201_1*: 2819 to 2836

PartII: Functional characterization of *MPF2*-like genes

3.12 *MPF2*-like genes are expressed in pollen of several Solanaceae

The expression of *MPF2*-like genes in pollen from different Physaleae and *S. tuberosum* was tested (see Fig. 3.12.1), because in *P. floridana MPF2* was suspected to play a role in male fertility, since RNAi knockdown plants from this species produced only little or no pollen at all, and they were sterile, unless when they were pollinated with wt pollen (He and Saedler 2005). Therefore the question arose, whether *MPF2*-like genes from other Physaleae could also play a role in male fertility in these species.

MPF2-like transcripts were visible after 35 PCR cycles in mature pollen of *P. floridana*, as expected, *W. solanacea, V. breviflora, I. australe* and surprisingly also in pollen of *S. tuberosum.* In mature pollen from *P. alkekengi* and *D. fasciculata* this was not the case.

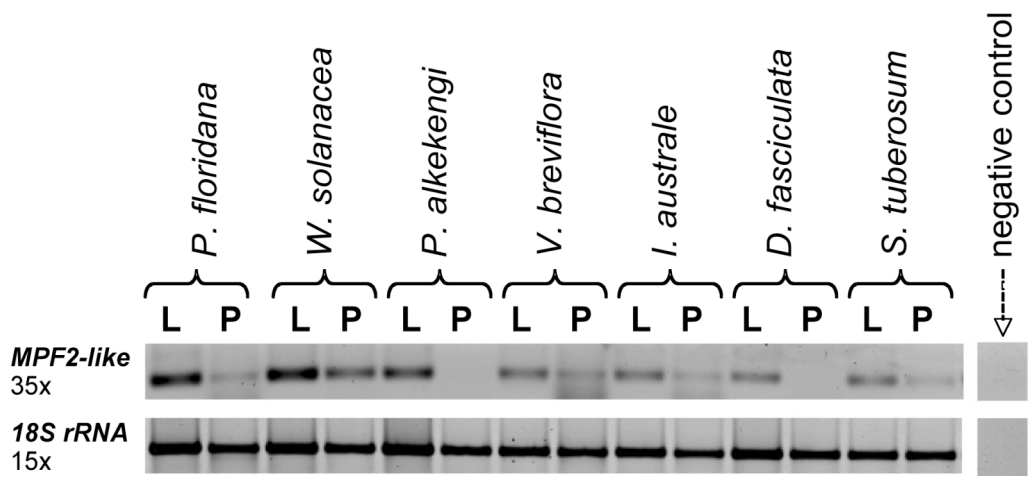


Figure 3.12.1: Comparison of *MPF2*-like gene expression in leaf and mature pollen

MPF2-like gene expression was tested in leaf (L) and mature pollen (P) from different Physaleae species and *S. tuberosum* by RT-PCR. As a control, *18S rRNA* was amplified.

3.13 Expression of *MPF2*-like genes in different tissues of Iochrominae

In this study the expression of *MPF2*-like genes in different floral and vegetative tissues of two Iochrominae species was tested (see Fig. 3.13.1), as the knowledge about spatial and temporal expression in these plants was limited to whole leaves and mixed sepal tissues of different floral stages.

Hu and Saedler (2007) showed that *V. breviflora* expresses an *MPF2*-like gene in sepals, while the closely related species *I. australe* does not. In the RT-PCR data presented here, *MPF2*-like transcripts were visible after 35 cycles, not only in young buds and sepals of immature and mature flowers of *V. breviflora,* but also in petals of immature and mature flowers of this species. However, no transcript could be detected in stamens and carpels. This was in contrast to previous assumptions based on expression patterns in *P. floridana* and *S. tuberosum*, where on the one hand *MPF2* was expressed in all four whorls and on the other hand *STMADS16* nowhere. Another result was that *I. australe* did not express the *MPF2*-like gene in most floral tissues, but surprisingly in stamens of mature flowers. In addition, both species showed expression in various vegetative tissues (leaf margin, leaf center, stem with leaf axils).

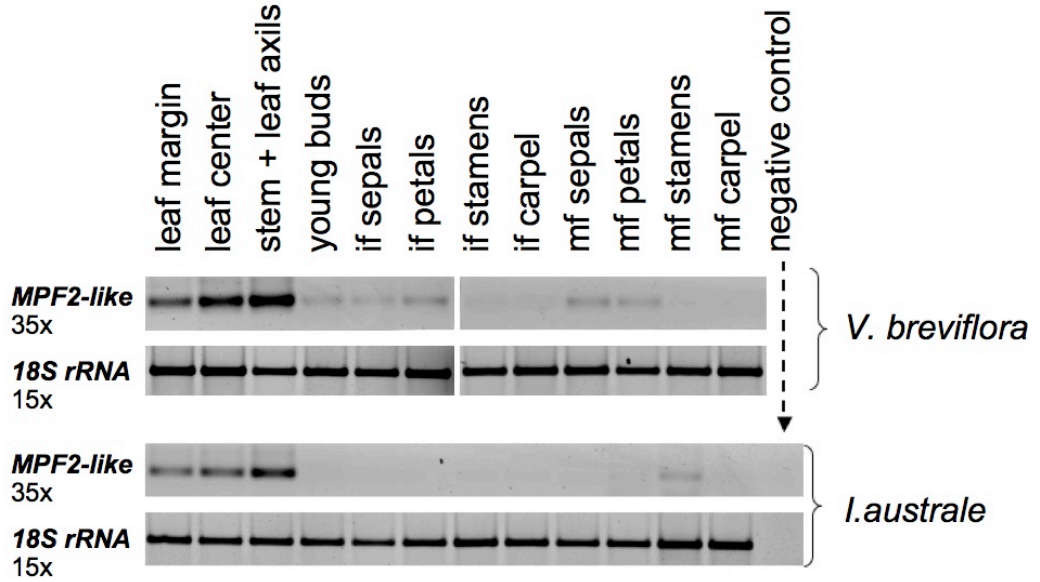


Figure 3.13.1: Comparison of *MPF2*-like gene expression in different tissues

RT-PCR data were obtained for different tissues and different stages (if = immature flower, mf = mature flower). As a control, *18S rRNA* was amplified.

3.14 *MPF2*-like upstream sequences drive strong GUS expression in pollen

MPF2-like upstream and intronic sequences had been isolated and analyzed with bioinformatics tools in an attempt to predict regions and motifs, which might play a role in transcriptional regulation of these genes. Finally the promoter activity of these regions was functionally tested by using 5'deletion promoter::*GUS* and promoter-intron::*GUS* fusion constructs in transgenic *A. thaliana.* The results are summarized in Table 3.14.2.

I. australe

Short *MPF2*-like upstream sequences from *I. australe* drove GUS expression in all four floral whorls (see Fig. 3.14.1b), while longer upstream sequences seemed to abolish expression in petals and in one allele also stigma expression. Nevertheless, the allelic variation present in the upstream sequences only led to mild variation in GUS staining patterns. Addition of the MADS-box and the first intron further restricted floral expression to solely pollen expression and up-regulated GUS expression in rosette leaf veins (see Fig. 3.14.1a). The resulting GUS pattern strongly resembled the native expression of *MPF2*-like gene transcripts in *I. australe*.

D. fasciculata

MPF2-like upstream sequences from *D. fasciculata* mainly led to GUS expression in sepals and in pollen. The sequence *pDfaM201_2* was also able to drive GUS expression in petals and carpels. Extended promoters drove expression in fewer whorls than the short versions. Expression in floral organs seemed to be repressed in the constructs containing the first intron. Besides the lacking leaf expression, the promoter-intron::*GUS* fusion constructs showed the expected pattern. However, plants containing the *pDfaM201_1*_ex-I::*GUS* construct (type 4) showed no GUS staining in any tissues. (Note: In the latter, the second exon was fused in frame to the *uidA* gene, in the same way like in *pDfaM201_1*_sh-I::*GUS*, the type 3 constructs, which drove GUS expression. This was confirmed by sequencing. However, these results should be further confirmed by RT-PCR.)

V. breviflora

Short *MPF2*-like upstream sequences from *V. breviflora* drove GUS expression in sepals, petals and pollen - and in one allele also in stigmas. Longer upstream sequences led to blue staining in sepals and pollen - and dependent on the allele also in stigmas or petals (like the native expression). Addition of the MADS-box and the first intron had a dramatic effect on the GUS pattern, changing it to expression in leaf axil buds, receptacle, pedicels, silique abscission zone and in some constructs also in sepals. So, the GUS expression pattern achieved only with upstream sequences was far more similar to the native expression pattern - except for the absent leaf expression.

All

The short promoter::*GUS* fusion constructs (type 1) mostly drove strong GUS expression in floral organs, regardless of the native expression pattern: all constructs showed intense GUS activity in pollen (Fig. 3.14.1c+d). Upstream sequences from all three Iochrominae species drove GUS expression in sepals, while those from *P. alkekengi* and *W. coccoloboides* did not. Further, expression was also found in petals, stigmas and leaf margin teeth for several of the short constructs.

Expression in transgenic plants containing extended promoter::*GUS* fusion constructs (type 2) was again mostly restricted to floral organs: Most constructs drove GUS expression in pollen and several also in sepals and stigmas.

By contrast, GUS activity patterns in transgenic plants seemed to be dramatically changed when the MADS-box and the first intron were added between the upstream sequence and the *uidA* gene: In lines containing the short promoter-intron::*GUS* fusion constructs (type 3) blue staining was mainly observed in the receptacle, the silique abscission zone and sometimes in pedicels, but only few constructs led to GUS expression in other floral organs like sepals, petals, stamens and carpels. On the other hand, in lines containing the extended promoter-intron::*GUS* fusion constructs (type 4) expression was predominantly observed in leaf axil buds and silique abscission zones.

In summary, only few constructs drove leaf expression, although RT-PCR showed that *MPF*2-like genes were expressed in this tissue in the native plants. While many of the promoter::*GUS* fusion constructs (type 1 and 2) showed GUS activity in pollen, sepals and stigmas, the constructs containing the first intron (type 3 and 4) showed a completely different GUS pattern. Yet, the constructs *pIauM201_1*_sh-I::*GUS,*

p*IauM201_1*_ex-I::*GUS* and *pWIsoM202_1*_ex::*GUS* produced GUS patterns that were very similar to *MPF2*-like expression patterns in the native plants.

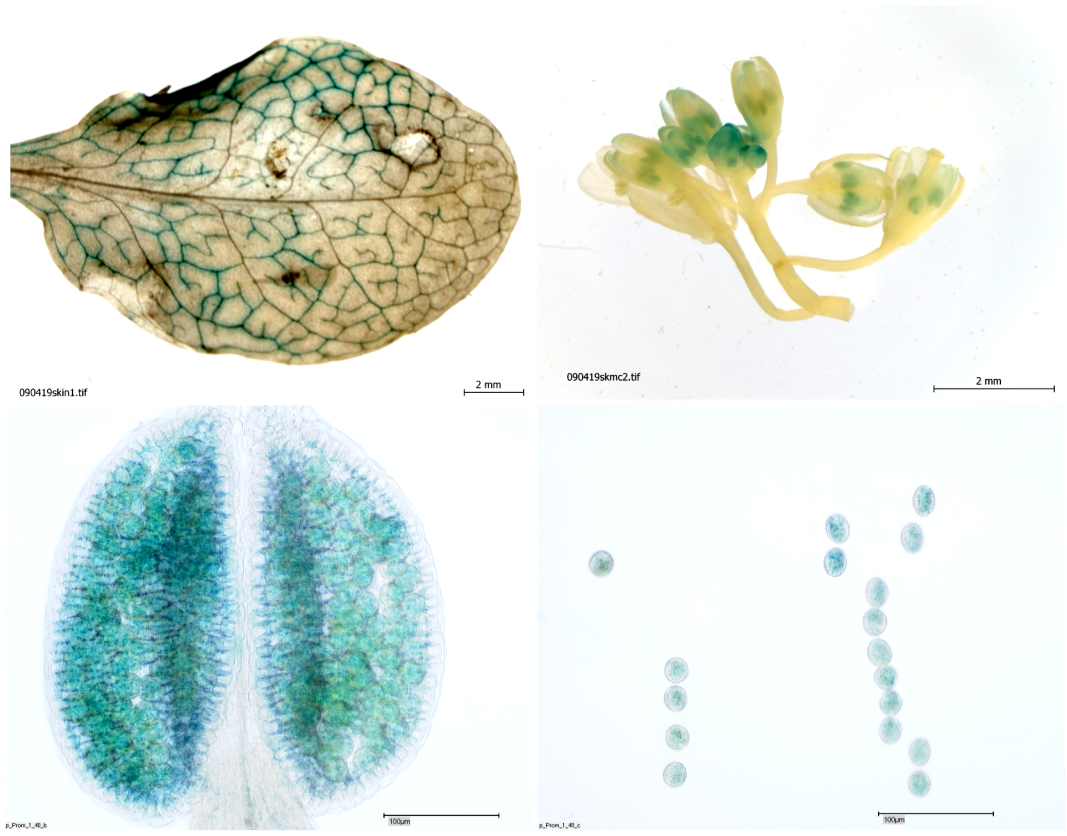


Figure 3.14.1: GUS expression patterns in transgenic *A. thaliana* containing putative regulatory sequences from *I. australe*

a) leaf veins (*pIauM201_1*_sh-I::*GUS*), scale bar = 2mm

b) flowers (*pIauM201_1*_sh::*GUS*), scale bar = 2mm

c) anther (*pIauM201_2*_ex::*GUS*), scale bar = 100 µm

d) pollen grains (*pIauM201_2*_ex::*GUS*), scale bar = 100 µm

Table 3.14.2: GUS expression patterns in transgenic *A. thaliana* containing putative regulatory sequences from Physaleae

Sequence from gene			Construct type	GUS+ transgenic lines	Leaf		Rosette leaf center, e.g. veins	Rosette leaf margin teeth	Cauline leaf	Cauline leaf tip	Stem	Leaf axil buds	Receptacle				Sepals						Petals						Stamen: Filament	Stamen: anthers/pollen						Carpel						Pedicels						Silique abscission zone			
					T								T	B	IF	MF	T	B	IF	MF	T	B	IF	MF	T	B	IF	MF	T	T	B	IF	MF	T	B	IF	MF	T	B	IF	MF	T	B	IF	MF
		***I. australe* expression**				+	+	+						(-)			-	(-)	-	-	-	(-)	-	-	-	(-)	-	-	-	+	+	(-)	-	-	-	(-)	-	-	-	(-)	-	-	-	(-)	-	-	-
IauM201_2	1		1	23/30 77%	5		2	3	2	0	1	0	1	1	1	1	23	23	18	6	13	6	12	8	1	0	9	6	10	9	20	16	15	6	2	1	0	0	0	0	0	0	1	3
IauM201_2	2		2	18/33 55%	0		0	0	0	0	1	1	0	0	0	0	17	17	5	0	0	0	4	0	0	0	9	4	0	8	8	3	8	3	0	0	0	0	0	0	0	0	0	0
IauM201_1	1		1	26/30 87%	21		2	21	0	13	0	0	1	0	0	0	20	19	12	0	4	0	10	4	1	0	20	10	4	1	1	5	1	5	0	0	0	0	0	0	0	0	0	0
IauM201_1	2		2	29/32 91%	24		0	15	0	18	0	11	0	0	0	0	7	7	1	0	0	0	12	0	0	0	24	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
IauM201_1	3		3	30/32 94%	25		25	0	0	0	0	15	13	0	0	0	1	0	0	0	0	0	5	1	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	16
IauM201_1	4		4	15/34 44%	7		7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
***D. fasciculata* expression**					+									(-)	(-)	(-)	-	(-)	(-)	(-)																											
DjfaM201_2	1		1	25/30 83%	12		1	12	1	3	0	2	0	0	0	0	22	22	18	5	11	0	0	0	0	0	24	3	0	1	0	9	15	10	1	1	0	0	0	0	1	1	0	7
DjfaM201_2	2		2	23/31 74%	1		0	1	0	1	0	0	0	0	0	0	18	18	5	0	1	0	1	0	0	0	19	0	0	0	0	9	9	9	5	0	0	0	0	0	0	0	0	0
DjfaM201_1	1		1	19/35 54%	1		0	0	1	0	0	1	0	0	0	0	15	15	0	0	0	0	0	0	0	0	9	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
DjfaM201_1	2		2	18/37 49%	0		0	0	0	0	0	1	0	0	0	0	2	2	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DjfaM201_1	3		3	7/40 18%	2		1	0	1	0	0	1	0	0	0	0	1	0	0	0	3	0	0	3	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0	1
DjfaM201_1	4		4	0/27 0%	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
***V. breviflora* expression**					+		+	+									+												+	(-)				-											
VbrM201_1	1		1	28/30 93%	18		4	18	1	11	0	2	3	0	1	3	28	27	28	15	14	0	10	10	2	0	20	10	2	0	0	5	1	5	2	0	0	0	0	0	1	1	0	2
VbrM201_1	2		2	16/31 52%	1		0	1	0	0	0	0	0	0	0	0	15	15	3	1	2	0	0	0	0	0	5	2	0	0	0	3	6	6	2	0	0	0	0	0	0	0	0	0
VbrM201_1	3		3	25/30 83%	0		0	0	0	0	0	19	18	1	13	18	11	11	3	0	0	0	0	0	0	0	4	4	0	1	0	23	17	3	2	0	0	0	0	0	17	20	0	0
VbrM201_1	4		4	24/35 69%	0		0	0	0	0	0	11	10	6	4	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	21	0	0
VbrM201_2	1		1	33/35 94%	3		3	0	1	1	0	0	1	0	1	1	16	16	2	0	17	1	16	14	0	0	33	11	3	3	2	23	3	3	2	1	0	0	0	0	0	0	0	0
VbrM201_2	2		2	10/33 30%	2		0	0	0	2	0	0	0	0	0	0	4	4	1	0	4	1	4	0	0	0	6	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
VbrM201_2	3		3	25/30 83%	0		0	0	0	0	0	25	19	6	1	18	6	6	1	0	2	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	14	18	0	0
VbrM201_2	4		4	24/33 73%	0		0	0	0	0	0	18	2	0	1	1	7	7	2	0	0	0	4	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	6	7	0	0

Sequence from gene	Construct type	GUS+ transgenic lines	Leaf	Rosette leaf veins	Rosette leaf margin teeth	Cauline leaf	Cauline leaf tip	Stem	Leaf axil buds	Receptacle				Sepals				Petals				Filament	Anthers/Pollen				Stigma				Pedicels	Silique abscission zone
			T							T	B	IF	MF	T	B	IF	MF	T	B	IF	MF		T	B	IF	MF	T	B	IF	MF		
***P. alkekengi* expression**			+											+												-						
PalM211_1	1	18/31 58%	2	2	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	18	13	16	11	0	0	0	0	1	1
PalM211_1	2	3/23 13%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	1	0
PalM211_1	3	27/30 90%	0	0	0	0	0	(9) cut	17	22	19	21	22	1	1	1	0	0	0	0	0	1	0	0	0	0	20	0	15	20	26	23
PalM211_1	4	25/33 76%	0	0	0	0	0	(14) cut	12	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	22	20
***W. solanacea* expression**			+											+												+						
WIsoM202_1	2	13/19 68%	6	6	1	1	2	(7) cut	11	5	0	2	5	4	4	4	1	4	0	3	3	0	5	5	0	0	6	2	5	6	9	12
WIsoM202_1	4	17/30 57%	1	1	0	1	0	1	17	8	0	8	7	7	5	7	6	2	0	2	2	0	0	0	0	0	0	0	0	0	6	10
***P. lanceifolia* expression**			+											+																		
PlaM235_1	2	31/32 97%	14	11	3	8	1	0	0	0	0	0	0	2	0	2	1	1	0	0	1	0	30	11	30	26	3	1	2	2	0	1
PlaM235_1	3	9/30 30%	0	0	0	0	0	(8) cut	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
***W. coccoloboides* expression**			+											+																		
WIcoM201_1	1	27/30 90%	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	16	0	16	13	1	26	14	26	23	3	0	2	3	0	2

	GUS+/transgenic lines
	= 100% and ≥ 3 independent lines
	≥ 70% and ≥ 3 independent lines
	≥ 40% and ≥ 3 independent lines
	≥ 10% and ≥ 3 independent lines

Table 3.14.2 shows the observed GUS expression patterns in different tissues of transgenic *A. thaliana*: GUS+/ transgen = percentage of the transgenic plants, which show GUS staining;

expression: + or - = presence or absence of transcript in this tissue in the native species;

T= total, B = Bud, iF = immature flower, mF = mature flower

3.15 Optimization of the method for transient expression in floral tissues

To have a more reliable and rapid tool for functional analysis of putative regulatory sequences of *MPF2*-like genes in the future, a protocol for a transient assay was established, which should allow testing the constructs in floral tissues of the native plants. Different Physaleae and Solanaceae were transiently infected with *A. tumefaciens* strains GV3101 and LBA4404 harboring p27IGUS (Shang *et al.* 2007), a vector containing an *uidA* gene with an intron (IGUS). This vector was provided by Yongjin Shang and prevents *Agrobacterium*-derived GUS expression. The protocol of Shang et al. (2007) was optimized for our species and the modified protocol (see chapter 2.22), based on infiltration and co-cultivation, allows infection of *S. tuberosum* and *P. alkekengi* leaves and floral tissues with *A. tumefaciens* strain GV3101. Typical results are shown in Fig. 3.15.1.

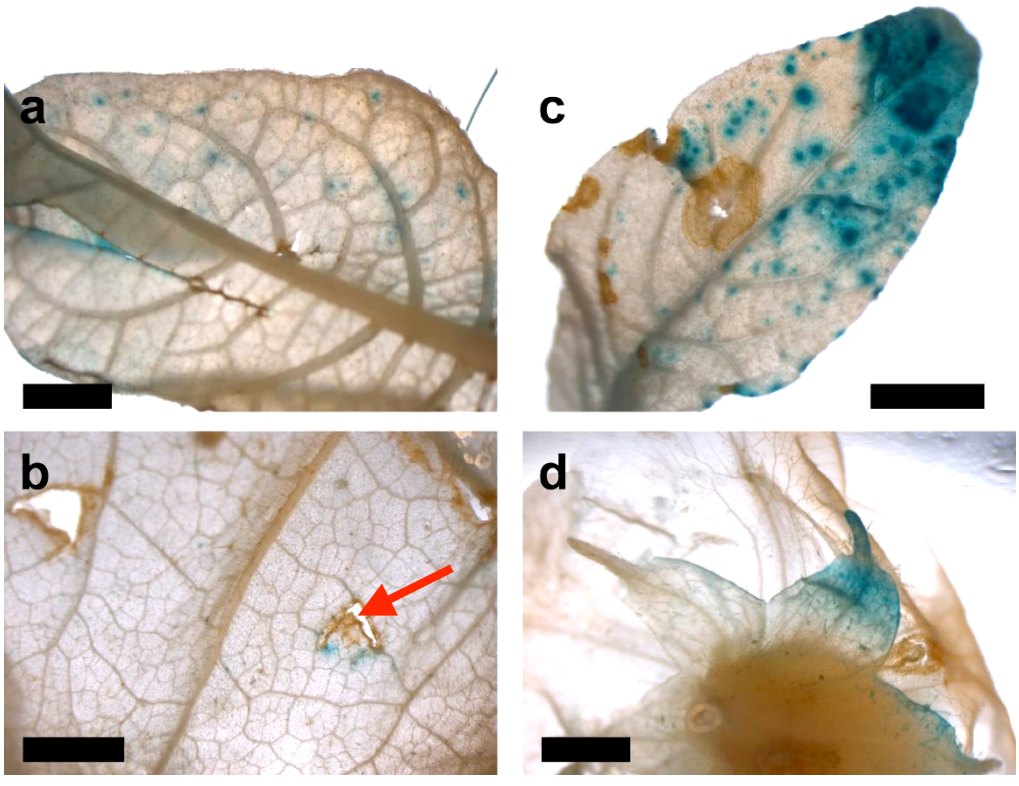


Figure 3.15.1: Transient GUS expression in leaves and calyx tissue from *P. alkekengi* and *S. tuberosum*

P. alkekengi leaf (a) and lantern (b), *S. tuberosum* leaf (c) and flower (d),

Scale bars = 2mm

4. Discussion

4.1 *MPF2*-like genes and the absence of ICS in Iochrominae

In this study 18 cDNA sequences of *MPF2*-like genes from 12 different Physaleae and one *Capsicum* species were isolated or extended. In several species more than one *MPF2*-like cDNA was found. However, these sequences always showed a high degree of sequence identity (96 to 100% sequence identity at the protein level) in contrast to the divergent *Withania MPF2*-like-A and B genes (84% sequence identity at the protein level; Khan 2009, Khan *et al.* 2009). Since most species in this study were wild plants a certain amount of heterozygosity could be expected and thus the different sequences found were classified as putative alleles, not as different genes. This assumption was also supported by phylogenetic reconstruction (see chapter 3.3). Other possible explanations for the observed sequence variation are recent duplication events or PCR-derived mutations. In contrast to species from the genus *Withania*, the Iochrominae and *Tubocapsicum* contained only one class of *MPF2*-like genes (see chapter 3.4). Based on the current sequence data this is probably also the case in the Physalinae and *W. coccoloboides*, but this has still to be confirmed by Southern Blot analysis.

In general, the translated MPF2-like protein sequences from the different Physaleae species in this study seemed to be very conserved. However, the MPF2-like protein sequences from Physalinae were longer than the ones from Iochrominae, *W. coccoloboides* and *C. baccatum*. This was due to an insertion of one amino acid in the C-terminus. The amino acid sequences of different groups, e.g. from the American *Physalis* species, contained certain characteristics, but no changes could be clearly correlated with the presence or absence of ICS in these species.

Based on phylogenetic reconstructions the *MPF2*-like cDNAs from Physaleae were subdivided into four to five sub-groups: the *MPF2*-like Iochrominae genes, the *MPF2*-like *W. coccoloboides* genes, the *MPF2*-like-B genes and a group containing the *MPF2*-like-A and the *MPF2*-like Physalinae genes. According to their sequence identities the *MPF2*-like Iochrominae genes and the *MPF2*-like *W. coccoloboides* genes were more similar to *MPF2*-like-B genes than to *MPF2*-like-A or *MPF2*-like Physalinae genes. This higher structural homology might reflect more similar properties of the proteins. Thus it could be hypothesized that *MPF2*-like proteins from

Iochrominae and *W. coccoloboides* might not be able to trigger ICS formation, since overexpression of *MPF2*-like-B genes did in contrast to *WSA206* and *MPF2* not promote sepal growth in *A.* thaliana (He and Saedler 2007, Khan 2009, Khan *et al.* 2009). This might explain the absence of this morphological structure in *V. breviflora* and *W. coccoloboides*, which both lack ICS, despite expression of *MPF2*-like genes in their calyces. Overexpression of *MPF2*-like genes from Iochrominae species in *A. thaliana* or *S. tuberosum* might give further clues about the functional properties of this subgroup.

Interestingly, it has been found recently that different groups of *MPF2*-like genes in Physaleae might have experienced different types of selection during their evolution (Khan *et al.* 2009). It appears that the branch leading to the *Withania MPF2*-like-A genes and *Physalis MPF2*-like genes has experienced positive Darwinian selection (highlighted in bold, see Fig. 4.1.1). By contrast, all other branches, including the ones leading to *MPF2*-like-B and the Iochrominae genes, seem to have evolved under purifying selection. These findings correlate with the presence of ICS in *Physalis* and *Withania* and its absence in Iochrominae.

Accordingly, in the course of the evolution of the ICS the progenitor of the MPF2-like–A proteins from *Withania* and the MPF2-like proteins from *Physalis* seems to have experienced positive Darwinian selection before the slit of the two genera. Afterwards purifying selection obviously took over, conserving the proteins and thereby also the newly gained morphological structure ICS. Currently several hypotheses about the adaptive value of the lantern are discussed, concerning the kind of selection pressure that might have led to the emergence or maintenance of ICS. One possible explanation is that the lantern produces an improved microclimate around the developing berry, which provides a higher humidity in dry habitats. This hypothesis is in accordance with the habitat preferences of the studied species: It has been noted that Iochroma species not showing ICS occur in scrub or cloud forest between 2200 and 2900 m in the Andes and that their flowering peaks during the rainy season (Smith *et al.* 2008). Further, *Tubocapsicum*, which as well lacks ICS, has been also found in relatively humid environments like evergreen wood and along streams in Asia (Merril 1923, Hunziker 2001). By contrast, *Withania* species, which enclose their berries, such as *W. somnifera* and *W. coagulans* favor dryer habitats (Hepper 1991). Also several *Physalis* species seem to prefer warm sunny habitats e.g. *P. peruviana* “needs full sun

… very little [rain] when the fruits are maturing" (Morton 1987), though the old world species *P. alkekengi* can even endure frost in winter.

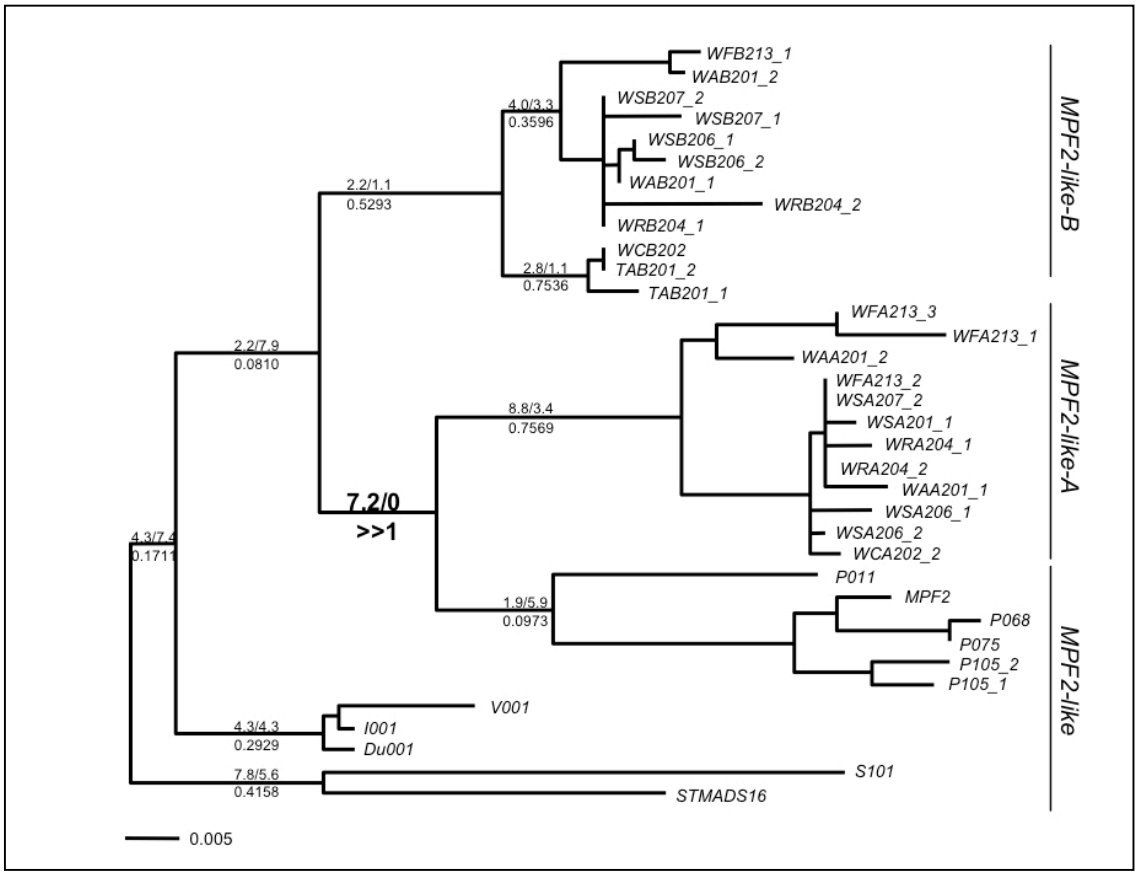


Figure 4.1.1: Darwinian Selection of *MPF2*-like proteins in Physaleae (adopted from Khan *et al.* 2009)

"Numbers above the major branches of the unrooted ML tree are averages of non-synonymous and synonymous changes, while ω = dN/dS values are given below the branches. Numbers highlighted in bold indicate the branch under positive Darwinian selection. Bootstrap and posterior probability values of MP, ML and Bayesian inference for all of the major branches were above 85%." (Khan *et al.* 2009)

4.2 Phylogenetic relationships among *MPF2*-like genes

The topologies of the phylogenetic trees based on coding and non-coding regions of *MPF2*-like genes did largely agree, accordingly no hints for promoter swapping events in the Physaleae were found (see chapter 3.3 and 3.8). Phylogenetic reconstructions using upstream and protein sequences only helped to define different subclades of

MPF2-like genes, while the phylogenetic tree of *MPF2*-like cDNAs suggested that the group containing the Iochrominae genes is basal to those from the Withaninae and Physalinae.

The main phylogenetic groups and their arrangement were in agreement with previous reports on phylogenetic relationships in Physaleae (Olmstead *et al.* 1999, Whitson and Manos 2005, Smith and Baum 2006, Hu and Saedler 2007). The observation that all American *Physalis* and *M. solanaceus* formed a subgroup within the Physalinae and that this subgroup did not include *P. alkekengi* reflects the geographical separation of this Old World species. In the phylogeny presented in this study (see Fig. 3.8.1), which was based on upstream sequences of the *MPF2*-like genes, *P. alkekengi* separates *W. solanacea* from the American *Physalis* as proposed by Olmstead *et al.* (1999) and Whitson and Manos (2005). By contrast, Hu and Saedler (2007) proposed an opposite arrangement.

The two *Witheringia* sequences did not cluster together in all three phylogenies based on *MPF2*-like sequences, even though both species have been grouped in the same genus (Sousa-Pena 2001). This finding is in agreement with the study of Hu and Saedler (2007), who suggested that the *W. coccoloboides* and *W. solanaceae* should not be classified in the same genus, based on phylogenetic reconstructions using MPF2-like protein sequences and three additional markers. Studies involving more *Witheringia* species could help to clarify the phylogeny within this genus and elucidate its position within the Physaleae.

In summary, this study showed that also non-coding sequences can be used to infer phylogenetic relationships - at least for the classification of major groups.

4.3 *MPF2*-like genes and male fertility

The expression study on different floral tissues of *I. australe* and *V. breviflora* revealed *MPF2*-like gene expression in buds, sepals, petals, and mature pollen of *V. breviflora*. By contrast, *I. australe* did only show expression in stamens of mature flowers (see Fig. 3.13.1). These findings are in accordance with previous reports (Hu and Saedler 2007), but suggest also that the spatial *MPF2*-gene expression patterns are more variable in floral organs of Physaleae than previously expected, based on observations in *P. floridana* and *S. tuberosum* (He and Saedler 2005). This variation could indicate

that the presence of *MPF2*-like genes in certain floral tissues of the Iochrominae has no useful (or even negative) effects, and was therefore lost during evolution.

Since *MPF2* seems to be important for pollen production and thus for male fertility in *P. floridana* (He and Saedler 2005), *MPF2*-like gene expression in mature pollen of other Physaleae was studied. Surprisingly, *MPF2*-like gene expression could be shown in pollen of several Physaleae and surprisingly also in pollen of *S. tuberosum*. The fact that in this study *MPF2*-like gene expression was detected in pollen of *V. breviflora* and *S. tuberosum*, but previously not in stamen of these species (He and Saedler 2005, this study) can be explained easily, since pollen comprises only a small fraction of the mass of an entire stamen. The observation that in mature pollen of *P. alkekengi* and *D. fasciculata* no *MPF2*-like genes seemed to be expressed could indicate that (i) the gene is not expressed, (ii) that expression was below the threshold of detection or (iii) that gene expression is more relevant during earlier stages of pollen maturation. However, in *A. thaliana AGL24,* the putative ortholog of *MPF2,* is only expressed during later stages of floral development in pollen and at the adaxial surface of the gynoecium (Yu *et al.* 2004).

Nevertheless, as *MPF2* seems to play an important role in male fertility in *P. floridana* (He and Saedler 2005), this could be also the case in other Physaleae and even in *S. tuberosum*. This is supported by the finding that all tested promotor::*GUS* constructs containing upstream sequences of *MPF2*-like genes from Physaleae were able to drive GUS expression in pollen of *A. thaliana* (see Fig. 3.14.2). Hence, if expression of *MPF2*-like genes in pollen is essential for male fertility in Physaleae, this trait should have been favored by selection during evolution. Expression in other floral organs could have arisen as a by-product of expression in pollen and could have been maintained due to “hitch-hiking” effects. Even the inflated calyx syndrome might have initially originated as a by-product of the selected increased male fertility, even though it might have become advantageous in other aspects later on, e.g. preventing desiccation of fruits in dry habitats. Further research is needed to better understand the role of *MPF2*-like genes in male fertility in Physaleae and other Solanaceae, e.g. by generating transgenic *MPF2*-like RNAi knockdown lines in other Physaleae species and in *S. tuberosum*.

4.4 Intragenic regions of *MPF2*-like genes

The genomic loci of *MPF2*-like genes were analyzed with the program Mulan in two *Solanum* species and two Physaleae species to detect conserved and divergent regions in their introns. The two *Solanum* species did not show *MPF2*-like gene expression in sepals (*STMADS16* from *S. tuberosum* and *SlyM2* from *S. lycopersicum*), while the Physaleae species expressed the gene in sepals (*WSA206* from *W. somnifera* and *MPF2* from *P. floridana*). The calculated conservation profiles revealed conserved as well as divergent regions in introns and the 3'UTR. The latter was only conserved between the two *Solanum* species.

The observation that the introns 1, 3 and 5 contained several overall conserved regions suggests that key *cis*-regulatory elements could be located there. Such *cis*-elements might be involved e.g. in the regulation of vegetative expression of *MPF2*-like genes.

In addition, the introns 1, 5, 7, 8, and the 3'UTR contained regions that were only conserved among species with a similar *MPF2*-like gene expression pattern in sepals. These regions might contain *cis*-regulatory elements responsible for differential expression of *MPF2*-like gene in sepals.

In conclusion, these results support the hypothesis that also intronic regions and the 3'UTR could play a role in transcriptional regulation of *MPF2*-like genes in Physaleae and other Solanaceae. This is further supported by the fact that intronic *cis*-regulatory elements have been described for several other MADS-box transcription factor genes, e.g. *PLE* (Davies *et al.* 1999), *AG* (Hong *et al.* 2003) and *AGL24* (Liu *et al.* 2007).

In Physaleae, the first intron might be the most promising candidate, as it was the largest intron and contained several conserved and divergent regions. The GUS expression patterns elicited by promoter::*GUS* and promoter-intron::*GUS* fusion constructs (see Table 3.14.2) showed that the intragenic region containing the MADS-box and the first intron had a strong regulatory effect, and thus suggest that the intron does contain key *cis*-regulatory elements. This will be discussed in more detail in chapter 4.7. However, the observation that many extended promoter-intron::*GUS* fusion constructs did not lead to the expected native-like expression patterns in transgenic *A. thaliana* also raised the question whether some *cis*-regulatory elements were missing in the tested constructs. Such elements could be located in other introns: For instance, Liu *et al.* (2007) showed that in *A. thaliana* AP1 binds to a CArG-box in the third intron of *AGL24*, which is the putative ortholog of *STMADS16* (Garcia-

Maroto *et al.* 2000), to repress its expression in young floral meristems. However, this CArG motif could not be detected in the third intron of *STMADS16* (data not shown). The divergence between native *MPF2*-like gene expression and observed GUS patterns can also be caused by the heterologous system *A. thaliana.* Future studies to discriminate between these two explanations might involve constructs in which an entire *MPF2*-like locus is fused to the *uidA* gene, or transient tests in the native species.

4.5 CArG-motifs and heterotopic expression

Upstream sequences and the first introns of *MPF2*-like genes from Physaleae and *S. tuberosum* were screened for MEF2- or N10-type CArG-boxes and other published CArG-motifs, since it had been hypothesized that the absence of N10-type CArG-boxes in the upstream sequence of *MPF2* from *P. floridana* could be causal for heterotopic expression of *MPF2* in sepals of this species (He and Saedler 2005). Three strictly defined MEF2- or N10-type CArG-boxes were found in 2.5 kb upstream sequence of *STMADS16* from *S. tuberosum*, yet none or only one motif of this type was found in *MPF2*-like upstream sequences from Physaleae. In this study no correlation between the presence and absence of these motifs (or other CArG motifs) and the occurrence of *MPF2*-like gene expression in sepals of Physaleae species was found. Thus the absence of expression in sepals in two of the Iochrominae species does not seem to be caused by MEF2- or N10-type CArG-boxes in its promoter region. However, regulatory changes underlying the absence of expression in sepals might not be the same in the two Iochrominae and *S. tuberosum*, which does not belong to the Physaleae. This hypothesis is based on the assumption that expression in sepals is probably a plesiomorphic trait in Physaleae and its absence in two of the Iochrominae may be caused by a secondary loss (Hu and Saedler 2007). Thus, it can so far not be excluded that the strict MEF2- or N10-type CArG-boxes in the upstream sequence of *STMADS16* are causal for the absence of *MPF2*-like gene expression in sepals of *S. tuberosum.*

4.6 Conserved *cis*-regulatory elements in upstream regions

GUS expression patterns

The proximal part of the upstream region of *MPF2*-like genes could drive GUS expression in pollen and also often in sepals, petals and stigmas of transgenic *A. thaliana* plants (see type 1 constructs, Table 3.14.2). Longer promoters led to similar GUS expression patterns in floral organs, although expression in petals was abolished in some cases (see type 2 constructs, Table 3.14.2). These findings suggest that the proximal part of the upstream region probably contains motifs that can up-regulate *MPF2*-like gene expression in pollen and other floral organs, while the more distal region might contain some negative regulatory elements repressing expression in certain floral organs. However, the included upstream regions were often not sufficient to create GUS patterns that mimic the native expression patterns. This could indicate that important regulatory regions are located outside of these regions or that the heterologous system *A. thaliana* is not suitable for functional analysis of *MPF2*-like promoter sequences. Functional promoter analyses using transient expression in the native species could therefore help to answer this question.

GUS expression patterns *vs.* overall conserved motifs in the upstream regions

The *in silico* promoter analysis revealed that only an approximately 350 bp long region flanking the inferred transcription initiation sites was conserved in all analyzed upstream sequences of *MPF2*-like genes. This region contained six motifs, which were conserved in all Physaleae and *S. tuberosum* (see Fig. 3.11.3, Fig. 3.11.6 and Fig. 4.8.1). These motifs included a binding site for the MADS-domain proteins AG, AGL1, AGL2, AGL3 and AGL15, a binding site for the HD-Zip I proteins ATHB1 and ATHB5 and a binding site for the MYB transcription factor ATMYB2. Hence *MPF2*-like genes might be under the control of MADS, HD-Zip I and MYB transcription factors in Solanaceae. Described functions of these transcription factors are the following: MADS-domain transcription factors are known to define floral meristem and floral organ identity (Schwarz-Sommer *et al.* 1990, Sommer *et al.* 1990, Yanofsky *et al.* 1990, Theissen *et al.* 2000), e.g. AG specifies floral meristem, carpel and stamen identity, while the SEPALLATA proteins AGL2 and AGL3 are required for the specification of floral organ identity and for floral meristem determinacy (Pelaz *et al.* 2000, Honma and Goto 2001, Pelaz *et al.* 2001a, Pelaz *et al.* 2001b, Ditta *et al.*

2004). AGL15 may play a role during post-germinative development. Hence, the observed GUS expression in floral tissues of transgenic *Arabidopsis* could be caused by a MADS-domain protein binding to the conserved CArG-box and up-regulating expression in floral tissues. The MYB transcription factor ATMYB2 is known to regulate the expression of salt- and dehydration-responsive genes in *A. thaliana* (Urao *et al.* 1993, Abe *et al.* 2003), so maybe these stresses can influence *MPF2*-like gene expression. ATHB5 is a positive regulator of ABA-responsiveness (Johannesson *et al.* 2003) and ATHB1 is involved in leaf development (Aoyama *et al.* 1995). Also *MPF2*-like genes are expressed in leaves (He and Saedler 2005, Hu and Saedler 2007) and play a role in leaf development (He and Saedler 2005). Therefore an ATHB1-like protein could bind to the ATHB binding site in the promoter and up-regulate their expression in leaves. Unfortunately, this was not supported by GUS expression patterns in transgenic *A. thaliana*, as most promoters did not drive *GUS* expression in leaves.

Bioinformatic analyses suggested that the core promoter might comprise only the conserved region flanking the transcription start. However this was not supported by the GUS expression patterns in *A. thaliana*, because native expression was not recapitulated, e.g. leaf expression was mostly absent.

4.7 Conserved *cis*-regulatory elements in intragenic regions

GUS expression patterns

GUS expression patterns in transgenic plants seemed to be dramatically changed when the MADS-box and the first intron were added between the upstream sequence and the *uidA* gene (see type 3 and 4 constructs, Table 3.14.2). This finding suggests that key *cis*-regulatory elements might be located in this intragenic region.

GUS expression patterns *vs.* conserved motifs in the MADS-box and the first intron

The programs DIALIGN and multiTF helped to identify nine overall conserved motifs: In the MADS-box a binding site for bZIP910, a basic leucine zipper (bZIP) transcription factor, was detected and the intron contained eight overall conserved motifs, including a binding site for INDETERMINATE1 (ID1). These findings suggested that expression of *MPF2*-like genes in Solanaceae might be controlled by bZIP and ID1-like transcription factors. *bZIP910* is expressed in flowers of *A. majus*,

but its function remains still speculative (Martinez-Garcia *et al.* 1998). The *ID1* gene encodes a zinc finger protein, which is expressed in immature leaves and controls the transition to flowering in maize (Colasanti *et al.* 1998, Colasanti *et al.* 2006). Whether these transcription factors can bind to the detected putative *cis*-regulatory elements and do control expression of *MPF2*-like genes *in planta* has still to be confirmed, but they are candidates for putative *trans*-regulators of *MPF2*-like genes in Solanaceae.

4.8 Loss of *MPF2*-like gene expression in *I. australe* and *D. fasciculata*

One of the major goals of this thesis was to elucidate the role of *cis*-regulatory elements in differential transcriptional regulation of *MPF2*-like MADS-box transcription factors in Physaleae. The Iochrominae species *V. breviflora* and *I. australe*, which differ in expression of *MPF2*-like genes in sepals, seemed to be a suitable system, where the changes underlying a secondary loss of *MPF2*-like gene expression in sepals could be studied.

GUS expression patterns

In transgenic *A. thaliana* plants short promoter::*GUS* fusion constructs (type 1) containing *MPF2*-like upstream sequences from *V. breviflora* as well as from *I. australe* could drive expression in all four floral whorls. Also the extended promoter::*GUS* fusion constructs (type 2) from both species led to GUS expression in sepals and pollen, but some alleles abolished GUS expression in petals and carpels. Addition of the MADS-box and the first intron of *I. australe* further restricted floral expression to solely pollen expression and also up-regulated GUS expression in rosette leaf veins. The resulting GUS pattern strongly resembled the native *MPF2*-like gene expression in *I. australe*. Unfortunately, the addition of the MADS-box and the first intron of *V. breviflora* produced paradoxical *GUS* expression patterns, although some of these constructs still drove GUS expression in sepals.

These results suggest that the proximal upstream regions contain elements up-regulating floral expression in all floral whorls and that in some alleles the distal upstream regions contain elements, which can down-regulate expression in petals and carpels. Further, the intragenic region (MADS-box and first intron) of *IauM201_1* from *I. australe* probably has a strong regulatory effect and can down-regulate floral expression in sepals and up-regulate expression in rosette leaf veins. This effect was

probably caused by the intron and not by the MADS-box of *IauM201_1*, since the latter region is highly conserved between different Iochrominae species. Therefore the first intron of *IauM201_1* might contain *cis*-elements, which can repress expression in sepals and which are responsible for the loss of *MPF2*-like gene expression in this species.

GUS expression patterns vs. divergent motifs

The bioinformatic analyses revealed several motifs, which were either only conserved in upstream sequences of *MPF2*-like genes from *V. breviflora* or in upstream sequences of *MPF2*-like genes from *I. australe* and *D. fasciculata* (see Fig. 4.8.1). However, as upstream sequences from all Iochrominae generated similar GUS expression patterns in floral organs of transgenic *A. thaliana*, these motifs did not seem to be involved in differential *MPF2*-like gene expression in sepals of Iochrominae.

Further, the first introns of *MPF2*-like genes from *I. australe* and *D. fasciculata* contained eight conserved motifs, which were all absent in *V. breviflora* and other Physaleae. Since the first intron of an *MPF2*-like gene from *I. australe* was shown to repress GUS expression in sepals (and to up-regulate leaf expression) in transgenic *A. thaliana*, these motifs were hot candidates for *cis*-regulatory elements underlying the absence of *MPF2*-like gene expression in *I. australe* and *D. fasciculata.* Among the conserved motifs were: a binding site for the MADS-domain protein MADSB, a *Sinapis* homologue of AGAMOUS-LIKE 8 (AGL8) / FRUITFULL (FUL) (Menzel *et al.* 1996); a binding site for the MADS-domain transcription factor AGL3; a binding site for the common plant regulatory factor 2 (CPRF2), a light-regulated bZIP-transcription factor from parsley (*Petroselinum crispum* L.); a binding site for ID1 and a binding site for RAV1, an AP2/B3 domain transcription factor. Maybe one of these transcription factors can bind to the first introns of *MPF2*-like genes in *I. australe* and *D. fasciculata* and can thereby repress their expression in sepals.

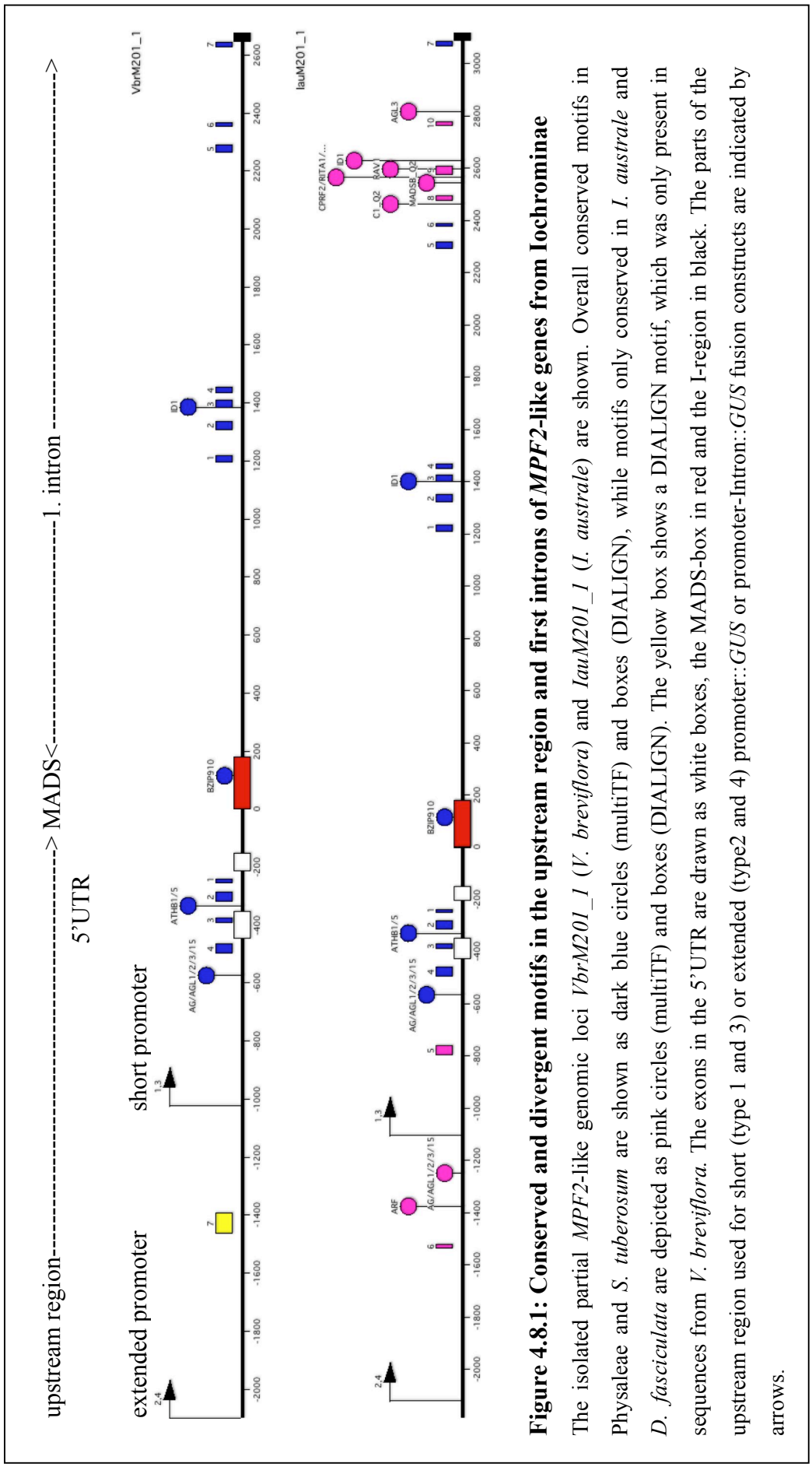


Figure 4.8.1: Conserved and divergent motifs in the upstream region and first introns of *MPF2*-like genes from Iochrominae

The isolated partial *MPF2*-like genomic loci *VbrM201_1* (*V. breviflora*) and *IauM201_1* (*I. australe*) are shown. Overall conserved motifs in Physaleae and *S. tuberosum* are shown as dark blue circles (multiTF) and boxes (DIALIGN), while motifs only conserved in *I. australe* and *D. fasciculata* are depicted as pink circles (multiTF) and boxes (DIALIGN). The yellow box shows a DIALIGN motif, which was only present in sequences from *V. breviflora.* The exons in the 5'UTR are drawn as white boxes, the MADS-box in red and the I-region in black. The parts of the upstream region used for short (type 1 and 3) or extended (type2 and 4) promoter::*GUS* or promoter-Intron::*GUS* fusion constructs are indicated by arrows.

4.9 Outlook

In this study key *cis*-regulatory regions of *MPF2*-like MADS-box genes have been identified. A detailed overview about overall conserved and divergent motifs in these regions was given. These predictions were based on a considerable large sequence dataset of wild and cultivated Solanaceae, which should allow drawing general conclusions for the tribe Physaleae and maybe even for other Solanaceae. The next step should be the functional analysis of the identified motifs (i) to confirm, which of the detected motifs are functional *cis*-elements, (ii) to elucidate their individual functions and (iii) to identify those motifs, which underlie differential transcriptional regulation of *MPF2*-like gene expression. This could be done using the optimized protocol for transient expression in floral tissues of Solanaceae (see chapter 3.15) and promoter::I*GUS* and promoter-intron::I*GUS* fusion constructs, in which individual motifs are mutated. Finally, this should further improve knowledge on the role of *cis*-regulatory elements in transcriptional regulation of *MPF2*-like genes.

5. Abstract

Species with large inflated lantern-like calyces occur in several genera of the family Solanaceae. In *Physalis*, this morphological novelty has been termed “Chinese lantern“ or “Inflated-Calyx-Syndrome“ (ICS). It has been shown that in *Physalis floridana* the MADS-box transcription factor *MPF2* is expressed in floral organs, while expression of its ortholog *STMADS16* from *Solanum tuberosum* is solely restricted to vegetative tissues. Expression of *MPF2* in floral organs is not only important for formation of the ICS, but also plays a role in male fertility in *P. floridana*. However, whether the recruitment of this transcription factor into a new context was achieved by changes in *cis*-regulatory elements or *trans*-acting factors is not known. Expression of *MPF2*-like genes in sepals seems to be a plesiomorphic trait in the Physaleae, but in several Iochrominae like *Iochroma australe* and *Dunalia fasciculata* secondary losses occurred. This was apparently not the case in its close relative *Vassobia breviflora*.

This project was initiated to provide insights concerning the role of *cis*-regulatory elements in differential transcriptional regulation of *MPF2*-like MADS-box transcription factors in Physaleae with an emphasis on the Iochrominae. *MPF2*-like genes and their upstream and first intron sequences were isolated from several species. Bioinformatic analyses indicated that a short region upstream of the MADS-domain and regions in the first intron were conserved in all studied species and contained several known transcription factor binding sites. Further, *I. australe* and *D. fasciculata* shared additional conserved motifs, which were absent in other Physaleae. Functional analysis using promoter::*GUS* and promoter-intron::*GUS* fusion constructs in transgenic *A. thaliana* suggested that the conserved region upstream of the MADS-box might contain *cis*-elements, which can up-regulate floral expression of *MPF2*-like genes, while the first intron of *I. australe* can repress expression in sepals. Thus *cis*-elements in this region might be responsible for the loss of expression in this species.

A more detailed expression study revealed that the spatial patterning of floral expression of *MPF2*-like genes is much more variable in Iochrominae than expected. Detection of *MPF2*-like gene expression in mature pollen of several Physaleae species suggests a more common role of these genes in male fertility. Together with further results this gave rise to the hypothesis that initially expression in pollen, improving male fertility, might have been the advantageous and therefore selected trait, while the

widespread *MPF2*-like gene expression in sepals of non-ICS species might have been the result of “hitch-hiking” effects.

6. Zusammenfassung

Viele Genera der Familie der Nachtschattengewächse (Solanaceae) beherbergen Arten, deren stark vergrößerte Kelchblätter eine laternenartige Hüllstruktur ausbilden. Diese morphologische Neuheit wird bei *Physalis* „Chinesische Laterne“ oder auch „Inflated-Calyx-Syndrome“ (ICS) genannt. In *Physalis floridana* wurde gezeigt, dass der MADS-Box Transkriptionsfaktor *MPF2* in floralen Organen exprimiert wird, während die Expression des orthologen Gens *STMADS16* in *Solanum tuberosum* auf vegetative Gewebe beschränkt ist. Die Expression von *MPF2* in der Blüte spielt nicht nur bei der Laternenbildung eine Rolle, sondern scheint auch ein wichtiger Bestandteil des männlichen Fertilitätsprogramms von *P. floridana* zu sein. Es ist bisher nicht bekannt, ob die Rekrutierung dieses Transkriptionsfaktors und seine Integration in einen neuen Kontext durch Änderungen in *cis*- oder *trans*-regulatorischen Elementen erfolgten. Im Tribus Physaleae scheint es sich bei der Expression von *MPF2*-ähnlichen Genen in Sepalen um ein plesiomorphes Merkmal zu handeln, welches in einigen Iochrominae-Arten wie *Iochroma australe* und *Dunalia fasciculata* wahrscheinlich durch sekundäre Veränderungen wieder verloren wurde. Dies war offensichtlich in der nahverwandten Art *Vassobia breviflora* nicht der Fall.

Das vorgestellte Projekt wurde initiiert, um zu erforschen, welche Rolle *cis*-Elemente bei der Regulierung der Transkription von *MPF2*-ähnlichen MADS-Box Transkriptionsfaktoren im Tribus Physaleae und insbesondere im Untertribus Iochrominae spielen. *MPF2*-ähnliche Gene, ihre stromaufwärts gelegenen putativen *cis*-regulatorischen Bereiche sowie das erste Intron in der codierenden Sequenz wurden isoliert und mit verschiedenen bioinformatischen Programmen analysiert. Die Ergebnisse zeigten, dass eine kurze Region vor der MADS-Box und weitere Regionen im ersten Intron in allen untersuchten Arten konserviert waren. Diese Regionen enthielten einige bekannte Bindemotive für Transkriptionsfaktoren. Des Weiteren wurden in *I. australe* and *D. fasciculata* konservierte Motive gefunden, die in allen anderen untersuchten Physaleae-Arten fehlten. Die funktionelle Analyse mittels Promoter::*GUS* und Promoter-Intron::*GUS* Fusions-Konstrukten in transgenen *Arabidopsis thaliana*-Pflanzen deutete darauf hin, dass die hochkonservierte Region vor der MADS-Box *cis*-Elemente zu enthalten scheint, welche die Expression des Gens in floralen Organen hochregulieren können, während das erste Intron von

I. australe scheinbar die Expression des Gens in Sepalen unterdrücken kann. *cis*-regulatorische Elemente in diesem Bereich könnten dafür verantwortlich sein.

Eine detailliertere Expressionsstudie zeigte, dass die räumliche Ausbreitung der floralen Genexpression in Iochrominae-Arten sehr viel variabler zu sein scheint, als zuvor angenommen wurde. Die Entdeckung von Expression *MPF2*-ähnlicher Gene in Pollen mehrerer Physaleae-Arten deutete an, dass diese Gene generell eine sehr wichtige Rolle im männlichen Fertilitätsprogramm von Physaleae-Arten spielen könnten. Dieses und weitere Ergebnisse führten zu der Hypothese, dass eventuell primär die Genexpression in Pollen, welche die männliche Fertilität verbesserte, das vorteilhafte und deswegen selektierte Merkmal darstellte, während die auch in ICS-losen Arten recht verbreitete Genexpression in Sepalen ein Ergebnis sogenannter „hitch-hiking"-Effekte sein könnte.

7. References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell*, **15**, 63-78.

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R. and Tabin, C.J. (2004) Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, **305**, 1462-1465.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

Aoyama, T., Dong, C.H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., Morelli, G. and Chua, N.H. (1995) Ectopic Expression of the Arabidopsis Transcriptional Activator Athb-1 Alters Leaf Cell Fate in Tobacco. *Plant Cell*, **7**, 1773-1785.

Baranowskij, N., Frohberg, C., Prat, S. and Willmitzer, L. (1994) A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. *Embo J*, **13**, 5383-5392.

Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol Biol*, **20**, 1195-1197.

Brudno, M., Steinkamp, R. and Morgenstern, B. (2004) The CHAOS/DIALIGN WWW server for multiple alignment of genomic sequences. *Nucleic Acids Res*, **32**, W41-44.

Carroll, S.B., Grenier, J.K. and Weatherbee, S.D. (2001) *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden: Blackwell Science.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*, **16**, 735-743.

Codex Aniciae Julianae (before 512) f.359v, Austrian National Library. In *Garden Eden: Masterpieces of Botanical Book Illustration* (Lack, W.H. ed: Taschen, pp. 31.

Coen, E.S. and Meyerowitz, E.M. (1991) The War of the Whorls - Genetic Interactions Controlling Flower Development. *Nature*, **353**, 31-37.

Colasanti, J., Tremblay, R., Wong, A.Y., Coneva, V., Kozaki, A. and Mable, B.K. (2006) The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genomics*, **7**, 158.

Colasanti, J., Yuan, Z. and Sundaresan, V. (1998) The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. *Cell*, **93**, 593-603.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H.J., Angenent, G.C. and van Tunen, A.J. (1995) The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell*, **7**, 1859-1868.

D'Arcy, W. (1991) Taxonomy, Chemistry, Evolution. In *Solanaceae III* (Hawkes, J.G., Lester, R.W., Nee, M. and Estrada, R.N. eds). Kew, 1991: Royal Botanic Gardens.

Darwin, C. (1859) *On the Origin of Species, A Facsimile on the First Edition* Cambridge, MA: Harvard University Press.

Davies, B., DiRosa, A., Eneva, T., Saedler, H. and Sommer, H. (1996a) Alteration of tobacco floral organ identity by expression of combinations of *Antirrhinum* MADS-box genes. *Plant J*, **10**, 663-677.

Davies, B., EgeaCortines, M., Silva, E.D., Saedler, H. and Sommer, H. (1996b) Multiple interactions amongst floral homeotic MADS box proteins. *Embo J*, **15**, 4330-4343.

Davies, B., Motte, P., Keck, E., Saedler, H., Sommer, H. and Schwarz-Sommer, Z. (1999) *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *Embo J*, **18**, 4023-4034.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. and Yanofsky, M.F. (2004) The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr Biol*, **14**, 1935-1940.

Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell*, **10**, 1075-1082.

Egea-Cortines, M., Saedler, H. and Sommer, H. (1999) Ternary complex formation between the MADS-box proteins *SQUAMOSA*, *DEFICIENS* and *GLOBOSA* is involved in the control of floral architecture in *Antirrhinum majus*. *Embo J*, **18**, 5370-5379.

Filichkin, S.A., Leonard, J.M., Monteros, A., Liu, P.P. and Nonogaki, H. (2004) A novel endo-beta-mannanase gene in tomato LeMAN5 is associated with anther and pollen development. *Plant Physiol*, **134**, 1080-1087.

Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J.P., Meller, J., Elber, R., Alpert, K.B. and Tanksley, S.D. (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85-88.

Garcia-Maroto, F., Ortega, N., Lozano, R. and Carmona, M.J. (2000) Characterization of the potato MADS-box gene STMADS16 and expression analysis in tobacco transgenic plants. *Plant Mol Biol*, **42**, 499-513.

He, C.Y., Münster, T. and Saedler, H. (2004) On the origin of floral morphological novelties. *Febs Lett*, **567**, 147-151.

He, C.Y. and Saedler, H. (2005) Heterotopic expression of *MPF2* is the key to the evolution of the Chinese lantern of *Physalis*, a morphological novelty in Solanaceae. *P Natl Acad Sci USA*, **102**, 5779-5784.

He, C.Y. and Saedler, H. (2007) Hormonal control of the inflated calyx syndrome, a morphological novelty, in Physalis. *Plant J*, **49**, 935-946.

He, C.Y., Sommer, H., Grosardt, B., Huijser, P. and Saedler, H. (2007) PFMAGO, a MAGO NASHI-like factor, interacts with the MADS-domain protein MPF2 from Physalis floridana. *Mol Biol Evol*, **24**, 1229-1241.

Hepper, F.N. (1991) Old World Withania (Solanaceae): a taxonomic review and key to the species. . In *Solanaceae III: taxonomy, chemistry, evolution.* (Hawkes, J.G., Lester, R.N., Nee, M. and Estrada, E. eds). Kew, UK Royal Botanic Gardens, pp. 211-227.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res*, **27**, 297-300.

Hileman, L.C., Kramer, E.M. and Baum, D.A. (2003) Differential regulation of symmetry genes and the evolution of floral morphologies. *P Natl Acad Sci USA*, **100**, 12814-12819.

Hong, R.L., Hamaguchi, L., Busch, M.A. and Weigel, D. (2003) Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell*, **15**, 1296-1309.

Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.

Honys, D. and Twell, D. (2003) Comparative analysis of the Arabidopsis pollen transcriptome. *Plant Physiol*, **132**, 640-652.

Hu, J.Y. and Saedler, H. (2007) Evolution of the inflated calyx syndrome in solanaceae. *Mol Biol Evol*, **24**, 2443-2453.

Huijser, P., Klein, J., Lonnig, W.E., Meijer, H., Saedler, H. and Sommer, H. (1992) Bracteomania, an Inflorescence Anomaly, Is Caused by the Loss of Function of the Mads-Box Gene Squamosa in Antirrhinum-Majus. *Embo J*, **11**, 1239-1249.

Hunziker, A.T. (2001) *The Genera of Solanaceae* Ruggell, Lichtenstein: A.R.G. Gantner Verlag K.G.

Ishiguro, S. and Nakamura, K. (1992) The nuclear factor SP8BF binds to the 5'-upstream regions of three different genes coding for major proteins of sweet potato tuberous roots. *Plant Mol Biol*, **18**, 97-108.

Johannesson, H., Wang, Y., Hanson, J. and Engstrom, P. (2003) The Arabidopsis thaliana homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol Biol*, **51**, 719-729.

Kagaya, Y., Ohmiya, K. and Hattori, T. (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res*, **27**, 470-478.

Kanno, A., Saeki, H., Kameya, T., Saedler, H. and Theissen, G. (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol Biol*, **52**, 831-841.

Keys, D.N., Lewis, D.L., Selegue, J.E., Pearson, B.J., Goodrich, L.V., Johnson, R.L., Gates, J., Scott, M.P. and Carroll, S.B. (1999) Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science*, **283**, 532-534.

Khan, M.R. (2009) Molecular evolution of the inflated calyx syndrome (ICS) in Withania (Solanaceae). In *PhD thesis (Genetics)*: University of Cologne.

Khan, M.R., Hu, J.Y., Riss, S., He, C. and Saedler, H. (2009) MPF2-like-A MADS-box genes control the inflated calyx syndrome in Withania (Solanaceae): roles of Darwinian selection. *Mol Biol Evol.*

Kim, S.H., Mizuno, K. and Fujimura, T. (2002) Isolation of MADS-box genes from sweet potato (Ipomoea batatas (L.) lam.) expressed specifically in vegetative tissues. *Plant and Cell Physiology*, **43**, 314-322.

Knapp, S. (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Journal of Experimental Botany*, **53**, 2001-2022.

Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T. and Yu, H. (2007) Specification of Arabidopsis floral meristem identity by repression of flowering time genes. *Development*, **134**, 1901-1910.

Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J. and Coen, E. (1999) Control of organ asymmetry in flowers of *Antirrhinum*. *Cell*, **99**, 367-376.

Luo, D., Carpenter, R., Vincent, C., Copsey, L. and Coen, E. (1996) Origin of floral asymmetry in *Antirrhinum*. *Nature*, **383**, 794-799.

Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992) Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature*, **360**, 273-277.

Marcellini, S. and Simpson, P. (2006) Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *Plos Biol*, **4**, e386.

Martinez-Garcia, J.F., Moyano, E., Alcocer, M.J. and Martin, C. (1998) Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. *Plant J*, **13**, 489-505.

Mayr, E. (1960) The emergence of evolutionary novelties. In *Evolution after Darwin* (Tax, S. ed. Chicago: University of Chicago Press, pp. 349-380.

Merril, E.D. (1923) *An Enumeration of Philippine Flowering Plants* Manila.

Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, **15**, 211-218.

Morton, J.F. (1987) Cape Gooseberry. In *Fruits of warm climates.* Miami, FL., pp. 430-434.

Olmstead, R.G., Sweere, J.A., Spangler, R.E., Bohs, L. and Palmer, J.D. (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In *Solanaceae IV: advances in biology and utilization.* (Nee, M., Symon, D.E., N., L.R. and Jessop, J.P. eds). Kew: Royal Botanic Gardens, pp. 111-137.

Ovcharenko, I., Loots, G.G., Giardine, B.M., Hou, M., Ma, J., Hardison, R.C., Stubbs, L. and Miller, W. (2005) Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res*, **15**, 184-194.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.

Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.L. and Yanofsky, M.F. (2001a) APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J*, **26**, 385-394.

Pelaz, S., Tapia-Lopez, R., Alvarez-Buylla, E.R. and Yanofsky, M.F. (2001b) Conversion of leaves into petals in Arabidopsis. *Curr Biol*, **11**, 182-184.

Pollock, R. and Treisman, R. (1991) Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev*, **5**, 2327-2341.

Raff, R.A., Parr, B.A., Parks, A.L. and Wray, G.A. (1990) Heterochrony and other mechanisms of radical evolutionary change. In *Evolutionary Innovations* (Nitecki, M.H. ed. Chicago: University of Chicago Press, pp. 71-98.

Riechmann, J.L., Wang, M. and Meyerowitz, E.M. (1996) DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res*, **24**, 3134-3141.

Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, **132**, 365-386.

Sambrook, J. and Russell, D.W. (2001) *Molecular cloning: a laboratory manual*: Cold Spring Harbor Laboratory Press.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990) Genetic-Control of Flower Development by Homeotic Genes in *Antirrhinum majus*. *Science*, **250**, 931-936.

Shang, Y., Schwinn, K.E., Bennett, M.J., Hunter, D.A., Waugh, T.L., Pathirana, N.N., Brummell, D.A., Jameson, P.E. and Davies, K.M. (2007) Methods for transient assay of gene function in floral tissues. *Plant Methods*, **3**, 1.

Shore, P. and Sharrocks, A.D. (1995) The MADS-box family of transcription factors. *Eur J Biochem*, **229**, 1-13.

Smith, S.D. and Baum, D.A. (2006) Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). *American Journal of Botany*, **93**, 1140-1153.

Smith, S.D. and Baum, D.A. (2007) Systematics of iochrominae (Solanaceae): Patterns in floral diversity and interspecific crossability. *Proceedings of the VIth International Solanaceae Conference, Solanaceae VI: Genomics Meets Biodiversity*, 241-254

Smith, S.D., Hall, S.J., Izquierdo, P.R. and Baum, D.A. (2008) COMPARATIVE POLLINATION BIOLOGY OF SYMPATRIC AND ALLOPATRIC ANDEAN IOCHROMA (SOLANACEAE). *ANN. MISSOURI BOT. GARD.* , **95**, 600-617.

Sommer, H., Beltran, J.P., Huijser, P., Pape, H., Lonnig, W.E., Saedler, H. and Schwarz-Sommer, Z. (1990) *Deficiens*, a Homeotic Gene Involved in the Control of Flower Morphogenesis in *Antirrhinum majus* - the Protein Shows Homology to Transcription Factors. *Embo J*, **9**, 605-613.

Tang, W. and Perry, S.E. (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. *J Biol Chem*, **278**, 28154-28159.

Tatematsu, K., Ward, S., Leyser, O., Kamiya, Y. and Nambara, E. (2005) Identification of cis-elements that regulate gene expression during initiation of axillary bud outgrowth in Arabidopsis. *Plant Physiol*, **138**, 757-766.

Theissen, G. (2006) The proper place of hopeful monsters in evolutionary biology. *Theory Biosci*, **124**, 349-369.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Munster, T., Winter, K.U. and Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol Biol*, **42**, 115-149.

Theißen, G., Becker, A., Kirchner, C., Münster, T., Winter, K.-U. and Saedler, H. (2002) How land plants learned their floral ABCs: the role of MADS-box genes in the evolutionary origin of flowers. In *Developmental Genetics and Plant Evolution* (Cronk, Q.C.B., Bateman, R.M. and Hawkins, J.A. eds). London: Taylor & Francis, pp. 173-205.

Theissen, G. and Saedler, H. (2001) Plant biology - Floral quartets. *Nature*, **409**, 469-471.

Troebner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lonnig, W.E., Saedler, H., Sommer, H. and Schwarz-Sommer, Z. (1992) *Globosa* - a Homeotic Gene Which Interacts with *Deficiens* in the Control of *Antirrhinum* Floral Organogenesis. *Embo J*, **11**, 4693-4704.

Urao, T., Yamaguchi-Shinozaki, K., Urao, S. and Shinozaki, K. (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell*, **5**, 1529-1539.

Usuka, J., Zhu, W. and Brendel, V. (2000) Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics*, **16**, 203-211.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W. and Giovannoni, J. (2002) A MADS-box gene necessary for fruit ripening at the *tomato ripening-inhibitor* (*Rin*) locus. *Science*, **296**, 343-346.

Wake, D.B. and Roth, G. (1989) The linkage between ontogeny and phylogeny in the evolution of complex systems. In *Complex Organismal Functions: Integration and Evolution in Vertebrates* (Wake, D.B. and Roth, G. eds). Chichester, London, New York: Wiley, pp. 361-377.

Wang, R.L., Stec, A., Hey, J., Lukens, L. and Doebley, J. (1999) The limits of selection during maize domestication. *Nature*, **398**, 236-239.

Weigel, D. and Meyerowitz, E.M. (1994) The Abcs of Floral Homeotic Genes. *Cell*, **78**, 203-209.

Whitson, M. and Manos, P.S. (2005) Untangling Physalis (Solanaceae) from the physaloids: A two-gene phylogeny of the physalinae. *Systematic Botany*, **30**, 216-230.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A. and Meyerowitz, E.M. (1990) The Protein Encoded by the *Arabidopsis* Homeotic Gene *Agamous* Resembles Transcription Factors. *Nature*, **346**, 35-39.

Yoshioka, S., Taniguchi, F., Miura, K., Inoue, T., Yamano, T. and Fukuzawa, H. (2004) The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. *Plant Cell*, **16**, 1466-1477.

Yu, H., Ito, T., Wellmer, F. and Meyerowitz, E.M. (2004) Repression of *AGAMOUS-LIKE 24* is a crucial step in promoting flower development. *Nat Genet*, **36**, 157-161.

Zheng, J.B., Zhou, Y.H., Maity, T., Liao, W.S.L. and Saunders, G.F. (2001) Activation of the human PAX6 gene through the exon 1 enhancer by transcription factors SEF and Sp1. *Nucleic Acids Research*, **29**, 4070-4078.

8. Supplement

8.1 Supplementary tables

Table 8.1.1: Genbank accession numbers of downloaded *MPF2*-like ESTs and mRNA sequences containing partial coding sequences

Cultivar	Species	EST/mRNA, partial cds
Cap002	*Capsicum baccatum L.*	EF395183 (*CbaM201_1*)
Du001	*Dunalia fasciculata (Miers) Sleumer*	EF395185 (*DfaM201_3*)
I001	*Iochroma australe Griseb.*	EF395186 (*IauM201_2*)
M001	*Margaranthus solanaceus Schltdl.*	EF395189 (*MsoM201_1*)
P001	*Physalis acutifolia (Miers) Sandwith*	EU292995 (*PacM201_3*)
P011	*Physalis alkekengi L. var. alkekengi*	EF395198 (PalM211_1)
P022	*Physalis crassifolia Benth.*	EF395221 (*PcrM222_1*)
P030	*Physalis fuscomaculata Dunal*	EF395208 (*PfuM230_1*)
P035	*Physalis lanceifolia Nees*	EF395213 (*PlaM235_1*)
P039	*Physalis minima L.*	EF395211 (*PmiM239_1*) EF395212 (*PmiM239_2*)
P055	*Physalis philadelphica*	EF395215 (*PphM255_1*)
TA496	*Solanum lycopersicum*	AI482810 AI482813 AW929235
Micro-Tom	*Solanum lycopersicum*	DB685265
R11-12	*Solanum lycopersicum*	AI773429

Table 8.1.2: Genbank accession numbers of downloaded *MPF2*-like mRNAs containing complete cds

Cultivar	Species	mRNA, complete cds
P105	*Physalis peruviana*	AY643731 (*MPP3*) AY643732 (*MPP4*)
P106	*Physalis floridana /Physalis pubescens*	AY643734 (*MPF2*)
S032	*Solanum tuberosum*	AY643733 (*STMADS16*)
T001	*Tubocapsicum anomalum (Franch. & Sav.) Makino*	FM956485 (*TAB201*)
V001	*Vassobia breviflora (Sendtn.) A.T.Hunz*	EF395177 (*VbrM201_1*)
W006	*Withania somnifera (L.) Dunal*	FM956486 (*WSA206*) FM956487 (*WSB206*)
Wi001	*Witheringia coccoloboides (Dammer) A.T.Hunz*	EF395175 (*WIcoM201_3*)
Wi002	*Witheringia solanacea L'Her.*	EF395174 (*WIsoM202_3*)
-	*Solanum macrocarpon*	AY643730 (*MSM2*)

Table 8.1.3: Genbank accession numbers of downloaded *MPF2*-like genomic sequences

Cultivar	Species	Genomic sequences
Heinz 1706	*Solanum lycopersicon*	Clone containing locus CU468638 (*SlyM2*)
P106	*Physalis floridana/pubescens*	Locus + 5'upstream AY643735 (MPF2)
S032	*Solanum tuberosum*	Locus + 5'upstream AY643736 (*STMADS16*)
T001	*Tubocapsicum anomalum (Franch. & Sav.) Makino*	Locus FM956484 (*TAB201*)
W006	*Withania somnifera (L.) Dunal*	Locus FM956482 (*WSA206*) Locus FM956483 (*WSB206*)

Table 8.1.4: Genbank accession numbers of *MPF2*-like sequences from *P. acutifolia*

Cultivar	Species	Type	Sequences
P001	*Physalis acutifolia*	mRNA, containing complete cds	FN356424 (*PacM201_1*)
		5' upstream sequence	FN356448 (*pPacM201_1*) FN356449 (*pPacM201_2*)

The sequences were isolated in this study, but not used for further analysis, since the *MPF2*-like gene expression pattern in sepals of *P. acutifolia* is still unclear.

Table 8.1.5: *SlyM2* - hypothetical *MPF2*-like mRNA from *S. lycopersicum* (cultivar: Heinz 1706)

AACAACAAGTGAATAGTACTTATTCCATCTCCAAAATTAGGTCTTACATATTTCT
ATACTCAATTATTGGCCCATTTTTGTATAGGTTTTCTAACTAGATCTAAAGTTTTC
CTCTTTTATTAGGGTTTCTTACAATTTCTTACTATTTATATCTTTTGTGAGATTTCT
TAGAAAGTTTGAGGAAAGATAATTATCTTCTTCGACCTGTTAAGG**ATG**GCAAGG
GAAAAGATAAAAATAAAAAAGATAGATAATATAACAGCAAGACAAGTAACATT
TTCAAAGAGGAGAAGAGGGCTTTTTAAGAAAGCTGAAGAGCTTTCTGTTCTTTGT
GATGCTGATGTTGCTCTTATCATTTTTTCTGCTACTGGAAAGCTATTTGATTTTGC
TAGCACCAGCATGAAGGATATTCTTGGAAAGTATAAGTTGCAATCAGCTAGCCTT
GAGAAAGTTGACCAACCTTCCCTTGATTTACAGCTAGAGAATAGCCTCAACATG
AGATTAAGCAAGCAGGTAGCTGATAAAACTCGTGAGCTCAGGCAGATGAGAGGT
GAGGAACTTGAAGGATTGAGTTTAGAAGAATTACAACAAATTGAGAAAAGACTT
GAAGCTGGTTTCAACCGTGTGCTTGAGATTAAGGGTAAACGAATTATGGATGAA
ATTACCAACCTCCAAAGAAAGGGTGCTGAGCTGATGGAAGAAAACAAACAATTG
AAACACAAAATGGAAATTATGAAAAAAGGGAAATTGCCTTTAGTGACTGACATG
GTGATGGAAGAGGGCCAATCATCTGAGTCTATAATTACAACTAATAATCCTGATC
AAGATGATTCATCAAATGCATCTTTGAAGTTAGGTGGTACTACTGCAGTTGAAGA
TGATTGCTCAATTACATCTTTAAAGTTAGGGCTACCATTCAGC**TAA**CAAACTCAA
AGGAGGAAGCTTTTGTTGAAGATTGTACTAATAATAATTTTTGGATGATAACAAT
CAGTGGCATATATATATATATTTGCCTTCTTTTGGAAAAAATATTACATTGGAAT
ATAATTGTCACTGCTTAATTCTACTATGTATTATATGTATGTATGTGGACAACTTG
ACAAAGGGCATAATGCAGATTGATTAACCATCTCATTTTAGGTCTCGTGTGTGGG
CTTCAATTTCTATAATATAGTATATGTATTTACGTGAA

5'UTR in blue, coding region in red, 3'UTR in black, start + stop codons underlined

Table 8.1.6: List of taxa (adapted from Hu and Saedler 2007)

Cultivar	Species	Acc. No.	Native *	Source **
Cap002	*Capsicum baccatum L.*	904750135	PE, wild	BGN
Du001	*Dunalia fasciculata (Miers) Sleumer*	904750145	BO, wild	BGN
I001	*Iochroma australe Griseb.*	814750022	AR, wild	BGN
M001	*Margaranthus solanaceus Schltdl.*	904750101	MX, wild	BGN
P001	*Physalis acutifolia (Miers) Sandwith*	974750059	US, wild	BGN
P011	*Physalis alkekengi L. var. alkekengi*	914750015	AT, wild	BGN
P022	*Physalis crassifolia Benth.*	974750142	US, wild	BGN
P030	*Physalis fuscomaculata Dunal*	904750141	not wild	BGN
P035	*Physalis lanceifolia Nees*	944750048	not wild	BGN
P039	*Physalis minima L.*	884750095	not wild	BGN
P055	*Physalis philadelphica*	PI1 97691 97GI	MX, wild	PGRU
P105	*Physalis peruviana*	P105	not wild	MPIZSB
P106	*Physalis floridana*	P106	not wild	MPIZSB
S032	*Solanum tuberosum*	P40	wild	MPIZSB
T001	*Tubocapsicum anomalum (Franch. & Sav.) Makino*	904750027	CN , wild	BGN
V001	*Vassobia breviflora (Sendtn.) A.T.Hunz*	804750001	not wild	BGN
W002	*Withania coagulans (Stocks) Dunal*	914750053	not wild	BGN
W006	*Withania somnifera (L.) Dunal*	894750106	CY, wild	BGN
W007	*Withania somnifera (L.) Dunal*	904750143	not wild	BGN
Wi001	*Witheringia coccoloboides (Dammer) A.T.Hunz*	814750081	CO, wild	BGN
Wi002	*Witheringia solanacea L'Her.*	814750082	PA, wild	BGN

* Native = abbreviation name of the country, where the accession was collected (wild = wild origin): AR, Argentina; AT, Austria; BO, Bolivia; CO, Colombia; CN, China; CY, Cyprus; MX, Mexico; PA, Panama; PE, Peru; US, USA.

** Source = seed banks providing the seeds: BGN, Botanical & Experimental Garden of Radboud University, Nijmegen, Netherlands; MPIZSB, seed bank of the Max-Planck-Institute for Plant Breeding Research; PGRU, Plant Genetic Resources Unit, USDA-ARS.

Table 8.1.7: Oligonucleotide primers (-F: forward primer, -R: reverse primer)

Name	Sequence 5'->3'	Remarks
Primers for RAGE (*MPF2*-like upstream sequences)		
P1-R	GAGCAACATCAGCATCAAAAGAACAGAAAG	1. PCR, universal,
P2-R	CCTCTTCTCCTCTTTGAAAATGTCACTTGTC	Nested, universal
Primers for 5'RACE (*MPF2*-like cDNAs)		
P1-R	GAGCAACATCAGCATCAAAAGAACAGAAAG	universal
P2-R	CCTCTTCTCCTCTTTGAAAATGTCACTTGTC	universal
PhR1-R	CWGABTGCAAMTTATATTTTCCAAGG	Physalis
PhR2-R	KGAYATYHTTCATGCTGGAGCTA	Physalis
PhR3-R	TGCTGGAGCTAGMRAAWTCAAAGAG	Physalis
22R1-R	GACTGCAACTTATATTTTTCAATGAT	P022
C2R3-R	AGCAAATTCAAAAAGCTTTCCAGTA	Cap002
C2R2-R	AAGGATATCATTCATGCTGGAGCTA	Cap002
C2R1-R	CTCATGGCTAGATGTCTGCAACTTA	Cap002
Primers for 3'RACE (*MPF2*-like cDNAs)		
3R01-F	GTGAACAGTGCTGATTCCATCTCC	Du001, Wi002
3R02-F	ATGGCAAGAGAGAAGATCAAGATAAGGAAG	P011
3R03-F	ATGGCAAGAGAGAAGATCAAGATAAAGAAG	Du001
3R04-F	ATGGCAAGAGAGAAGATAAAGATAAAGAAG	Wi001, Wi002
16F2-F	ACAGCAAGRCARGTRACATTTTCAAAG	Wi001, P011
16F3-F	GTTCTYTGTGATGCTGATGTTGCTCTYAT	universal
Primers for *MPF2*-like upstream sequence and intron amplification		
P2-R	CCTCTTCTCCTCTTTGAAAATGTCACTTGTC	MADS, universal
Vfw1-F	ATATTTCATGCCTGGAAATGATGACAAAAG	Iochrominae
Vrv1-R	GTTGGTCAACTTTYTCAAAGCTAGCTGAMT	I-region, Iochrominae
P11F-F	AGAGGAAGAAGAGGGAACTTACCACAAATG	P011
P11R-R	CAACTTTCTCAAGGTTAGCTGACTGCAACT	I-region, P011

Wi16-F	CAAAGCAACAGTAGTGCATTAACAA	Wi001
Wi2F-F	GAAAAATGGCCAAAAATAGCAAACATAAGG	Wi002
Wi2R-R	ATCTTTCTCAAGGTTAGCTGAGTGCAACTT	I-region, Wi002
p35f-F	CCTTTAAGTTTGGTATTTACACGAGA	Physalis, M001
M107-F	TTTCGTTCCATTTTCACTAATTTTCTCACA	Physalis, M001
35f2-F	TTAGCCTCCAATGCATTAAATATTTGTCCT	P035
106r-R	AGCAGACTGCAACTTATATTTTCCAAGGAT	I-region, Physalis

Primers for RT-PCR (Expression analysis)		
M-F	CTCAACRTGAGATTAMGCAAGCA	*MPF2*-like; universal
M-R	AGASTCAGATGATTGRCCTTCTT	*MPF2*-like; universal
18S-F	TGCAGTTAAAAAGCTCGTAGTTG	18S rRNA; universal
18S-R	ACATCTAAGGGCATCACAGAC	18S rRNA; universal

Primers for Iochrominae promoter constructs (HindIII)		
VfwC-F	TTTTAAGCTTATATTTCATGCCTGGAAATGATGACA	Type 2+4 Iochrominae
DIVK-F	ATGATCGAAGCTTTCATTTTCTTAATTTTCA	Type 1+3 Iochrominae
Vr1C-R	TCTTAAGCTTGATCTTCTCTCTTGCCAATCTAAAAA	Type 1+2; *VbrM201_1/2*, *DfaM201_1*
VrMC-R	TCTTAAGCTTGATCTTCTCTCTTGCCAATCTTA	Type 1+2; *IauM201_1/2, DfaM201_2*
VrIN-R	TTTAAGCTTGTCAACTTTTTCAAAGCTAGCTGACTG	Type 3+4; Iochrominae

Primers for *P. alkekengi* promoter constructs (HindIII)		
P11F-F	ATAAAGCTTAGAGGAAGAAGAGGGAACTTACCACAA	Type 2+4
P11K-F	ACAAAGCTTTAAAGAAGGCATGTCCCATAA	Type 1+3
P11R-R	TCTTAAGCTTGATCTTCTCTCTTGCCAACCTAAAAA	Type 1+2
P11I-R	AAAAAGCTTCAACTTTCTCAAGGTTAGCTGACTG	Type 3+4

Primers for *W. solanacea* promoter constructs (AvrII⇒XbaI**)**		
W2FA-F	GAACCTAGGCCAAAAATAGCAAACATAAGG	Type 2+4
W2TA-R	TATCCTAGGCTTCTCTCTTGCCAAATTGATCC	Type 2
W2IA-R	TCTTTCTCCTAGGTAGCTGAGTGCAACTT	Type 4

Primers for *W. coccoloboides* promoter constructs (HindIII)		
kfw2-F	GCTTCAAAGCTTCAGTAGTGCATTAACAACCA	Type 1/2
KMAD-R	TCTTCTCTCAAGCTTTATTGATCCCCCTTT	Type 1/2

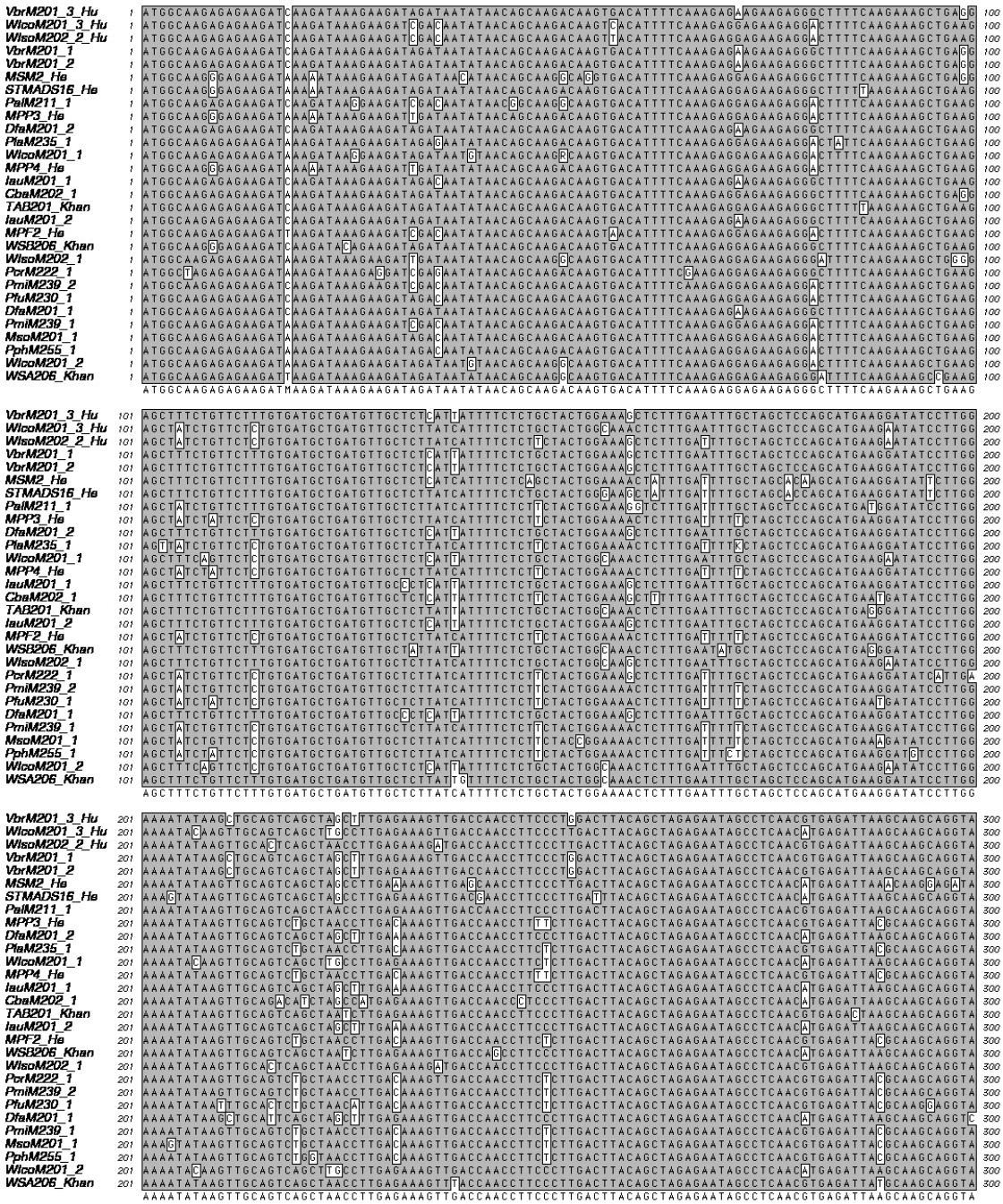
Primers for *P. lanceifolia* promoter constructs (XmaI)		
35FC-F	ATTCACCCGGGATTTTAGCCTCCAATG	Type 2
35FK-F	TGAAATAAAAATAGACCCGGGAGTGGTTAAT	Type 3
35RM-R	TCTTCTCCCGGGCCAATCTATACAAAA	Type 2
35RI-R	AACTTATATTTTCCCGGGATATCCTTCAT	Type 3

Primers for *MPF2*-like-A gene (*WSA206*) probe for Southern blotting	
K122-F	GCCCTTAACTAACAGACATGGTGATG
K123-R	GTCTCCATCCAAAAGTTATTAGTACAATC

Primers for *MPF2*-like-B gene (*WSB206*) probe for Southern blotting	
K124-F	CCTTAGTAATAGACATGGATTGCATGGTG
K125-R	CTTCAACAAAGCTTCCTTCATTGAGTTAG

8.2 Supplementary figures

Figure 8.2.1: Alignment of *MPF2*-like cDNA sequences



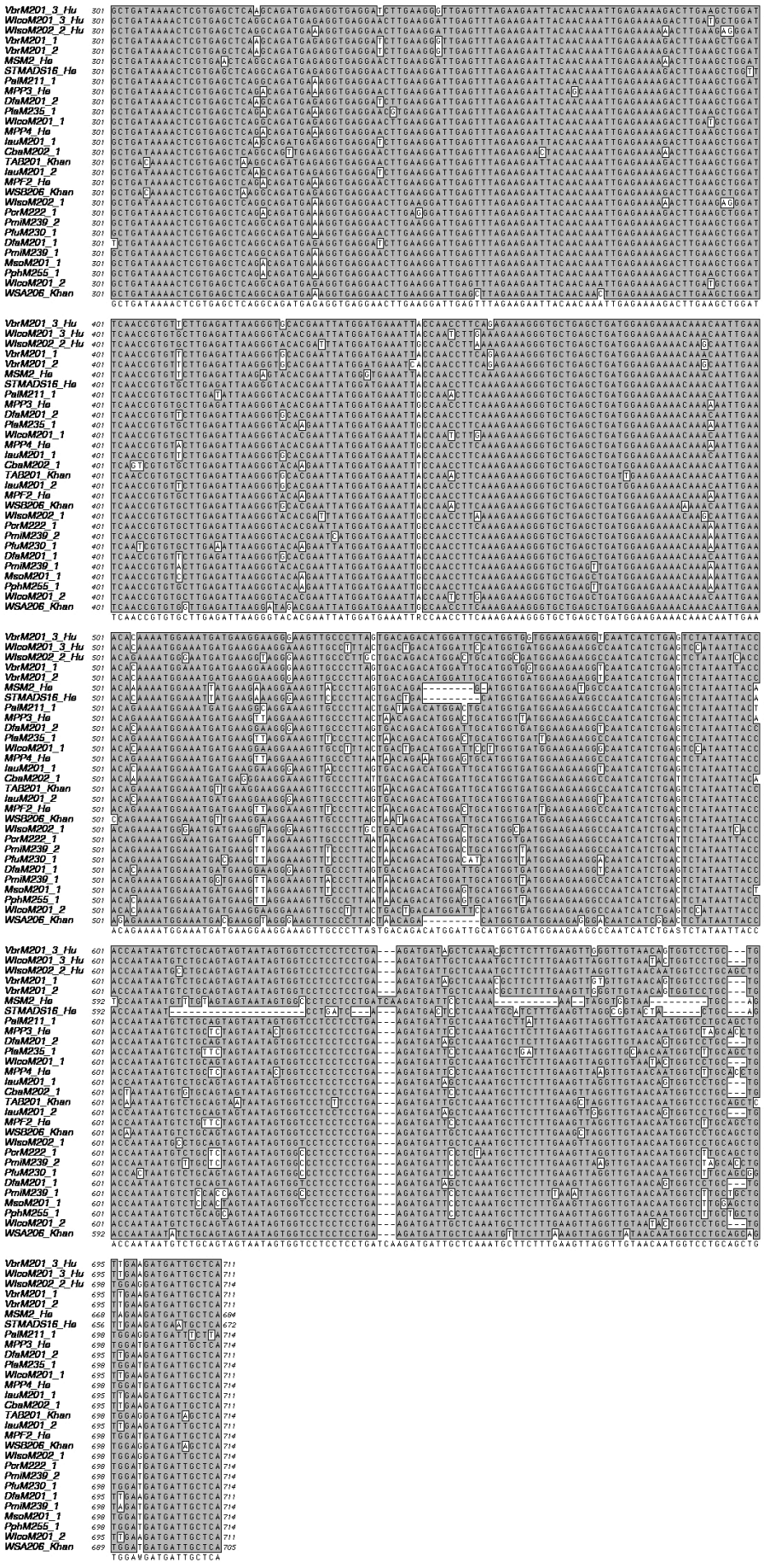


Figure 8.2.2: Alignment of MPF2-like protein sequences

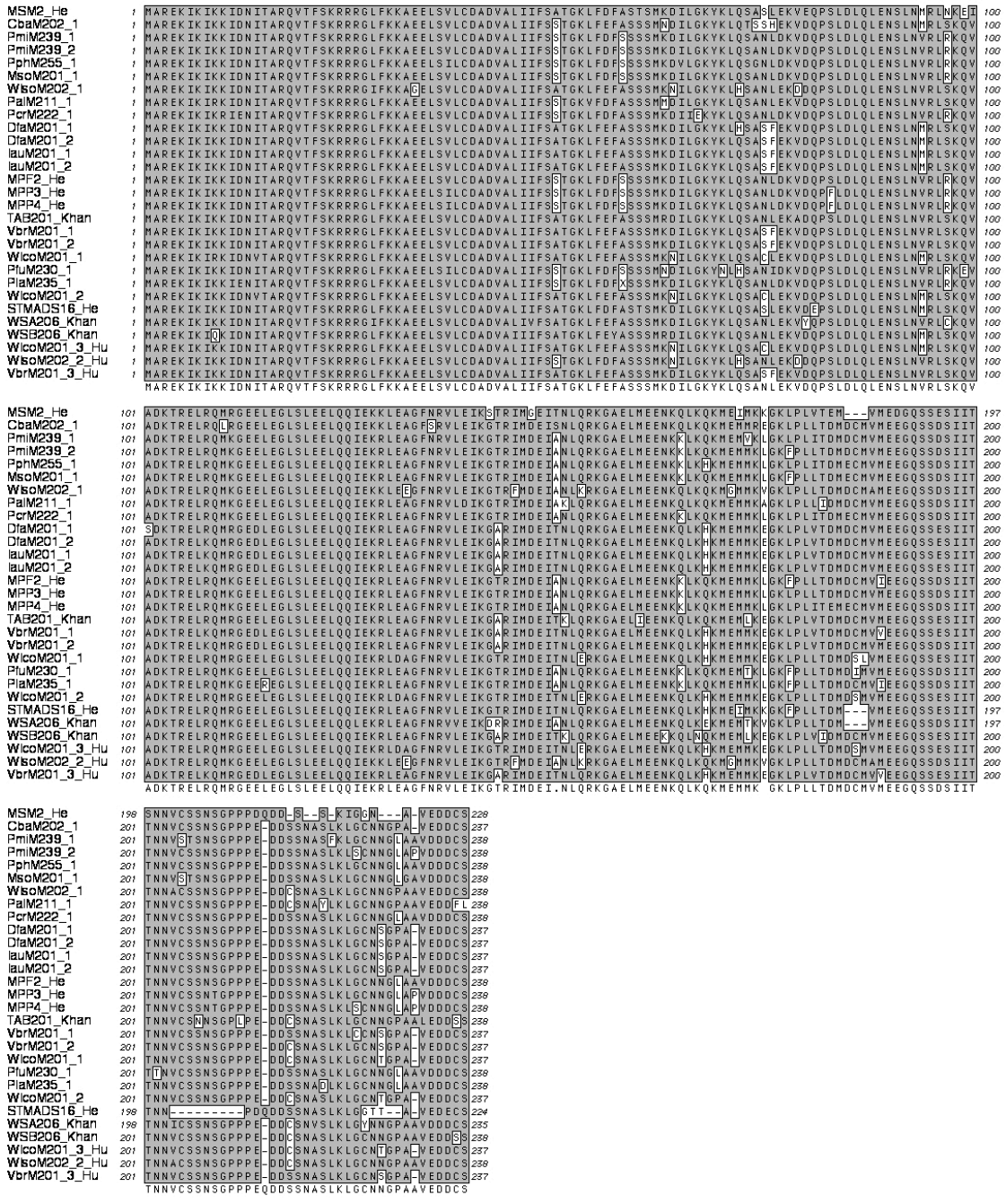
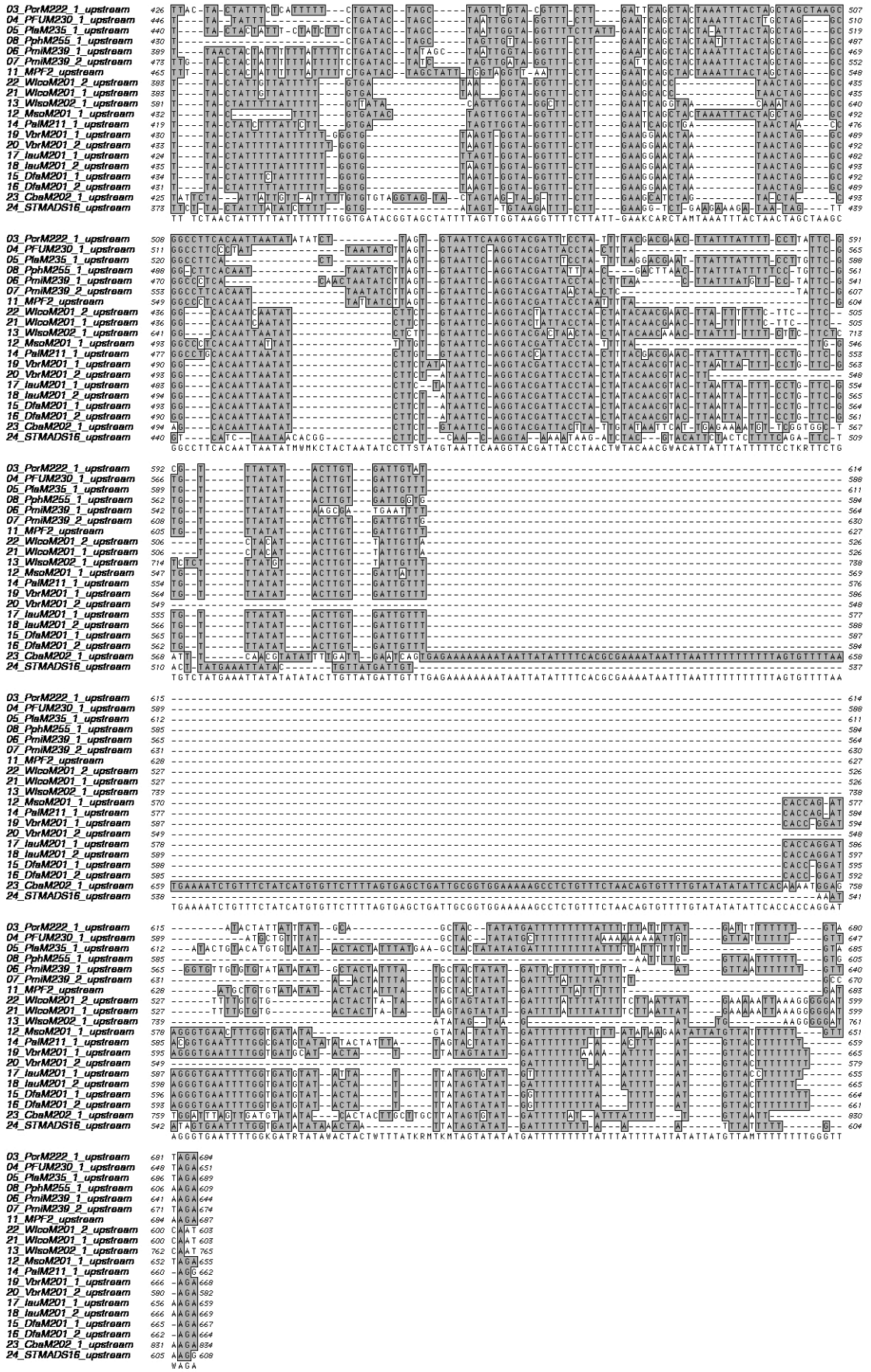


Figure 8.2.3: Alignment of upstream sequences of *MPF2*-like genes


03_PcrM222_1_upstream
04_PFUM230_1_upstream
05_PlaM235_1_upstream
08_PphM255_1_upstream
06_PmiM239_1_upstream
07_PmiM239_2_upstream
11_MPF2_upstream
22_WlcoM201_2_upstream
21_WlcoM201_1_upstream
13_WlsoM202_1_upstream
12_MsoM201_1_upstream
14_PalM211_1_upstream
19_VbrM201_1_upstream
20_VbrM201_2_upstream
17_lauM201_1_upstream
18_lauM201_2_upstream
15_DfaM201_1_upstream
16_DfaM201_2_upstream
23_CbaM202_1_upstream
24_STMADS16_upstream

8.3 Abbreviations

%	percent
°C	degree Celsius
3'	three prime end of a DNA fragment
35S	35S promotor from the Cauliflower Mosaic virus
5'	five prime end of a DNA fragment
Acc.	Accession
AGL	*AGAMOUS-LIKE*
BLAST	Basic Local Alignment Search Tool
bp	base pair
CArG	C A/T-rich G motif
cDNA	complementary DNA
cds	coding sequence
CTAB	hexadecyl-trimethyl-ammonium bromide
Col	*Arabidopsis thaliana* Columbia accession
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	desoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
ECR	Evolutionarily Conserved Region(s)
EDTA	ethylenediaminetetraacetic acid
et al.	et alii/aliae (Latin: “and others“)
Fig.	Figure
GUS	ß-Glucuronidase
h	hour
kb	kilobase
µ	micro (10^{-6})
m	milli (10^{-3})
M	molar (mol/m3)
MADS	acronym for the four founder proteins MCM1 (*S. cerevisiae*), AGAMOUS (*A. thaliana*), DEFICIENS (*A. majus*) and SRF (*H. sapiens*)
MES	2-(N-morpholino) ethanesulfonic acid

min	minute
MS	Murashige & Skoog
ORF	open reading frame
PCR	polymerase chain reaction
RACE	rapid amplification of cDNA ends
RAGE	rapid amplification of genomic DNA ends
RT	room temperature (20 to 23.5 °C with an average of 21 °C)
rpm	rounds per minute
RT-PCR	reverse transcriptase PCR
SRE	serum response element
UTR	untranslated region
Vol	volume
wt	wild type

8.4 Danksagung

Mein herzlichster Dank gilt:

meinem Betreuer und Gruppenleiter **Prof. Dr. Heinz Saedler** für die Möglichkeit, diese Arbeit am Max-Planck-Institut für Züchtungsforschung durchzuführen, die Überlassung des Laborarbeitsplatzes und des Themas und interessante Anregungen,

Prof. Dr. Martin Hülskamp für die Übernahme des Koreferats,

Prof. Dr. Ute Höcker für die Übernahme des Prüfungsvorsitzes während der Disputation,

Dr. Heiko Schoof für die Übernahme des Beisitzes während der Disputation,

Dr. Thomas Münster für die Durchsicht des Manuskripts, hilfreiche Kommentare und Hilfe bei den Phylogenierekonstruktionen,

Prof. Dr. Chaoying He für interessante Diskussionen während seiner Zeit am MPIZ, das Mitteilen von unveröffentlichten Daten, sowie für die Durchsicht des Manuskripts und hilfreiche Kommentare,

Daniela Liebsch für ihre stete und unerschöpfliche Hilfsbereitschaft bei allen Fragen und Problemen, inklusive vieler konstruktiver Vorschläge beim Schreiben der Arbeit und moralischer Unterstützung,

Britta Grosardt für ihre Hilfe beim Erlernen vieler Standardtechniken in der Anfangsphase und das Überlassen zahlreicher Materialien,

Dr. Jinyong Hu und **Ramzan Khan** für die freundschaftliche Zusammenarbeit und für das Mitteilen von unveröffentlichten Daten,

Dr. Ulrike Göbel dafür, dass sie es geschafft hat, dass CREDO für die *in silico* Analysen wieder zur Verfügung stand,

Dr. Claus-Peter Stelzer für hilfreiche Vorschläge beim Schreiben der Arbeit und moralische Unterstützung,

außerdem **meinen Kollegen, meiner Familie und allen anderen, die mich während dieser Zeit in irgendeiner Form unterstützt haben!**

8.5 Eidesstattliche Erklärung

„Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Heinz Saedler betreut worden.“

Köln, den _____

8.6 Lebenslauf

Zur Person

Name	Simone Riß
Adresse	Kolibriweg 14, 50829 Köln
Geburtsdatum	02.03.1980
Geburtsort	Meppen
Nationalität	deutsch
Familienstand	ledig

Schulausbildung

1986 – 1990	Marienschule, Meppen
1990 – 1999	Gymnasium Marianum, Meppen
1999	Allgemeine Hochschulreife

Studium

1999 – 2005	Studium an der Westfälischen Wilhelms-Universität Münster Diplomstudiengang Biologie Hauptfach: Zoologie, Nebenfächer: Mikrobiologie und Jura
2004 - 2005	Diplomarbeit am Institut für Evolution und Ökologie der Tiere, Universität Münster Betreuer: Prof. Dr. Nico Michiels Thema: Analysis of innate immunity in the nematode *Caenorhabditis elegans*
04/2005	Abschluss als Diplom-Biologin
seit 02/2006	Doktorarbeit am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung Molekulare Pflanzengenetik Betreuer: Prof. Dr. Heinz Saedler Thema: Isolation and analysis of *MPF2*-like MADS-box genes from Physaleae and characterization of their *cis*-regulatory regions

Publikationen

Khan, M.R., Hu, J.Y., Riss, S., He, C. and Saedler, H. (2009) MPF2-like-A MADS-box genes control the inflated calyx syndrome in Withania (Solanaceae): roles of Darwinian selection. *Mol Biol Evol*. (Epub ahead of print)

Lüttmann, K., Anthes, N., D'Souza, T.G., Riss, S. and Michiels, N.K. (2006) Population size estimate of a reef-flat aggregation of Chromodoris annulata (Opisthobranchia, Chromodoridae). *Journal of Molluscan Studies*, **72**, 214-216.