Reciprocal Loss of CArG-Boxes and Auxin Response Elements Drives Expression Divergence of MPF2-Like MADS-Box Genes Controlling Calyx Inflation

Muhammad Ramzan Khan1,2, Jinyong Hu1, Ghulam Muhammad Ali2

1 Max-Planck-Institute for Plant Breeding Research, Department of Molecular Plant Genetics, Cologne, Germany; 2 National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Centre, Islamabad, Pakistan

Abstract
Expression divergence is thought to be a hallmark of functional diversification between homologs post duplication. Modification in regulatory elements has been invoked to explain expression divergence after duplication for several MADS-box genes, however, verification of reciprocal loss of cis-regulatory elements is lacking in plants. Here, we report that the evolution of MPF2-like genes has entailed degenerative mutations in a core promoter CArG-box and an auxin response factor (ARF) binding element in the large 1st intron in the coding region. Previously, MPF2-like genes were duplicated into MPF2-like-A and -B through genome duplication in Withania and Tubocapsicum (Withaninae). The calyx of Withania grows exorbitantly after pollination unlike Tubocapsicum, where it degenerates. Besides inflated calyx syndrome formation, MPF2-like transcription factors are implicated in functions both during the vegetative and reproductive development as well as in phase transition. MPF2-like-A of Withania (WSA206) is strongly expressed in sepal, while MPF2-like-B (WSB206) is not. Interestingly, their combined expression patterns seem to replicate the pattern of their closely related hypothetical progenitors from Vassobia and Physalis. Using phylogenetic shadowing, site-directed mutagenesis and motif swapping, we could show that the loss of a conserved CArG-box in MPF2-like-8 of Withania is responsible for impeding its expression in sepal. Conversely, loss of an ARE in MPF2-like-A relaxed the constraint on expression in sepal. Thus, the ARE is an active suppressor of MPF2-like gene expression in sepal, which in contrast is activated via the CArG-box. The observed expression divergence in MPF2-like genes due to reciprocal loss of cis-regulatory elements has added to genetic and phenotypic variations in the Withaninae and enhanced the potential of natural selection for the adaptive evolution of ICS. Moreover, these results provide insight into the interplay of floral developmental and hormonal pathways during ICS development and add to the understanding of the importance of polyploidy in plants.

Introduction
Gene or genome duplications and subsequent expression divergence due to modifications in regulatory elements create morphogenetic novelties [1]. However, our current understanding of the selective regulatory changes arising post-duplication is preliminary and fragmented. Among transcription factors families, the MADS family has greatly expanded in plants by a series of duplications that enabled the genes to diversify in structure and function [2]. Alterations in the expression profiles of MADS-domain proteins subsequent to regulatory changes are closely linked to the origin of developmental and morphological novelties, most conspicuously the floral organs of angiosperms. The inflated calyx syndrome (ICS) is a spectacular floral morphological novelty exhibited by a few genera of Solanaceae [3]. After pollination, Withania and Physalis sepal resume growth and give rise to a balloon-like structure, i.e. ICS or “Chinese lantern” encapsulating the berry [4]. The genus Withania consists of 11 species, which are worldwide in distribution. It displays a variety of inflated calyces ranging from the half open balloons of W. aristata and W. frutescens containing needle-like and teeth-like projections, respectively, to open fleshy “lanterns” in W. riebeckii, and completely closed semi-succulent papery lanterns in W. somnifera and W. coagulans [5]. By sharp contrast, Tubocapsicum features only a rudimentary calyx and is considered to be an “evolutionary loss mutant” of the ICS. This genus consists of two species, mainly endemic to humid regions of Eastern Asia. Phylogenetically, Withania and Tubocapsicum are sister genera to each other and are placed in the sub-tribe Withaninae along with 7 other genera [4].

The unique and single copy MADS-box gene MPF2—an ortholog of the potato gene STMADS16 - controls ICS formation in Physalis [6]. Previously, changes in its promoter region were shown to account for the expression of MPF2 in floral organs. However, in the tetraploid Withaniae, probably due to allotetraploidization, MPF2 is duplicated into MPF2-like-A and -B genes, of which only the former controls ICS formation in Withania. This functional divergence is a consequence of differences in gene expression potentially mediated by their divergent promoters [3].
It is generally believed that variations in gene expression are an important source of phenotypic diversity. The identification of regulatory changes underlying specific expression differences, however, is more difficult and hitherto, little progress has been made in connecting expression divergence with regulatory sequence variation. In Withania, the evidence for divergent expression of MPF2-like duplicates suggests an important control exerted by cis-regulatory regions, and prompted us to assess, if divergent elements in the promoter and intronic regions are responsible. Indeed, a conservation-based approach revealed that MPF2-like-A promoters lack an ARF (auxin response factor) binding site in the large 1st intron in the coding region whereas its homolog MPF2-like-B is devoid of a CArG-box near the transcriptional start site. This CArG-box is responsible for expression in sepals, which is essential for ICS formation. By contrast, the ARF binding element in the 1st large intron suppresses the expression in flower/calyx as evident from site-directed mutagenesis using reporter genes driven by MPF2-like cis-elements. Based on structural and functional characterization of MPF2-like cis-regulatory elements, we posit that adaptive evolution of MPF2-like genes was achieved at least in part by degenerative mutations in the core promoter CArG-box and the ARF binding site in the large 1st intron of the coding region. Moreover, our data provide insight into auxin signaling as an input pathway for MPF2-like genes.

**Results**

**Withania Calyx Cells Grow after Pollination**

Withania and Tubocapsicum though display contrasting phenotypes with regard to calyx inflation, are close relatives of Physalis and sister genera to each other. Therefore, in order to observe to what extent the calyx of Withania and Tubocapsicum increases in size after fertilization, we compared their flowering and fruiting calyces, i.e. before and after fertilization (Fig. 1A). The measurements revealed an exorbitant increase (4 to 5 times) in the fruiting calyx of Withania. By sharp contrast, the fertilizing calyx of Tubocapsicum was even smaller than the flowering calyx showing its degeneration after pollination (Fig. 1B). By analogy, Vassobia a closely related genus to Tubocapsicum also exhibited no post-fertilization calyx inflation. These data suggest that after pollination or during fruit development both the flowering and fruiting calyces of Withania can change in size and architecture. Figure 1C shows that post anthesis when fruits are developing; unlike in Withania where cells increase in size and become lobed, Tubocapsicum sepal cells are slightly stretched out but neither larger nor lobed. From this data we can infer that only in the Withania calyx the morphology of cells changes post fertilization.
**MPF2-like-B of Withania (WSB206) is not Expressed in the Calyx**

RNA was isolated from leaf and floral tissues of *Withania* and *Tubocapsicum*, and subjected to quantitative real-time RT-PCR (see Materials and Methods). Figure 1D demonstrates that all the MPF2-like genes are strongly expressed in leaves. The transcript signals of *Withania* MPF2-like-A (WSA206) and *Tubocapsicum* MPF2-like-B (TAB201) are detectable in all the floral organs, i.e. calyx, stamen, carpel and fruit. Conversely, the precocious expression of *Withania* MPF2-like-B (WSB206) is weak in stamens, carpels and fruits. Surprisingly, WSB206 transcript is absent in sepalas, which precludes its role in IGS formation. Hence, the expression of the Withanidae MPF2-like gene duplicates MPF2-like-A and -B have diverged. These genes have originated presumably after genome merger (alloetraploidization) in the Withanidae. Which gene was contributed by which progenitor would be exciting to elucidate.

Therefore, in order to get an indication about their hypothetical progenitors, the phylogeny of Physaleae (a tribe in Solanaceae that includes Physalinae, Withaninae, Iochrominae etc.) was established.

**Combined Expression Patterns of MPF2-like Duplicates of Withania Replicate their Hypothetical Progenitor**

We constructed a maximum likelihood phylogeny of MPF2-like genes using the PAUP4.0b10 software (Fig. 2A). Phylogenetic analysis clearly differentiated the MPF2-like genes into three distinct clades designated as MPF2-like-A, MPF2-like-B and MPF2-like. Multiple copies of the genes found in *Withania* clustered into their respective gene clades, i.e. MPF2-like-A and -B, respectively. MPF2 from *Physalis* was found in a sister position with respect to the MPF2-like-A subclade, while Iochrominae genes (V001, 1001 and Du001) occupied a basal position to both the MPF2-like-A and MPF2-like-B clades. These phylogenetic analyses allow us to speculate that for the alloetraploidization of MPF2-like genes of Withanidae, *Physalis* might be the contributor of MPF2-like-A and Iochrominae of MPF2-like-B, respectively. This inference is supported by their protein sequence similarities. The pairwise alignment showed that V001 and WSB206 (MPF2-like-B) are 98% similar. Therefore a closely related gene might have been the progenitor of the MPF2-like-B protein. Interestingly, MPF2 is 37% identical with both WSA206 (MPF2-like-A) and WSB206 (MPF2-like-B). As both WSA206 (MPF2-like-A) and MPF2 play an important role in calyx inflation, thus, MPF2 is clearly closely related to MPF2-like-A (WSA206) functionally [5] (Fig. S1).

One of the characteristics of divergence after duplication is that combined expression patterns of the duplicates should be reminiscent of their progenitors. Therefore expression patterns of *Physalis* and *Vassobia* were studied in leaf and calyx tissues (Fig. 2B). Quantitative real-time RT-PCR revealed that both MPF2 and V001 were strongly expressed in leaves like the Withanidae duplicates. Although weaker, the transcript signals of both genes were also detectable in the calyx [4]. Thus, combined expression of the duplicates (*Withania* MPF2-like-A and -B) seems to be equivalent to the expression of unduplicated homologs in the closely related species of *Physalis* and *Vassobia*. Clearly, these observations support the occurrence of expression divergence after the genome duplication in Withanidae and allude towards their hypothetical diploid progenitors.

Previously, promoter mutations were shown to be implicated in heterotopic expression of MPF2 of *Physalis* in floral organs in comparison with STMADS16 of *Solanum*, which is strictly vegetative in expression [6]. In most of the cases, expression divergence in paralogs observed after duplication, is because of modifications in cis-regulatory elements [7], [8]. Therefore, we decided to analyze the promoters and large 1st intron sequences of MPF2-like genes with various bioinformatic tools to identify the putative cis-elements responsible for their divergent expression.

**Phylogenetic Shadowing Reveals Reciprocal Loss of cis-Regulatory Elements in MPF2-like-A and -B Genes Post Duplication**

We aligned MPF2-like promoter sequences (more than 2 kb upstream of the translational start site) as well as sequences of the large 1st intron in the coding region (more than 2 kb downstream of translational start site) from various genera. MULAN and mVISTA analyses revealed 2 highly conserved homology blocks B1 and B2 in the promoter region and 3 blocks B3, B4 and B5 in the large 1st intron (Fig. 3A). The first conserved block (B1) in the promoter includes 5’ UTRs and the TSS and may be designated as the core promoter. ClustalW alignments were used to analyze, whether these conserved blocks contain functional regulatory motifs potentially responsible for divergent expression. Figure 3B demonstrates that five of these putative sequences did not contain documented transcription factor binding sites for MADS-box genes. Therefore we called them shadows S1–S5 (Fig. S2). Sequence analysis of entire ~2 kb MPF2-like promoters revealed different categories of MADS-domain binding motifs (CARG-boxes) including an N10-like (C(A/T)nG), CC(A/T)nGG, and CC(A/T)nG CARG motifs, (the subscript represent the number of A/T base pairs). These CARG motifs were frequently found throughout the MPF2-like promoter and 1st intron (Table 1 and Fig. S3). A CC(A/T)nG type CARG-box (CCATAAAAAAAG) in the vicinity of the TSS was identified in all the sequences except MPF2-like-B of *Withania* (WSB206). Besides this core-promoter CARG-box, which is a reported binding site for MADS-box proteins such as AG, AGI, SEP1, SEP4 and AGL15 in Arabidopsis [9], binding sites for homeodomain proteins (AtHBI, AtHBS), AGPi and B-binding protein were identified in all the MPF2-like-A and -B promoters. Interestingly, all these latter transcription factor-binding sites are located within the 5’-UTR region, while the CARG-box is found upstream of the TSS at fairly constant distance (approximately 10 to 12 helical turns of the DNA).

Remarkably, in the 1st conserved block (B3) of the large 1st intron in the coding region, a 200 bp highly conserved region was identified in all the MPF2-like sequences at variable positions. Within this 200 bp conserved stretch of DNA, ClustalW alignment revealed an auxin response factor (ARF) binding element (ARE; cgtgcttc) to be conserved in all the sequences except MPF2-like-A (WSA206) of *Withania* (WSB206). Besides this core-promoter CARG-box, which is a reported binding site for MADS-box proteins such as AG, AGI, SEP1, SEP4 and AGL15 in Arabidopsis [9], binding sites for homeodomain proteins (AtHBI, AtHBS), AGPi and B-binding protein were identified in all the MPF2-like-A and -B promoters. Interestingly, all these latter transcription factor-binding sites are located within the 5’-UTR region, while the CARG-box is found upstream of the TSS at fairly constant distance (approximately 10 to 12 helical turns of the DNA).

In conclusion, the data show the loss of an otherwise conserved core-promoter CARG-box in MPF2-like-B and of an ARF binding site in the large 1st intron of MPF2-like-A from *Withania*, respectively. The reciprocal loss of cis-regulatory elements is a characteristic of divergence after duplication. Remarkably, both the *Physalis* and *Vassobia* MPF2-like genes - the presumed progenitors of MPF2-like-A and -B genes, respectively - contain the core-promoter CARG-box as well as the ARE in the large 1st intron.

**Loss of Core-promoter CARG-box Blocks Expression of MPF2-like-B Genes in Sepals**

Using an in vivo reporter assay the regulatory potential of the MPF2-like core promoter CARG-box was evaluated through site-directed mutagenesis and motif swapping. For this purpose chimeric promoter-reporter constructs containing normal and
Figure 2. Phylogenetic reconstruction of *MPF2-like* genes and expression patterns of their progenitors. A) A maximum likelihood tree of *MPF2-like* sequences from various Solanaceous species is established using *STMADS16* as an out-group. The ML tree reconstruction is carried out with PAUP 4.0b10. The robustness of the tree structure is evaluated by 1000 replicates of bootstrap searches using maximum parsimony (MP) and maximum likelihood (ML) in PAUP and Bayesian posterior probability is indicated as *. The multiple sequences for a gene are also indicated. B) The expression patterns of *MPF2-like* duplicates replicate their subsumed progenitors. RNA isolated from leaves and sepals of *Physalis* and *Vassobia* was subjected to real-time RT-PCR analysis. The values given for leaf (empty bar) and calyx (light grey bar) represent the relative expression normalized with respect to 18 S rRNA. Error bars indicate the standard deviation.
doi:10.1371/journal.pone.0042781.g002
Figure 3. Bioinformatics analysis and phylogenetic shadowing of MPF2-like regulatory elements. A) Mulan analysis of MPF2-like promoters and large 1st introns including MADS-box. Two conserved blocks (B1, B2) are identified in the promoter region and three (B3–B5) in the large 1st intron in the coding region. The first conserved block (B1) in the promoter region contains a CArG-box, which is present in all the sequence except MPF2-like-B of Withania (WSB206) and STMADS16. This CArG-box is a binding site for MADS-domain proteins such as AG, SEP1, SEP4, AGL1 and AGL15. The wild type sequence (in the dotted box) of this CArG-box is shown for WSA206 gene and mutated one for WSB206. In the large 1st intron first conserved block (B3; Fig. S4) contains an ARF binding site at variable positions in all the MPF2-like sequences except MPF2-like-A of Withania (WSA206). The normal ARF binding site (in the dotted box) is shown for WSB206, while the mutated one for WSA206 is also depicted. A rough scale is indicated. The position of translational start site is marked as +1. B) ClustalW multiple alignment of the core promoter conserved block (B1) of MPF2-like promoter sequences (~350 bp to ~650 bp). Five conserved sequences stretches are identified (Fig. S3) and called as shadows 1 to 5 (S1 to S5). Only S3 to S5 are shown in this alignment. Backward arrow marks the putative transcription start site. Dotted red box encloses the conserved CArG-box sequence.

doi:10.1371/journal.pone.0042781.g003
mutated CArG-boxes were transformed into Arabidopsis stably (YFP and GUS were used as reporter genes). The transformation of regulatory elements in heterologus system is now routine particularly in animals. In animal system the human regulatory elements have been successfully tested in zebra fish [11], which are evolutionarily distant relatives (450 million years: [8], [12]. It is speculated that evolutionary distance between Arabidopsis and Withania might falls within this limit as they are different at family level. Figure 4A depicts that very strong signals are detected in the nucleus of Arabidopsis sepals when transformed with WSA206*YFP, while YFP signals are diminished if the CArG-box is mutated (WSA206*YFP	1*YFP). Remarkably, the introduction of a CArG-box in WSB206*YFP plants causes a dramatic increase in YFP signal, while originally WSB206*YFP lack or show little intensity in sepals. However, the native MPF2-like-B promoter lacking the CArG motif drives expression in sepals. Similarly, the signal strength is weaker in case of plants transformed with mutated TAB201PS575*M:YFP constructs, lacking an intact CArG-motif. The role of this CArG-box in controlling MPF2-like expression in sepals was further validated by stable promoter-GUS analysis in transgenic Arabidopsis (Fig. 4b). In case of WSA206*YFP	1:GUS plants the GUS expression was strongly detectable in young sepals but only weak if the CArG-box was mutated (WSA206*YFP	1*YFP). On the other hand, introduction of a CArG-box in WSB206*YFP	1:GUS rescued the expression in young sepals, while originally WSB206*YFP	1:GUS plants showed little or no expression in the sepals. A slight decrease in GUS signals in the case of TAB201PS575*M:GUS (lacking the CArG-box) was also observed in comparison with TAB201PS575*:GUS that exhibited stronger signals. These data clearly suggest that the core-promoter CArG-box is crucial for activating expression of MPF2-like genes in sepals. This is supported by the lack of expression of MPF2-like-B (not containing a CArG-box) in the sepals of native plant (Fig. 1D).

Given our results it is paramount that effects of the entire promoter, i.e. more than ~2 kb upstream of the translational start site, on the expression of MPF2-like genes should be delineated.

### MPF2-like-B Promoter Drives Strong GUS Expression in Transgenic Arabidopsis Flowers Including Sepals

Three constructs in which GUS/YFP reporter expression is driven by the more than 2 kb long promoter of WSA206, WSB206, and TAB201, respectively, were used to transform Nicotiana transiently (reporter gene YFP) and Arabidopsis stably (reporter gene GUS). Figure 5A illustrates that strong GUS signals are detected in all floral organs including sepals as well as in leaves for the promoter of WSA206 (MPF2-like-A of Withania). These expression patterns are reminiscent of the native expression of WSA206 in Withania. In the case of WSB206 (MPF2-like-B of Withania), surprisingly, we detected very strong GUS signals in almost all the floral organs including sepals. Furthermore, when the GUS reporter was replaced with YFP and constructs were used to transiently transform Nicotiana, the YFP signals were equally strong for both the WSA206 and WSB206 promoters. This is an extraordinary result, which is in contrast with the native expression of MPF2-like-B. It is already mentioned that MPF2-like-B expression is not present in sepals. How MPF2-like-B promoters stimulate GUS expression in the heterologous systems to a stronger extent in flowers/sepals compared to the native background is an interesting question. One explanation could be that there are functional regulatory elements in the MPF2-like-B introns that are suppressing the effect of activator elements in the promoter responsible for floral expression including sepals. As most of the regulatory elements reported are enriched near the TSS [13], and as the large 1st intron in the coding region of MPF2-like genes is located not very far from the TSS, we embarked on analyzing the effect of this intron on expression.

### Large 1st Intron Suppresses the Expression of MPF2-like-B Genes of Withania (WSB206)

To investigate whether the complete large 1st intron affects the expression of MPF2-like genes we used two types of constructs (Fig. S5). In the first type of constructs only promoter region (more than ~2 kb upstream of the translational start site) was fused with the coding sequences of GUS and YFP reporter genes, respectively (MPF2-like	1*:GUS/YFP). In contrast, in the second type of constructs the complete 1st intron was additionally attached to the promoter sequence (MPF2-like	1*:GUS/YFP). Figure 5ABC shows that YFP signal intensity is equal in almost all the MPF2-like-B promoter constructs except for the promoter of MPF2-like-B of Withania (WSB206) where it is slightly lower. However, YFP is barely detectable if the large 1st intron is combined with the promoter for WSB206. In the other cases (WSA206*YFP and TAB201*YFP) no dramatic differences in YFP expression could be observed. These results were also confirmed by histological assay using GUS as a reporter gene (Fig. 6ABC). When the complete 1st intron was attached to the WSA206 promoter (WSA206*PS575*:GUS) no noticeable change in the GUS expression pattern was observed and conversely, there was a dramatic reduction in GUS expression in case of combining WSB206 1st intron and promoter (WSB206*PS575*:GUS). This reduction in GUS expression was observed in most of the floral

### Table 1. Specifications and positions of different cis-regulatory elements of the MPF2-like genes.

| MPF2-like gene | Promoter length (bp) | Transcription start site | Position of CArG-box in promoter | Intron length (bp) | Position of CArG-box in the large 1st intron | Position of ARE |
|----------------|----------------------|--------------------------|-----------------------------------|-------------------|---------------------------------------------|-----------------|
| WSA206         | 2606                 | ~320                     | ~430, ~2050, ~2550                | 2544              | 474, 2092                                    | 694*            |
| WSB206         | 3348                 | ~415                     | 4600*+M, ~698, ~1754              | 1831              | 1133, 1627                                   | 430             |
| TAB201         | 2308                 | ~436                     | ~560, 1290, ~1557, ~1851          | 3704              | 697, 1493, 2010, 2888, 2995, 3048            | 1214            |
| MPF2           | 2237                 | ~472                     | ~580, 1488, ~1784, ~2091          | 2159              | 549, 855, 1429                               | 786             |
| STMADS15       | 2588                 | ~459                     | ~32, ~249, ~4600*+M, ~937, ~1226, | 2539              | 234, 314, 558, 717, 921, 1296                | 1187            |
|                 |                      |                          | ~2147                           |                   |                                             |                 |
| V001_1         | 2056                 | ~594                     | ~440, ~836, ~1285                | 3386              | 186, 812, 945, 2377, 2539, 2781, 2896, 3311 | 1344            |

The position of translational start site is +1. *M indicates the position of lost or mutated CArG-box in WSB206 and STMADS15 core promoters. whereas mutated ARF binding site in the large 1st intron of WSA206 is indicated as +M. doi:10.1371/journal.pone.0042781.t001

Large 1st Intron Suppresses the Expression of MPF2-like-B Genes of Withania (WSB206)
organs, particularly in sepals. This is congruent with the expression of \textit{WSB206} in \textit{Withania}.

From these data it can be deduced that the large 1st intron suppresses the expression of \textit{MPF2-like-B} genes of \textit{Withania} (\textit{WSB206}). Phylogenetic footprinting already indicated that an ARE is present in the 200 bp conserved region of all the \textit{MPF2-like} large 1st introns except \textit{MPF2-like-A} of \textit{Withania} (\textit{WSA206}). The potential activity of an element can be missed if it is not analyzed in isolation. It will be interesting to study whether this ARE is responsible for suppressing the expression of \textit{MPF2-like} genes, particularly the \textit{MPF2-like-B} of \textit{Withania}. For this purpose, reporter assay using site-directed mutagenesis and motif swapping of these elements was undertaken.

**ARF Binding Site Suppresses the Expression Activated by CArG-box**

The effects of the ARE and the CArG-box on tissue specific expression of the GUS reporter gene in transgenic \textit{Arabidopsis} were tested for promoters of \textit{MPF2-like} genes (\textit{WSA206}, \textit{WSB206} and \textit{TAB201}). For this purpose a total of 18 constructs - 6 constructs for each of the three promoters (\textit{WSA206}, \textit{WSB206} and \textit{TAB201}) - were employed. Figure 6ABC (left panels) shows a heat map for qualitative expression observed for 30 plants for each construct. It depicts that in the case of the \textit{WSL206} long promoter construct (\textit{WSA206I2373:PL2606:GUS}) very strong GUS expression in both the leaf and floral organs including sepals was observed. Similarly, when combining the complete large 1st intron with this promoter (\textit{WSA206I2373:PL2606:GUS}), no deviation from this GUS expression pattern was visible. To investigate, whether the absence or presence of the mentioned ARE influences GUS expression, we introduced this site and used a short promoter containing the core-promoter CArG-box (\textit{WSA206I712:PS445*:GUS}). Interestingly, GUS signals became weak indicating that the ARE is suppressing the GUS expression in flowers, especially in sepal tissue. On the other hand, removal of the CArG-box (\textit{WSA206I712:PS445*:GUS}) strongly reduced the expression in sepals even if the ARE was not present. Introduction of an ARE (\textit{WSA206I712:PS445*M:GUS}), however, further reduced the expression below the detection level in sepals supporting a role of the ARE as a suppressor element with an antagonistic effect on expression compared to the activating CArG-box.

\textit{WSB206} cis-elements can help to improve our understanding of the interplay between the CArG-box and ARE. As already mentioned, the \textit{WSB206} large 1st intron contains a native ARE element but lacks the core promoter CArG-box. GUS expression was very strong in all the organs if only the long promoter was fused with the GUS cds (\textit{WSB206PL3129:GUS}) but after attaching the complete 1st intron (\textit{WSB206I1827:PL3129:GUS}) there was a dramatic decrease in GUS signal intensity and expression in sepals became invisible. In order to further characterize the interaction between CArG-box and ARE, a short promoter without the CArG-box was combined with the intronic region (\textit{WSB206I458:PS471*:GUS}). The resulting reduction in GUS expression in sepals could be a combined effect of the absence of the CArG-box and the presence of the ARE. However, if a CArG-box was introduced to the promoter and the ARE was mutated, GUS expression became very strong in all the floral organs including sepals. \textit{TAB201} cis-elements...
exhibited results not very different from WSA206 with slight variations in GUS expression levels. These results are further supported by quantitative GUS expression analysis. For this purpose only calyx tissue was selected for determining the GUS activity by using fluorometric analysis. Figure 7A shows that GUS activity is similar to qualitative GUS expression for almost all the constructs for WSA206, WSB206 and TAB201 cis-elements with subtle variations.

Figure 5. MPF2-like promoter activates the expression of YFP/GUS, while large 1st intron suppresses it. A) MPF2-like long promoter was attached with GUS and YFP reporter genes and expressed stably and transiently in Arabidopsis and Nicotiana, respectively. The GUS signals driven by Withania MPF2-like-A and -B promoter are equally strong in the inflorescence though native expression for MPF2-like-B is barely detectable. Similarly YFP signals are very strong in both types of promoters. The scale bar represents 4 mm and 30 µm for GUS and YFP pictures, respectively. B) A transient YFP expression assay was performed under the control of MPF2-like long promoter and complete large 1st intron (MPF2-like<sup>PL</sup>:YFP; MPF2-like<sup>I:PL</sup>:YFP). Three days after infiltration, leaves of N. benthamiana were scanned under Leica LCS SP2 AOBS<sup>R</sup>, Confocal Laser Scanning Microscope (CLSM) for YFP signal detection. At least 10 images selected randomly to quantify the luminescence with the Leica software LCS Lite. Promoter strength was determined as the relative intensity of YFP fluorescence of nuclear area of MPF2-like promoter YFP constructs in comparison with nuclear YFP fluorescence intensity of a 35 S CaMV promoter YFP construct. In the case of WSB206-like<sup>I:PL</sup>:YFP there is a tremendous decrease in YFP signal intensity after attaching MPF2-like large 1st intron with promoter region. The error bars represent the standard deviation. C) Confocal images showing YFP expression driven under MPF2-like cis-elements in the nuclei of the transiently transformed Nicotiana leaves. The scale bar represents 30 nm.

doi:10.1371/journal.pone.0042781.g005
The interplay between ARE and CArG-box was also tested transiently in *Nicotiana* using the YFP reporter gene driven by WSA206, WSB06, TAB201, C promoter and large 1st intron elements (Fig. S6). The results demonstrated that loss of the ARE relaxed constraints on the expression of the YFP reporter exerted by MPF2-like promoter or CArG-box elements.

**Application of Exogenous Auxin Suppresses the Expression of GUS Reporter Genes**

Figure 6ABC (right panels) shows that transgenic *Arabidopsis* plants harboring only promoter-GUS constructs failed to respond to external auxin and exhibited no change in their GUS expression patterns. Furthermore, all the transgenic plants lacking an ARE in the construct including intronic sequence also showed no response to exogenous auxin. However, in cases where an ARE exists naturally or was introduced artificially, suppression of GUS signal intensity was observed. Hence, without auxin treatment the AREs mediate a suppression of GUS expression and this effect is more pronounced upon auxin treatment. In other words, an increase in auxin supply leads to an increased suppression of GUS expression. These results are supported by quantitative analysis of GUS activity in sepal tissues of the transgenic *Arabidopsis* plants after auxin treatment (Fig. 7B). This is a very interesting result, because it indicates an influence of auxin signaling on ICS development via mediating divergent expression of MPF2-like genes.

These cumulative data allow us to conclude that the ARE is an active suppressor of MPF2-like expression in sepal, which in turn is activated by the CArG-box. Thus reciprocal loss of core promoter CArG-box and intronic ARE in MPF2-like-B and -A genes, respectively, is responsible for expression divergence of MPF2-like genes and eventually influences ICS formation.

**Materials and Methods**

**Plant Materials**

A total of 13 accessions of Withania including 5 species (W. aristata, W. frutescens, W. riebeckii, W. coagulans and W. somnifera) and 1 accession each for Tubocapsicum anomalum, Vassobia breviflora, and Physalis floridana were grown in...
Figure 7. Interplay of auxin response factor element (ARE) and core promoter CArG-box with regard to expression of MPF2-like genes in sepals. A) Quantification of GUS enzymatic activity in the calyx of transgenic *Arabidopsis* by fluorometric method. Expression of the GUS gene is driven by MPF2-like cis-elements. B) Quantification of GUS enzymatic activity in the calyx of transgenic *Arabidopsis* after auxin spray by fluorometric method. The expression of the GUS gene is driven by MPF2-like cis-elements.

doi:10.1371/journal.pone.0042781.g007
Measurement of Calyx Inflation and Scanning Electron Microscopy (SEM) of Withania and Tubocapsicum Sepals

In order to record the degree of inflation of the calyx after pollination, the length and width of flowering and fruiting calyces (10 each) of W. somnifera, T. anomalum, P. floridana and V. breviflora were measured using a Vernier scale. Data were statistically analyzed and inflation degree of the calyx was determined. Furthermore, morphological differences in the sepal cells of W. somnifera and T. anomalum were investigated by digital scanning electron microscopy (DSM940, Zeiss, Oberkochen, Germany; [4]).

Expression Analysis

Expression of MPF2-like genes in Withania, Tubocapsicum, Vassobia and Physalis was investigated by real-time RT-PCR using the Bio-Rad iQ5 Real Time PCR Detection System. Total RNA extraction and first strand cDNA synthesis were performed [4]. A single forward primer flanking the conserved region coding for K- and C-domain was designed for Arabidopsis thaliana (ecotype Columbia) and Nicotiana benthamiana were grown under standard conditions. A few selected lines of transgenic Arabidopsis plants were also grown at the National Institute for Genomics and Advanced Biotechnology, Islamabad, Pakistan.

Promoter and Large 1st Intron Sequence Isolation

A total of 16 MPF2-like promoter sequences more than 2 kb in length (genomic sequence upstream of the MADS-box comprising 5’UTR) were isolated from various accessions of Withania, Tubocapsicum and Vassobia by RAGE (rapid amplification of genomic DNA ends) using the Universal Genome Walker Kit (Clontech; [5]). Their corresponding MPF2-like large 1st intron sequences were isolated by PCR amplification by using primers designed to bind to promoter region and second exon of MPF2-like genes (Table 2) with Takara LA Taq polymerase.

Phylogenetic Analyses and Transcription Factor Binding Site Identification

Maximum likelihood (ML) tree reconstruction was carried out with PAUP 4.0b10 [4], [15]. To evaluate the robustness of the tree structure, 1000 replicates of bootstrap searches were performed using maximum parsimony (MP) and maximum likelihood (ML) in PAUP and Bayesian posterior probability was calculated [16]. MPF2-like cis-regulatory sequences were assembled using MacVector™ 7.2.3 (Accelrys Inc.) ggc/Wisconsin Package (University of Wisconsin) and AssemblyLIGN™ (Oxford Molecular Group Plc.) softwares. The ClustalW program in MacVector™ and mVista Shuffle-LAGAN were used to create multiple alignments (http://genome.lbl.gov/vista). In silico mapping of conserved transcription factor binding sites in MPF2-like promoters was carried out with TFSEARCH and PROSCAN.

Table 2. Isolation of MPF2-like promoter and large 1st intron sequences from five Withania, one Tubocapsicum and one species of Vassobia.

| Sr. Nr. | Accession | Species     | Gene    | Promoter length (bp) | 1st intron length (bp) |
|--------|-----------|-------------|---------|----------------------|------------------------|
| 1      | W001      | W. aristata | WAA201  | 2667                 | 2373                   |
| 2      | W001      | W. aristata | WAB201  | 3294                 | 1827                   |
| 3      | W002      | W. coagulans| WCA202  | 2609                 | 2541                   |
| 4      | W002      | W. coagulans| WCB202  | 2978                 | 1865                   |
| 5      | W004      | W. riebeckii| WRA204  | 2645                 | 2621                   |
| 6      | W004      | W. riebeckii| WRB204  | 3015                 | 1873                   |
| 7      | W006      | W. somnifera| WSA206  | 2606                 | 2479                   |
| 8      | W006      | W. somnifera| WSB206  | 3348                 | 1945                   |
| 9      | W007      | W. somnifera| WSA207  | 2645                 | 2589                   |
| 10     | W007      | W. somnifera| WSB207  | 3919                 | 1870                   |
| 11     | W013      | W. frutescens| WFA213  | 2645                 | 2452                   |
| 12     | W013      | W. frutescens| WFB213  | 2970                 | 2001                   |
| 13     | T001      | T. anomalum | TAB201  | 2308                 | 3704                   |
| 14     | T002      | T. anomalum | TAB202  | 2357                 | –                      |
| 15     | V001      | V. breviflora| V001_1  | 2107                 | 3386                   |
| 16     | V001      | V. breviflora| V001_2  | 2056                 | –                      |
| 17     | P106a     | P. floridana | MPF2    | 2238                 | 2165                   |
| 18     | S032a     | S. tuberosum | STMADS16 | 2588                 | 2539                   |

*Obtained from database; -, Could not be isolated.
doi:10.1371/journal.pone.0042781.t002
and large 1st intron sequences was carried out using MULAN (Multiple-sequence local alignment; [17]) and MultiTF (Local multiple sequence alignments, transcription factor binding sites). Putative cis-regulatory elements were predicted using the TRANSFAC database of transcription factors (http://www.gene-regulation.com/pub/databases.html).

**Promoter:GUS/YFP and Intront:promoter:GUS/YFP Vector Construction**

In order to investigate the effects of promoter and large 1st intron elements on MPF2-like expression divergence, promoter:GUS/YFP and intron:promoter:GUS/YFP expression analyses were carried out in *Nicotiana* and *Arabidopsis* transiently and stably, respectively. Initially, a total of 16 constructs - 8 for each of the two reporters for four genes *WSA206*, *WSB206*, *TAB201*, and MPF2 were generated. The 1st construct type (e.g. MPF2-like*P*MPF2-GUS/YFP) for each gene contains more than 2 kb of upstream genomic sequence. The 2nd construct type (e.g. MPF2-like*P*MPF2-GUS/YFP) harbors more than 2 kb of genomic region upstream of the translational start site fused with the complete 1st intron (from 2 to 4 kb) containing both the 5’ splice donor site of exon 1 (MADS-domain) and 3’ splice acceptor site of exon 2 (I-domain). This construct lacks the 180 bp of the MADS-box, and removing the ATG start codon prevents the expression of a truncated MPF2-like protein (containing only the MADS DNA binding domain). In order to decipher the role of the identified CArG-box close to the TSS, two special types of promoter constructs were generated in such a way that the promoter was considerably shortened and the naturally occurring CArG-box was included but in the other three types, CArG-box and ARE were mutated or introduced through site-directed mutagenesis where present or absent, respectively. Motif swapping of CArG-box and ARE was also performed between MPF2-like-A and -B cis-elements. More details about the pursued scheme of promoter:reporter and intron:promoter:reporter construct making are given in figure S5.

Gateway (Invitrogen, http://www.invitrogen.com) expression vector pXCG-mYFP was used as a backbone to generate promoter-YFP and promoter-intron-YFP constructs whereas pGVT-bar was employed to make promoter-GUS and promoter-intron-GUS constructs. For making GUS constructs, both the MPF2-like regulatory elements and pGVT-bar vector were digested with *SmaI*, *XmaI*, *HindIII*, *StuI* and *XbaI* restriction enzymes. After purification, ligation was done using T4 DNA ligase from NEB (New England Biolabs) at 15°C overnight. Sequenced plasmids containing inserts in the correct orientation and coding frame were transformed into the *A. tumefaciens* strain GV3101 (pMP90RK) for transient infiltration in *Nicotiana* and stable expression in *Arabidopsis*.

**Agroinfiltration in *Nicotiana* and YFP Signal Quantification**

For transient infiltration, liquid YEB medium containing appropriate antibiotics was inoculated with *Agrobacterium* strain GV3101 (pMP90RK) harboring constructs with MPF2-like cis-elements fused with the YFP reporter along with constructs expressing the p19 protein [18]. These cultures were mixed before infiltration into *Nicotiana* leaves as described by [19]. The upper two healthy leaves were co-infiltrated using a 1 ml plastic syringe. Three days after infiltration, leaves of *N. benthamiana* were scanned using a Leica LCS SP2 AOB5 Confocal Laser Scanning Microscope (CLSM) for YFP signal detection. At least 10 images showing YFP signals in the nuclei were selected randomly to quantify the luminescence with the Leica software LCS Lite (Leica Microsystems, Wetzlar, Germany). Promoter strength was determined as the relative intensity of YFP fluorescence in nuclei (mean of digital brightness values per nuclear area) for MPF2-like cis-element-YFP constructs in comparison with nuclear YFP fluorescence intensity of a 35 S promoter-YFP construct, which is the most intense signal, was set to 100%. The images were taken with the same hardware values for PMT gain, and offset and zoom.

**Transgenic Plant Production**

Transformation of all the MPF2-like promoter: GUS and intron:promoter:GUS constructs along with the empty pGVT-bar vector into *Arabidopsis* was mediated by the GV3101 (pMP90RG) strain of *A. tumefaciens*. For plant transformation, a floral-dipping protocol was followed [20]. After seed germination, the transgenic plants were selected by spraying 0.15% Basta (Bayer Crop Science, Monheim) solution twice. T3 homozygous single copy transgenic lines were selected for reporter gene analysis.

**GUS Histochemical and Fluorometric Analysis**

Histochemical GUS analysis was carried out to observe expression patterns of MPF2-like genes in different tissues of the transgenic *Arabidopsis* plants. Inflorescence, rosette leaves, siliques, small seedlings and entire plants of the transgenic *Arabidopsis* were stained with GUS staining solution [21]. More than 30 independent transgenic lines for each construct were examined for GUS activity. Calyx tissue was selected for GUS protein quantification. Quantitative assays were performed with 4-MUG as the substrate. Calyxes of the transgenic plants before and after auxin spraying were harvested and homogenized in a 50 mM sodium phosphate lysis buffer (pH 7.0) containing 0.1% Triton X-100, 0.1% sodium lauryl sarcosine, 10 mM EDTA and 10 mM β-mercaptoethanol. After centrifugation at 5000 rpm for 10 min, the GUS activity was evaluated using the resultant supernatant at 37°C in lysis buffer containing 1 mM 4-MUG. The reaction was terminated by adding 200 mM Na2CO3 to a final concentration of 160 mM. Fluorescence was quantified using a FLUOstar Galaxy multi-well plate reader with the excitation and the emission filters set at 365 nm and 455 nm, respectively. The GUS protein concentration was determined as described [22].

**DNA Sequencing**

For all the PCR products, plasmids and constructs DNA sequences were determined by the MIP1Z DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry.
Discussion

The Core Promoter CARG-motif is a Key Regulatory Element for MPF2-like Expression in the Calyx

MADS-domain proteins regulate expression of target genes by binding to CARG motifs in their promoter regions [23–25]. Most MADS-domain proteins prefer the so-called serum response element (SRE or SRF)-type CARG box, which has the consensus CC(A/T)6GG [26], [27]. Another closely related sequence motif is the MEF2-type CARG-box having the general consensus C(A/T)₄G, but being usually more strictly defined as C(TA)(A/T)₃T,G. However, plant MADS-domain proteins often show relatively broad DNA-binding preferences, recognizing SRF- and MEF2-type CARG-boxes in the promoter of Arabidopsis as well as intermediate types of motifs [28].

Previously it was hypothesized that absence of two N10-like CARG-boxes in the promoter of MPF2 of Physalis in comparison with STMA116 of Solanum accounts for the heterotopic expression of this gene in floral organs. This shift in vegetative to floral expression through cis-mutations ultimately leads to the formation of an inflated calyx [6]. However, this is merely an assumption that lacks functional evidence. In this study we have shown that different types of CARG-boxes (SRF and MEF2) frequently occur throughout 2 kb MPF2-like promoters and large 1st introns (see “Results” section). Thus, presence of CARG-motifs alone is not sufficient to predict floral expression and targets of MADS-domain transcription factors [29]. Given the large number of MADS transcription factors and considering their divergent functions in plant development, it is imperative to elucidate their target-sequence specificity, for instance, what are the different types of CARG-boxes and what are the cis-regulatory positions that can be preferable sites for binding of MADS-domain interactors. Genome-wide analysis of Arabidopsis SEP3 revealed the highest enrichment of SRF- and intermediate types of CARG-boxes in the promoter region 1 kb upstream close to the TSS [13]. Hence it is suggested that recognition sites are not randomly distributed in the genome. Phylogenetic shadowing of MPF2-like short promoters identified an intermediate type of CARG-box with consensus sequence CCATAAAAAG in the close proximity of the TSS. The putative TSS of MPF2-like genes based on the sequence of the longest cDNA is −320 bp, −415 bp, −436 bp, −472 bp and −459 bp for WS206, WSB206, TAB201, MPF2 and STMA116, respectively (Table 2). Hence, the position of our identified CARG-box in the regulatory region immediately upstream of the TSS in the core promoter of MPF2-like genes supports it to be functionally active.

The MPF2-like core promoter CARG-box is a potential binding site for the MADS-domain proteins AG, SEP1, SEP4, AGL3 and AGL15. MADS-domain proteins bind CARG-boxes as evident from electrophoretic mobility shift assays and chromatin immuno-precipitation experiments [30]. AG specifies floral meristem, carpel and stamen identity [31], [32]. The function of AGL2 and AGL3 is to define floral organ identity as well as to determine floral meristems [33], [34]. SEP4 has redundant functions in all the floral organs [33], while AGL15 plays a role in post-germinative development [35]. Hence the observed GUS expression patterns in floral tissue of transgenic Arabidopsis could be caused by MADS-domain proteins binding to the conserved CARG-box and up-regulating expression in floral organs. This is supported by recapitulation of MPF2-like transcript expression in floral organs of native plants. The core promoter CARG-box is found in all the MPF2-like promoters except MPF2-like-B of Withania (WS206) and STMA116 of Solanum. Notably, both these genes are not expressed in sepals. Moreover, promoter-reporter experiments comprising site-directed mutagenesis and motif swapping confirmed that the core promoter CARG-box controls expression of MPF2-like genes in sepals. As expression of MPF2-like genes is essential for ICS formation, the core promoter CARG-box seems to be the key regulatory element for expression divergence of MPF2-like genes in the evolution of ICS.

Auxin Signalling as an Input Pathway for MPF2-like Genes

Recently it was reported that ARF genes are not only targets of SEP3 MADS-box genes but also can be co-regulators [13]. ARF proteins recognize AREs in the intronic regions of target genes and modulate their expression. In Arabidopsis around 23 different types of ARFs have been identified so far. The amino acid contents in the variable middle region determine whether a particular ARF functions as a repressor or activator [36], [37]. Glutamine (Q) rich ARFs such as ARF5 and ARF7 activate transcription [12], [38]. All the ARFs bind to an auxin-response element (ARE), having the consensus sequence TGTGTC, in the regulatory regions of auxin response genes in an auxin-dependent manner [39]. In this study, the type of binding site
identified in MPF2-like 1st intron is characteristic for an ARF with a Q-rich domain. Hence it might activate the expression. However, another class of proteins, Aux/IAA proteins, interact with ARFs and act as negative regulators of transcription [40]. Studying GUS expression patterns in Arabidopsis by site-directed mutagenesis and motif swapping of MPF2-like cis-elements (promoters with a CArG-box and intron with an ARE element) suggested that ARFs bind the ARE and suppress the expression of MPF2-like genes in flowers including sepals. The loss of the ARE in MPF2-like-A relaxes the expression constraints particularly on sepals. These results are supported by the native expression patterns. Interestingly, upon auxin treatment, MPF2-like expression is even more repressed. This repression may be either due to direct binding of ARFs to the ARE or there might exist an indirect pathway involving interaction of ARE with Aux/IAA. It is important to mention that Arabidopsis is evolutionarily distant from Withania but does possess AGL24 and SVP as orthologs of Solanaceous MPF2-like genes. In the promoter and large 1st intron of AGL24 and SVP no ARF binding site exist (data not shown). Interestingly, there is no CArG-box near the transcriptional start site of AGL24, and in SVP it is present far away from the TSS. As SVP and AGL24 lack core-promoter CArG-boxes, no strong expression in sepal is expected. Consequently, due to the absence of a CArG-box in Arabidopsis orthologs (SVP and AGL24) the roles of AREs in affecting flower expression cannot be explored. Furthermore in Withania and Tubocapsicum, cytokinin and auxin treatment does not alter the expression pattern at the transcript level (unpublished results).

The minima and maxima of endogenous auxin concentration influence flower morphogenesis [41]. The effects of auxin are thought to be dependent on its concentration; with high and low doses eliciting different responses [42], [43]. These interactions can be explained through a simple model. Figure 8 shows that MPF2-like expression activated via the core promoter CArG-box is suppressed through the ARE in the large 1st intron. At lower concentration of auxin (endogenous in this case) ARFs might be bound by Aux/IAA proteins. This complex is unable to bind the ARE in the MPF2-like intron thereby allowing the transcription of MPF2-like genes to be activated by the core promoter CArG-box. An increase in auxin concentration (exogenous in this case) may cause the proteasome-mediated degradation of Aux/IAAs, as a result of which free ARF is able to bind the ARE in MPF2-like large 1st intron. Hence, the ARF binding site acts as suppressor of MPF2-like expression particularly in sepals in contrast with the CArG-box, which is an activator element. This interaction is, however, not simple and might involve complex regulatory networks. The unravelling of the exact mechanism of transcriptional control through the ARE depends upon the availability of mutant ARF and MPF2-like lines. Nevertheless, this is the first time that a role of auxin in ICS development has been suggested at transcription level.

Reciprocal Loss of cis-regulatory Elements in the Evolution of MPF2-like Genes

According to the duplication-degeneration-complementation (DDC) model, sub-functionalization of duplicated genes predicts the reciprocal loss of cis-regulatory elements in each paralog by degenerative mutations such that paralog’s combined expression patterns recapitulate the expression patterns of the ancestor [44], [45]. A few examples that invoke the sub-functionalization of genes in plants exist but verification of reciprocal loss of cis-regulatory elements is lacking in plants [46–49]. There are a few examples of gene pairs such as pax6a and pax6b in vertebrates [50], which are consistent with the DDC model. The story of hoxb5a and hoxb5b is complex; there is no evidence for a simple loss of regulatory elements and the interactions between regulatory elements are not well understood [51]. Similarly, the sub-functionalization of AGL6 and AGL13 MADS-box genes is also based on the complex interactions of enhancer and silencer elements [46]. The evolution of MPF2-like genes in Withania seems to be an example of sub- and/or neofunctionalization according to the DDC model. The duplicated MPF2-like-A and -B genes show degenerative mutations in the promoter and large 1st intron, respectively. The combined expression of the homologs and their contribution to function seem to be equivalent to the expression and functional contribution of unduplicated orthologs in closely related species.

Based on structural and functional characterization, the role of cis-regulatory elements in the evolutionary history of MPF2-like genes is summarized in Figure 9A. It is presumed that progenitors of MPF2-like genes contained a core promoter CArG-box (an asterisk marks CArG-box in MPF2-like promoters) near the TSS and an ARE (indicated by crossed circle) in the conserved block of the large 1st intron. After, hybridization and allotetraploidization, both the lineages followed a normal speciation procedure and acquired their evolutionary fates. In the lineage Withania, MPF2-like-A retained the CArG-box and lost the ARE site. Consequently, MPF2-like-A of (WASA206) Withania was expressed in sepals due to the presence of CArG-box and was not repressed due to absence of an ARF binding site. Hence, MPF2-like-A could generate ICS in Withania. However, this gene was completely lost from the genome in the lineage Tubocapsicum (indicated by dotted line and empty box), wherefore there is no possibility of ICS formation in this genus. In the case of MPF2-like-B of Withania (WSB206) the CArG-box was lost. Furthermore, by retaining the ARE, WSB206 failed to show any expression in sepals. Therefore, it has no capacity to generate ICS. On the other hand in Tubocapsicum, MPF2-like-B (TAB201) retained both the CArG-box and the ARF binding site. Consequently, it is expressed in the calyx due to relaxing of expression constraints but still impaired in ICS formation. No doubt, the expression of MPF2-like gene in the calyx is thought to...
Contribute to the ICS, but availability of plant hormones cytokinins and gibberellins is also essential [4], [6]. Cytokinins facilitate the transport of MPF2 to the nuclei of the calyx cells thus promoting cell division resulting in smaller cells. These cells in the presence of gibberellins enlarge and ultimately form ICS [32]. Hence, the function of MPF2-like proteins is also regulated at the post-translational level. Tabacopsis is sister to Withania but an "evolutionary loss mutant" of ICS. Therefore, TAB201 is expressed in the flower but has lost the capacity to interact with hormones probably due to secondary mutations [4]. Interestingly removal of the CArG-box from WSA206 and TAB201 blocked expression of these genes in sepalas, while its introduction in WSB206 rescued the sepal expression (Fig 9B). The ARE, on the other hand suppressed the expression of WSB206 in flowers including sepalas activated by CArG-box as revealed by site-directed mutagenesis and motif swapping. The expression in sepalas is a pre-requisite for ICS formation but it is not the sole requirement. Hence the reciprocal loss of regulatory elements has a major impact on the expression evolvability or in other words sub-functionalization of MPF2-like genes.

The present study found evidence, that a protein similar to MPF2-like from Vassobia could be the progenitor of MPF2-like genes, particularly MPF2-like-B. Surprisingly, V001 is expressed in the calyx but still does not feature ICS. How does this happen? Previously it was demonstrated that expression of MPF2-like genes in sepalas is plesiomorphic and so might be the ICS trait [4]. This means that the ancestor of all Solanaceae might have exhibited ICS but later on due to secondary mutations (changes in protein structure precluding their interaction with hormones, which is important for the proliferation of ICS like structures), Vassobia would later have lost the ICS. As MPF2-like-B is subsumed to be contributed by Iochrominae, i.e. Vassobia, it would also not possess the ability to trigger ICS formation. The differences in protein structure for the respective genes support these speculations (Fig. S1). Moreover, the interactions of these proteins with hormones were altered and hence no ICS would be possible. This is supported by lack of calyx proliferation when hormones are applied to Vassobia flower buds [4], [53].

Recently it is demonstrated that expression evolvability may be subjected to selection [54]. Due to the co-dominant nature of cis-acting motif changes, selection acts on them rapidly unlike for coding sequences where loss-of-function changes are recessive [55]. In Antirrhinum, differences in flower color patterning that result from differences in the expression of the MYB transcription factors encoding Rona1, Rona2, and Venosa can affect pollinator attraction, and may have been under selection during speciation [56]. In a genome-scale investigation evidence for positive selection on putative transcription factor binding sites in human proximal promoters was found [57]. Nevertheless, MPF2-like-A from Withania and MPF2 from Physalis are under positive selection [5]. To link observed promoter sequence and expression divergence with positive selection, aiming to assess the contribution of regulatory changes in the evolution of ICS, it will be essential to perform a broad range of evolutionary and genome-wide studies for these plant species. Nonetheless, the evolutionary history of MPF2-like genes allow us to conclude that expression divergence in MPF2-like genes due to reciprocal loss of cis-regulatory elements has added to genetic and/or phenotypic variations and enhanced the potential of natural selection for adaptive evolution. Moreover, this study highlights the contribution of allopolyploidy in plant evolution.

Supporting Information

Figures S1, S2, S3, S4, S5 and S6 are available at PLoS ONE online (http://http://www.plosone.org). New promoter and intronic sequences have been submitted to NCBI database under the accession numbers —. Sequence data from this article can be found in the National Centre for Biotechnology Information (NCBI) under the following accession numbers: MPF2 mRNA, (AY643734), MPF2 (AY643733:2) STMADS16 mRNA (AY643733) STMADS16 (AY643736), WSA206 mRNA (FM564866), WSB206 (FM564862), WSB206 mRNA (FM564867), WSB206 (FM564863), TAB201 mRNA (FM564865), TAB201 (FM56484).

Supporting Information

Figure S1 ClustalW formatted multiple alignment of the amino acid sequences of WSA206, MPF2, V001, TAB201 and WSB206 proteins. Blocked residues are conserved. Consensus sequence is also given at the bottom. (TIF)

Figure S2 ClustalW2 multiple alignment of the core promoter conserved block (B1) of MPF2-like promoter sequence (−150 bp to −650 bp). Five conserved sequences were identified and called shadow 1 to 5 (S1 to S5). Putative transcription start site is indicated. Dotted red box encloses the conserved CArG-box sequence. (TIF)

Figure S3 Conserved regulatory motifs in the promoter and large first intron of MPF2-like genes. Arrow indicates the direction of transcription. Grey box is the MADS-box and two empty boxes represent two exons in the upstream region. CARG-boxes occur frequently in the promoter and large 1st intron. Red X represents the position of CC(A/T)G CARG-box, which is a site for binding of AG, SEP1, SEP4, AGL1 and AGL15. Other types of CARG-boxes are shown as black X. Near the translational start site in the upstream region there are binding sites for Homeobox proteins such as ATHB1, ATHB5, AGP1 and P1. ARF binding site is shown as a crossed circle in the large 1st intron. A rough scale and legend are also given. (TIF)

Figure S4 ClustalW multiple alignment of the conserved block (Block 3) of MPF2-like large 1st intron sequence (1 bp to 500 bp). Two conserved sequences were identified and called shadow 1 to 2 (S1 to S2). Dotted red box indicates the conserved ARF binding site in MPF2-like-B large 1st intron. (TIF)

Figure S5 Schematic of MPF2-like promoter: GUS/YFP and intron:promoter: GUS/YFP constructs making. A) Shown are the 8 types of constructs using pGVT bar and pXCG-mYFP vectors containing GUS and YFP reporter genes as backbones, respectively. The red and white boxes represent the different lengths of MPF2-like promoters and introns. B) Twelve constructs using pGVT bar and pXCG-mYFP vectors containing GUS and YFP reporter genes as backbones are represented here to show the effects of CARG-box and ARF binding site on expression of these genes. PS, promoter short; I, MPF2-like large 1st intron; * CARG-box; ⊕, ARE; M, mutated; I*, introduced CARG-box; I@, introduced ARE. For details please see “Materials and Methods” section and “Table 2”. (TIF)
promoter and large 1st intron. Three days after infiltration, leaves performed using YFP reporter gene under the control of attached with YFP reporter.

Figure S6 Graph shows the interplay of CArG-box and of nuclear area of with nuclear YFP fluorescence intensity of a 35 S promoter YFP construct. PS, promoter short; I, MPF2-like large 1st intron; *a

References
1. Iman H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Review Genetics 11: 97–108.
2. Rijpkema AS, Gerats T, Vandenbusche M (2007) Evolutionary complexity of MADS complexes. Current Opinion in Plant Biology 10: 32–38.
3. He CY, Münster T, Saedler H (2004) On the origin of floral morphological novelties. FEBS Lett 567: 147–151.
4. Hu J, Saedler H (2007) Evolution of the inflated calyx syndrome in Solanaceae. Mol Biol Evol 24: 2485–2493.
5. Khan MR, Hu JY, Riss S, He CY, Saedler H (2009) MPF2-like-A MADS-box genes control the inflated calyx syndrome in Withania (Solanaceae): roles of Darwinian’s selection. Mol Biol Evol 26: 2463–2473.
6. He C, Saedler H (2003) Heterostyly expression of MPF2 is the key to the evolution of the Chinese lantern of Physalis, a morphological novelty in Solanaceae. Proc Nat Acad Sci USA 102: 5779–5784.
7. Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, et al. (2009) Reciprocal Silencing, Transcriptional Bias and Functional Divergence of Homologs in Polyploid Cotton (Gossypium). Genetics 182: 503–517.
8. Goode DK, Callaway HA, Cerda GA, Lewis KE, et al. (2011) Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. Development 138: 679–684.
9. de Folter S, Angenent GC (2006) tums meets ca in MADS science. Trends in Plant Science 11: 224–231.
10. Ulmasov T, Hagen G, Guilfoyle TJ (1997) Activation and repression of auxin-response factors. Proc Natl Acad Sci 94: 5044–5049.
11. Abhani AA, Paparizou Z, Malik S, Goode DK, Callaway H, et al. (2007) Human GLI3 Intragenic Conserved Non-Coding Sequences Are Tissue-Specific Enhancers. PLoNE 4: e466. doi:10.1371/journal.pone.0000366.
12. Kumar S, Hedges B (1990) A molecular timescale for vertebrate evolution. Nature, 392: 917–920.
13. Kaufmann K, Muino JM, Juárez-R, Airola CA, Smaczniak C, et al. (2009) Target Genes of the MADS Transcription Factor SEPALLATA3: Integration of Developmental and Hormonal Pathways in Arabidopsis flower. PLoS Biol 7 (4): e1000960. doi:10.1371/journal.pbio.1000960.
14. Tichopád A, Pecen L, Pliffe MW (2006) Distribution-sensitive insulator cluster analysis in SAS on real-time PCR gene expression data of steadily expressed genes. Comput Methods Programs Biomed Biomed 82: 44–50.
15. Swidford DL (2002) PAUP*: phylogenetic analysis under mixed models. Version 4.0b10. Sunderland (MA): Sinauer Associates.
16. Riechmann JL, Wang MQ, Meyerowitz EM (1996) DNA-binding properties of Arabidopsis MADS-domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucl Acids Res 24: 3134–3141.
17. West AG, Causier BE, Davies B, Sharrocks AD (1998) DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. Nuclear Acids Res 26: 5277–5287.
18. Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J Biol Chem 278: 21813–21819.
19. Chuck G, Hake S (2005) Regulation of developmental transitions.Curr Opin Plant Biol 8: 67–70.
20. de Folter S, Urbanska SL, van Zuijen LGC, Kaufmann K, Angenent GC (2007) Tagging of MADS domain proteins for chromatim immunoprecipitation. BMC Plant Biology doi:10.1186/1471-2229-7:47.
21. Cuen ES, Meyerowitz EM (1998) The war of the worlds: genetic interactions controlling flower development. Nature 353: 31–37.
22. Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. Cell 78: 203–209.
23. Pelaz S, Dittr GS, Bau mann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA-MADS box genes. Nature 405: 200–203.
24. Dittr G, Pinoyich A, Rohles P, Pelaz S, Yanofsky MF (2004) The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol 14: 1935–1940.
25. Fang SC, Fernandez DE (2002) Effect of Regulated Overexpression of the MADS Domain Factor AGL15 on Flower Senescence and Fruit Maturation. Plant Physiology 130: 176–189.
26. Okushima Y, Overvoorde P, Arimaa K, Alonsob JM, Chana A, et al. (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: Unique and overlapping functions of ARF7 and ARF19. Plant Cell 17: 444–463.
27. Tewari SR, Hagen G, Guilfoyle TJ (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell 15: 533–543.
28. Ulmasov T, Hagen G, Guilfoyle TJ (1999) Activation and repression of transcription by auxin-response factors. Proc. Natl Acad. Sci. USA 96: 5844–5849.
29. Hagen G, Guilfoyle TJ (2002) Auxin-responsive gene expression: genes, promoters, and regulatory factors. Plant Molecular Biology 51: 533–543.
30. Lascu E, Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. Plant Molecular Biology 49: 387–390.
31. Alonso R, Alonso L, Langhans M, Ulrich CI (2006) Role of auxin in regulating Arabidopsis flower development. Planta 223: 313–328.
32. Guilfoyle TJ, Hagen G (2007) Auxin response factors. Current Opinion in Plant Biology 10: 433–460.
33. Kim J, Harter K, Theologis A (1997) Protein–protein interactions among the Aux/IAA proteins. Proc Natl Acad Sci USA 94: 11786–11791.
34. Force AG, Lynch M, Pickett FB,Amores A, Yan YL, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.
35. Force AG, Cresko WA, Pickett FB (2004) Informational accretion, gene duplication, and the mechanisms of genetic module parcellation, 315–337 in Modularity in Evolution and Development, edited by G. Schöler and G. P. Wagner. University of Chicago Press, Chicago.
36. Davies BH, Motte P,Keck E, Saedler H, Sommer H, et al. (1999) Plena and Farinellı Redundancy and regulatory interactions between two Antirrhinum MADS box factors controlling flower development. The EMBO Journal 18: 4033–4044.
37. Causier B, Castillo R, Zhou J, Ingram R, Xue Y, et al. (2005) Evolution in action: following function in duplicated floral homeotic genes. Curr Biol 15: 1508–1512.
48. Irish V (2006) Duplication, diversification, and comparative genetics of angiosperm MADS-box genes. Advances in Botanical Research 44: 127–159.
49. Schauer SE, Schleuter PM, Baskar R, Gheyselinck J, Bolanos A, et al. (2009) Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. Plant J 59: 967–1000.
50. Jarinova O, Hatch G, Poitras L, Prudhomme C3, Grzyb M, et al. (2008) Functional resolution of duplicated hoxb5 genes in teleosts. Development 135: 3543–53.
51. Kleinjan DA, Bancewicz RM, Gastier P, Dahm R, Schonthaler HB, et al. (2008) Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. Proc Natl Acad Sci U S A 105: 18526–31.
52. He C, Saedler H (2007) Hormonal control of the inflated calyx syndrome, a morphological novelty, in Physalis. Plant J 49: 935–946.
53. Khan MR, Jinyong H, Chaoying H (2012) Plant hormones including ethylene are recruited in calyx inflation in Solanaceous plants. J Plant Physiology 169: 940–948.
54. Tirosh I, Barkai N, Verstrepen KJ (2009) Promoter architecture and the evolvability of gene expression. Journal of Biology 8:55 doi:10.1186/jbiol204.
55. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8: 206–216.
56. Shang Y, Vemall J, Mackay S, Bailey PC, Schwinn KE, et al. (2011) The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of Antirrhinum. New Phytol 189: 602–615.
57. Sethupathy P, Giang H, Plotkin JB, Hannenhalli S (2008) Genome-Wide Analysis of Natural Selection on Human Cis-Elements. PLoS ONE 10. doi: 10.1371/journal.pone.0003137.