

UNDERSTANDING FLOWER DEVELOPMENT
IN BARLEY (HORDEUM VULGARE)
THROUGH CHARACTERISATION OF
MULTIOVARY MOV MUTANTS AND THEIR
POTENTIAL USE IN HYBRID BREEDING

A thesis in fulfilment of the requirements for the
degree of Doctor of Philosophy at the University of Adelaide

Caterina Selva

School of Agriculture, Food and Wine

Faculty of Sciences

The University of Adelaide

SEPTEMBER 2019

“Everybody knows, but you’ve got to stop and think about it, to really get the pleasure about the complexity, the inconceivable nature of Nature”

Prof. Richard Feynman (1918 - 1988)

Nobel Prize in Physics



Table of Contents



Table of Contents	5
Abstract	11
Thesis Declaration	13
Acknowledgments	15
Thesis Overview	17
Chapter 1	21
STATEMENT OF AUTHORSHIP	23
HYBRID BREEDING IN WHEAT: THE TOOLS OF THE TRADE	25
Abstract.....	25
Introduction	25
Heterosis: an emergent property of nature	26
Existing male-sterility methods: present limitations and future outlooks	33
The tailoring of a wheat plant – “precision breeding”	45
Mutant resources: the joker in the pack.....	49
The future of breeding strategies	52
Challenges and perspectives.....	54
Acknowledgments	55

References.....	57
Chapter 2.....	65
THE ABC OF FLOWER DEVELOPMENT	67
It's a matter of class: the genes behind flower development.....	67
Living in a MADS world.....	72
The intimate relationship of B-class genes.....	75
Expanding our view: flower development in wheat and barley.....	79
Perspective.....	81
References.....	83
Proposed Study	87
Chapter 3.....	89
STATEMENT OF AUTHORSHIP	91
THE B-CLASS MADS-BOX GENE <i>HVMADS16</i> UNDERLIES THE BARLEY MULTIOVARY LOCUS <i>MOV1</i>	95
Abstract	95
Introduction	96
Results:.....	98
Discussion.....	110
Materials and Methods.....	119
Acknowledgments.....	126
References.....	127
Chapter 4.....	129
STATEMENT OF AUTHORSHIP	131
THE ROLE OF <i>HVSTAMENLESS1</i> IN REGULATING BARLEY FLOWER DEVELOPMENT	133

Abstract.....	133
Introduction	133
Results.....	134
Discussion	150
Materials and Methods.....	159
Acknowledgments	165
References	167
Chapter 5	169
STATEMENT OF AUTHORSHIP	171
MAPPING OF THE <i>MOV5</i> LOCUS IN BARLEY	173
Abstract.....	173
Introduction	173
Results.....	174
Discussion	190
Materials and Methods.....	200
Acknowledgments	201
References	203
Chapter 6	205
GENERAL SYNTHESIS	207
Solving the jigsaw of barley flower development.....	207
Applicability of barley <i>mov</i> mutants to hybrid breeding	214
Transferring knowledge to wheat.....	217
References	221

Conclusions	223
Contributions to Knowledge	225
Appendix A.....	227
SUPPLEMENTARY MATERIAL TO CHAPTER 3.....	229
Appendix B.....	241
SUPPLEMENTARY MATERIAL TO CHAPTER 4.....	243
Appendix C.....	257
SUPPLEMENTARY MATERIAL TO CHAPTER 5.....	259
Appendix D.....	273
STATEMENT OF AUTHORSHIP	275
UNCOVERING THE EVOLUTIONARY ORIGIN OF BLUE ANTHOCYANINS IN CEREAL GRAINS.....	281
Summary	282
Significance statement	283
Introduction	283
Results	286
Discussion.....	310
Conclusions	317
Experimental procedures	319
Accession numbers.....	323
Acknowledgements.....	324
Author contributions.....	324
Conflict of interests	324

References	325
LEGENDS FOR SUPPORTING INFORMATION.....	329

Abstract



In today's world we cannot ignore the challenge of feeding an increasing global population despite more erratic and extreme weather patterns. Among the many approaches suggested to address this challenge and improve crop yield, hybrid breeding seems a promising solution. Through the phenomenon of heterosis, hybrid breeding provides the opportunity to obtain more resilient plants and more grain, thus ensuring food stability. This is particularly promising for staple crops such as wheat and rice that represent the main form of sustenance worldwide.

Despite this, the application of hybrid breeding in autogamous cereals like wheat must overcome many difficulties. Specific modifications to the architecture of wheat flowers are required to maximise seed set obtained from cross-pollination, while inhibiting self-fertilization. Scientific obstacles include a complex genome organization and polyploidy, as well as the physiology and characteristics of the wheat flower itself. As a result, this project focuses on barley as a surrogate genetic model for wheat. Since barley is diploid, working with barley mutant resources greatly simplifies genetic analysis and potentially makes it easier to uncover mutant phenotypes otherwise hidden by the genic redundancy embedded within a hexaploid genome.

This thesis focuses on the characterization of three barley *multiovary* (*mov*) mutants named *mov1*, *mov2* and *mov5* which show abnormal flower development. Compared to a wild-type

barley floret, the *mov* mutants share the characteristics of having a complete or partial reduction of stamens, combined with an increase in the number of carpels. From the perspective of hybrid seed production, these mutants potentially present the dual advantage of being male-sterile and of being able to produce multiple seeds per floret. The main aims of the project were to identify which gene(s) confer the *mov* phenotype, understand the relationship between the identified genes and establish how the candidate genes interact with the known floral development network, particularly in the context of the ABC model. For each mutant, the approach taken relied mostly on forward genetics via mapping. Concomitantly, developmental and morphological aspects were explored using microscopy, as well as expression analysis and transient *in vivo* assays to determine interaction dynamics of the identified genes with known players of floral development.

Overall, plausible candidate genes have been identified for each of the three *mov* loci and their role in flower development has been explained. These results provide the basis for a model explaining flower development in barley, taking into account not only the main genetic actors, but also their regulators. The knowledge transfer from barley to wheat and the applicability of using floral development mutants for wheat hybrid breeding is discussed.

Thesis Declaration



21st September 2019

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Acknowledgments



Life is a journey. This journey has led me to the traditional country of the Kurna people of the Adelaide Plains and I recognise and respect their cultural heritage, beliefs and relationship with this land.

More than a journey, the Ph.D. is a rollercoaster and I have shared this ride with many colleagues and friends, too many to name, but who all joined me, for a little or a while, in this adventure. It has been instructive to see how vibrant, supportive and sometimes competitive the scientific community can be.

Among the many help I've received from colleagues, special mention goes to Margaret Pallotta, without whose help this thesis would surely be less readable and the meetings duller. I'm very grateful to Dr. Xiujuan Yang for her help in providing crucial components for my experimental endeavours.

Special thanks must go to my supervisors: Dr. Ute Baumann, Dr. Ryan Whitford and A/Prof. Dr. Matthew R. Tucker, who found the time to mentor me through the course of my scientific explorations. They have been enlightening in portraying different aspects of becoming a successful scientist.

Thanks to all the friends I've made along the way, my adoptive "dysfunctional family". Some of whom I shared the joys and pains of the Ph.D. with, others a beer at the pub or a campfire in the starry Australian outback.

Lastly, but certainly not for importance, to my family. I'm particularly indebted to my husband Matteo Riboni: companion, lover and friend. Thanks for the unconditional support and motivation, for our (not always peaceful) scientific conversations and for the constructive criticism in the face of my stubbornness.

To the reviewers and readers of this thesis. I'm aware this work does not represent a literary masterpiece, but I hope it will keep you interested until the end as you sip your cup of coffee.

Thesis overview



To help navigate this thesis, an overview outlining the contents of each chapter is provided below.

The aim of Chapter 1 is to set the context around which this work has been structured. It presents a Literature Review on hybrid breeding and expands on the feasibility of its use in wheat considering current limitations and advantages. Here, the reader will find a description of the prevailing strategies employed and aspects that might be considered to make hybrid breeding in wheat more accessible and cost-effective in the near future. This chapter proposes how modifying floral morphology could be an interesting avenue of pursuit towards increasing the effectiveness of hybrid wheat seed production.

Chapter 2 forms a General Introduction about the state-of-the-art knowledge on flower development and provides a basis for understanding the subsequent Results sections. Chapter 2 brings examples from across plant species, ranging from Gymnosperms to Angiosperms. Explanation of the ABC model will enable the reader to understand the vast diversity of floral structures and conceive how flexible and plastic the model is. The main processes of flower development are discussed, with particular attention to specific gene classes. The chapter ends with an overview of the current knowledge on wheat and barley flower development and highlights the gaps this project has attempted to address.

Chapters 3, 4 and 5 present the research Results and report the characterization of barley *multiovary (mov)* mutants *mov1*, *mov2* and *mov5*. Within the present work, each *mov* locus has been located and promising candidate genes have been proposed for each mutant based on genetic, transcriptomic and morphological data. Although the data has yet to be published, these chapters are presented in manuscript style. Due to the nature of the experiments, the reader may find some procedures and methods repetitive. With this in mind, an attempt has been made to keep repetition to a minimum. For the same reason, the introduction sections to Chapters 3, 4 and 5 only include current literature specific for each mutant and avoid discussing flower development further.

Chapter 6 contains a General Discussion, in which results from *mov1*, *mov2* and *mov5* are used to construct a unifying model for flower development in barley. The chapter also discusses the potential application of *mov* mutants in a hybrid seed production scenario and is followed by Chapter 7, which highlights the Conclusions that may be drawn from the present work, and Chapter 8, which lists the Contribution to Knowledge created by the current thesis.

Appendices A, B and C contain Supplementary Information to Chapters 3, 4 and 5, respectively. Appendix D presents an addition to the project, carried out in collaboration with Professor Chengdao Li and colleagues at Murdoch University, Perth, Western Australia. This side project explores the genetic basis of anthocyanin accumulation in the aleurone of grains of monocotyledons, with particular attention to wheat and barley. Grain colour could be used as a visual selective marker in breeding and is thus pertinent to the broader objective

of this work. The data in Appendix D has been accepted at *The Plant Journal* (September 2019).

Chapter 1



Statement of Authorship

Title of Paper	Hybrid breeding in wheat: the tools of the trade
Publication Status	Unpublished and unsubmitted review written in manuscript style.

By signing the Statement of Authorship, each author certifies that:

- i. each author's contribution to the manuscript is accurate; and
- ii. permission is granted to include the manuscript in the thesis

Principal Author

Name of Principal Author	Caterina Selva		
Contribution to the Paper	Collected and interpreted the literature. Wrote the manuscript.		
Percentage of Contribution	70 %		
Certification	This review is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this review. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	14/09/2019

Co-author Contributions

Name of Co-author	Dr. Matteo Riboni		
Contribution to the Paper	Assisted in figure preparation and manuscript layout; evaluated the manuscript. I hereby certify that the Statement of Authorship is accurate.		

Contribution to the Paper	Evaluated and edited the manuscript. I hereby certify that the statement of Authorship is accurate.		

Contribution to the Paper	Evaluated and edited the manuscript. I hereby certify that the statement of Authorship is accurate.		
Signature		Date	2. Sept. 2019

Name of Co-author	Dr. Ryan Whitford		
Signature		Date	4/09/2019

Name of Co-author	Prof. Dr. Matthew R. Tucker		
Signature		Date	4/09/2019

Hybrid breeding in wheat: the tools of the trade

Caterina Selva¹, Matteo Riboni¹, Ute Baumann¹, Tobias Würschum², Ryan Whitford¹,
Matthew R. Tucker¹

¹ School of Agriculture Food and Wine, University of Adelaide, Waite Campus, Urrbrae 5064, South Australia, Australia

² State Plant Breeding Institute, University of Hohenheim, Stuttgart 70593, Germany

Abstract

Hybrid breeding in wheat has the potential to deliver major increases in yield. This is a requisite to guarantee food security for increasing population demands and to counteract extreme environmental conditions. For successful hybrid breeding in wheat, efficient fertility control systems are needed to force outcrossing and avoid self-pollination. Recent technological advances may provide a suite of tools to achieve this. This review summarizes male-sterility systems for wheat hybridization, focusing on genetic resources that can be used to alter floral development and spike morphology. Importantly, mutant resources have the potential to deliver distinct male and female lines for enhanced outcrossing, a key requirement in the progress towards hybrid wheat breeding.

Introduction

As of 2019, we share the world with over 7.7 billion humans [1]. Although the population growth rate has declined over the last half-century [1], the global population has risen

leading to concerns about our ability to nourish everyone. Indeed, for the third consecutive year, there has been an incremental increase in world hunger, which reached a total of approximately 821 million undernourished people globally in 2017 [2]. This trend has been accentuated by climate variability and extreme weather events that increasingly account for the majority of all internationally reported disasters [2]. These climatic events negatively impact agricultural production by affecting crop yield, planted area and the number of crops grown annually. Specifically, among the many effects of climate extremes, field crops represent the agricultural sector that is most affected in terms of damage and losses [2].

In order to achieve food security in the face of increased climate variability and weather extremes, we need to breed high-yielding and resilient crops that are suited to sub-optimal growing conditions. This can be addressed in different ways depending on the physiology and plasticity of each crop. For example, some studies have focused on increasing photosynthetic efficiency [3] or resource use efficiency [4], whereas others have tried to identify, isolate and deploy gene sequences related to tolerance traits to either single or combined environmental stresses [5-7]. Another approach to improve yield and yield stability in crops is to exploit the phenomenon of heterosis.

Heterosis: an emergent property of nature

Heterosis or hybrid vigour refers to a widespread natural phenomenon whereby the first generation progeny of two inbred lines typically exhibits superior performance relative to the parents themselves (**Figure 1**). Heterosis is an emergent property of all living systems in the sense that the hybrid progeny shows properties that cannot be accounted by simple linear additivity of the individual properties of each parent [8]. In plants, hybrid performance

usually manifests itself as increased biomass due to more robust vegetative growth, a stronger root system and a greater uniformity in growth and flowering [9-12]. This results in larger and stronger plants that are able to better resist both biotic and abiotic stresses, thus leading to overall improved yields and greater yield stability on marginal cropland.

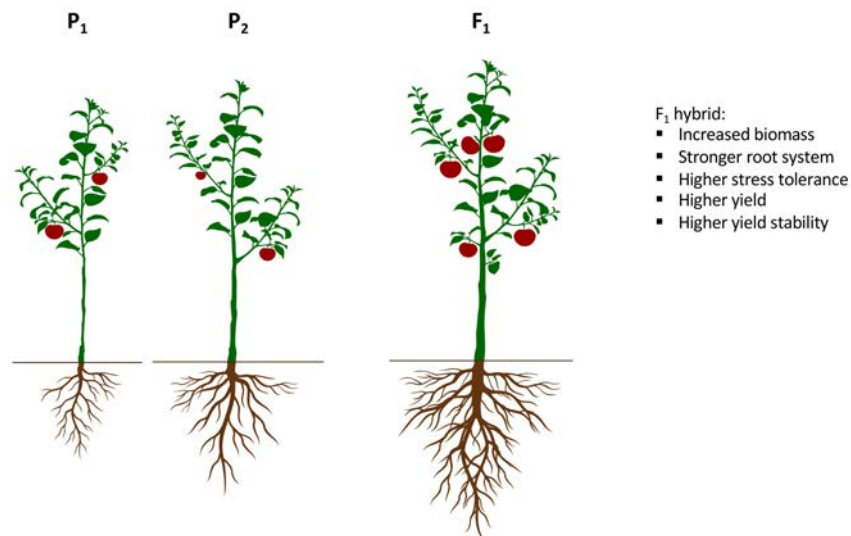


Figure 1. Schematic of heterosis. F₁ hybrid plants demonstrate superior performance compared to both parents (P₁ and P₂) which typically manifests as the characteristics listed.

Current models for heterosis

The hybrid vigour effect was first described in the 1876 book “The effects of cross and self fertilization in the vegetable kingdom” by Charles Darwin [13] where he compiled data for more than 60 different plant species and compared cross- and self-pollinated plants in terms of growth, development and fertility. However, the term “heterosis” was only coined in 1914 by George H. Shull and since then many models have been proposed to genetically explain heterosis, with the most common being the dominance, overdominance, pseudo-overdominance and epistasis models.

According to the dominance model (**Figure 2A**), heterosis is the overall accumulation of favourable alleles in the hybrid progeny as a result of complementation of deleterious recessive alleles present in the inbred parents [10,14-16]. The overdominance model (**Figure 2B**) proposes that heterozygosity at key individual loci is the contributing factor towards heterosis, allowing a synergistic interaction of different alleles at a given locus in hybrids to exceed the effects of homozygosity at the same locus in the parents [10,17,18-20]. In the pseudo-overdominance model (**Figure 2C**), heterosis is explained by complementation at different yet linked loci [16]. Finally, the epistasis model (**Figure 2D**) takes into consideration the sum of interactions between two or more favourable genes inherited from the two parental lines [21-23].

Although each of these genetic models have been extensively tested in many species, it is becoming increasingly apparent that these models are not mutually exclusive, meaning they can concurrently occur in the same plant. Indeed, the exact contribution of each of these genetic mechanisms to heterosis still remains elusive and can vary greatly according to the trait, cultivar, species, and parental combination under consideration [24]. For example, overdominance has been observed to be the major contributor to heterosis in tomato [25], whereas heterosis in rice (*Oryza sativa*) relies upon both dominance and overdominance effects [26-28]. By comparison, the overdominance and pseudo-overdominance models seem to largely explain the expression of heterosis in maize (*Zea mays*) [29-31].

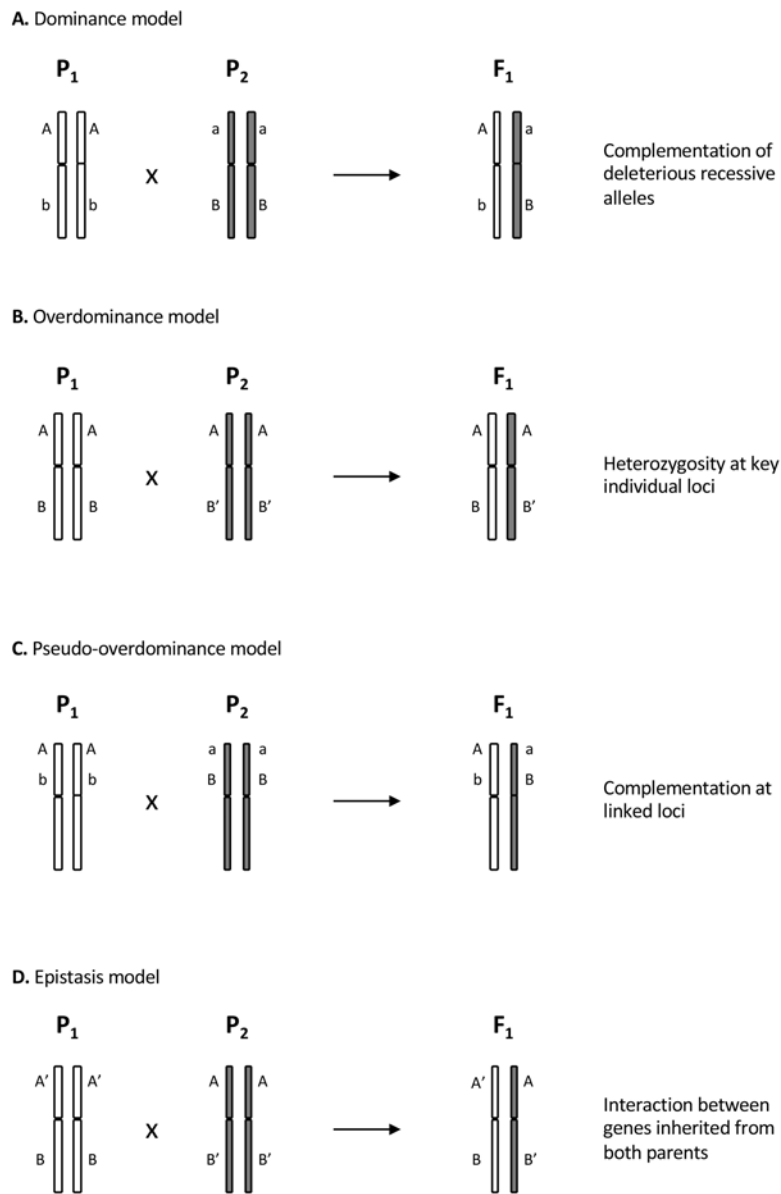


Figure 2. Genetic models of heterosis. A and B represent two distinct genes; a, A, A' indicate allelic variation for gene A, while b, B, B' indicate allelic variation for gene B for parents (P₁ and P₂) and F₁ hybrid progeny.

Overall, heterosis appears to be a very complex phenomenon which cannot be readily explained by a single unified model. For this reason, studies seeking to understand the basis of heterosis are now considering a more comprehensive systems-based approach. Examples of such a holistic view are studies that take into account variation in expression of regulatory networks [32,33], epigenetic modifications [34-38], genomic rearrangements [39,40],

energy-use efficiency [41-44], circadian-related pathways [45], as well as the role of small RNAs [46-48]. A systems-based approach is being facilitated by the “-omics” era which surveys a wider, more global representation of changes at the whole organism level, each of which is necessary in determining connections between heterotic performance and the underlying molecular events [49]. For example, a study assessing 1604 hybrids and their parents from European winter wheat elite lines showed that epistatic interactions are the major components for grain yield heterosis, overriding dominance effects [50].

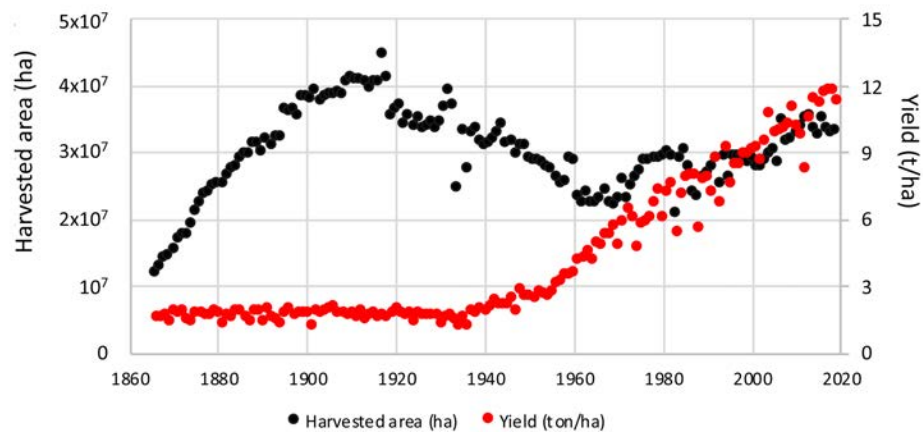
Status quo of plant hybrid breeding

Although the molecular mechanisms responsible for hybrid vigour still remain unclear, it is widely reported that the heterotic effect improves as the genetic diversity of the parental lines increases. Interestingly, it appears that heterosis is also maximised as ploidy increases [51]. This is particularly relevant to agricultural systems, as many of the major crops and food plants are polyploids (containing more than two sets of chromosomes). These include wheat (*Triticum aestivum*), maize, oat (*Avena sativa*), potato (*Solanum tuberosum*), peanut (*Arachis hypogaea*), sugarcane (*Saccharum* spp.), cotton (*Gossypium hirsutum*), canola (*Brassica napus*), strawberry and more.

Currently, hybrid breeding is widely used in broad-acre crops such as rice, maize, canola, sorghum (*Sorghum bicolor*) and sunflower (*Helianthus* spp.), as well as for many horticultural crops [52]. For maize, the advantages of hybrid maize were discovered in the early 1900s by Shull [18] with hybrid programs developed as of the 1920s [10]. It is estimated that heterosis in maize initially contributed to a yield improvement of at least 15 % [53]. The introduction of hybrid breeding, together with advances in agronomic practices and improved inbred lines

has resulted in a continuous improvement in maize yields. In the United States alone, maize average yields showed an increase of approximately 400 % from 1930 to 2002 and the increase still carries on today [9,10] (**Figure 3A**). Despite the self-pollinating nature of rice, hybrid rice which was originally introduced in China in 1964 is now commercially established and widely distributed across China, India, Vietnam, Bangladesh and Indonesia [54]. Similarly to maize, between 1977 and 1997 hybrid rice showed yield advantages of 20 - 30 % over conventional inbred rice varieties shortly after their introduction [55]. From the late 1970s to 2008, hybrid rice yields have increased by about 800 % and still continue to deliver yield benefits [56] (**Figure 3B**).

A. USA corn yield and harvested area (1866 – 2019)



B. China rice yield and harvested area (1951 – 2010)

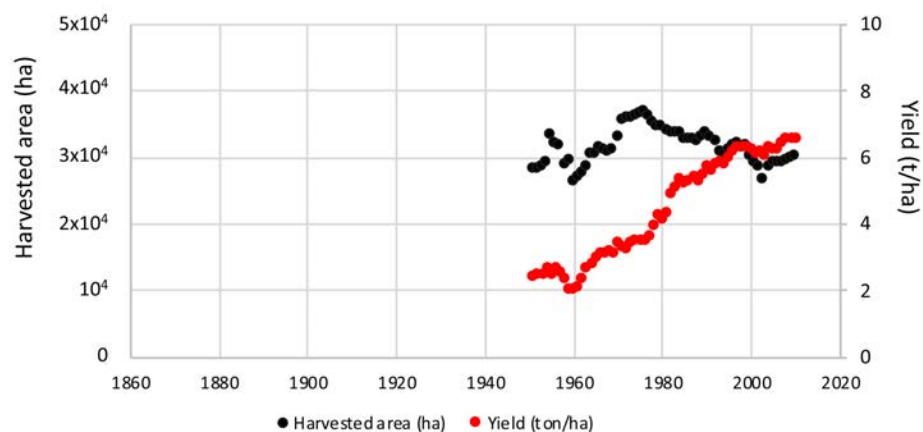


Figure 3. Historical changes in yield (red) and harvested area (black) for (A) corn in USA (1866 – 2019) and (B) rice in China (1951 – 2010). Data for corn was sourced from USDA-NASS (<https://quickstats.nass.usda.gov>) , while rice statistics were gathered from IRRI (http://ricestat.irri.org:8080/wrs/WRS_manila.html).

In wheat, heterosis was first recorded in 1919 [57] and since then attempts have been made to establish an effective hybrid breeding system [58]. Over the years, numerous studies have shown that wheat hybrids outperform line varieties in terms of stress tolerance and a predicted yield advantage of approximately 10 % [50,58-61]. For example, it was recently reported that hybrid wheat can deliver a higher grain yield, higher protein content and high quality over line varieties, possibly due to higher nitrogen use efficiency, if parental lines are carefully chosen [62]. This indicates that, as seen for other predominantly inbreeding species such as rice, heterosis can also be harnessed and exploited in wheat. Wheat is among the main commodities produced worldwide. Initially cultivated in the area of the Fertile Crescent about 10,000 BCE, it is now farmed on more land area than any other food crop [2]. This was mainly possible thanks to its high environmental adaptation and plasticity in processing. For these reasons, wheat plays a significant role in human nutrition as a major source of both vegetable-based protein and carbohydrates, as well as providing dietary minerals, B vitamins and fibre [63,64]. Apart from human nutrition, wheat grain is also used as animal feed and for fermentation into alcoholic beverages. Recently, the use of wheat as potential biofuel has been investigated [65-67]. Consequently, being able to boost wheat yields is of paramount importance when dealing with future food supply challenges.

Despite the importance of wheat as a staple crop and the improved performance shown by hybrids, hybrid wheat accounts for only a minor fraction of all commercial wheat production.

As less than 1 % of the global wheat area is currently planted to hybrid wheat [12], hybrid wheat represents a niche sector when compared to commercialization of other hybrid cereals like rice and maize. A major limitation to the exploitation of yield improvements from hybrid wheat is its high cost of seed production, a consequence of wheat's floral architecture. Wheat is an autogamous cereal, meaning florets preferentially undergo self-fertilization. Hybrid seed production, however, requires efficient cross-pollination between inbred lines. To facilitate the production of hybrids in wheat, obligate outcrossing is necessary, this can be achieved by using male-sterile plants as female parents.

Existing male-sterility methods: present limitations and future outlooks

In the early 1960s both private and public institutions invested greatly in wheat hybrid research [68]. However, the hybridization systems that were explored were unreliable and impractical and the yield gains of hybrids over conventional line breeding were too marginal to be deemed commercially advantageous. The lack of the expected return on investment meant that variety developers, growers and companies gradually lost interest, funding progressively dwindled and research in hybrid wheat was discontinued. Some of the obstacles faced, such as the high production and distribution cost of hybrid seeds and the lack of seed purity are still present today and are the main reasons for the delays in hybrid wheat commercialization. Nonetheless, different mechanisms were proposed to induce male-sterility in wheat plants (**Table 1**).

Emasculation can be performed by physically or mechanically removing anthers from wheat florets [69]. However, contrary to the mechanical de-tasselling of maize male flower organs, wheat emasculation is simply too labour-intensive and time consuming to be feasible for large scale production.

Chemical sterility control strategies

Male-sterility in wheat can also be induced by spraying plants with chemical hybridizing agents (CHAs), the name given to a group of chemicals that show a selective male gametocide effect. Although several CHAs have been developed over the years, only a few are being used for commercial wheat breeding today. This is because most CHAs exhibit phytotoxic effects which reduce seed set, can result in incomplete induction of male-sterility and/or partial female-sterility, reduced seed germination and seedling vigour [70,71]. The only CHA currently used in Europe is the plant growth regulator sintofen (commercially known as Croisor[®]100 from Saaten Union Recherche, France). On the other hand, clofencet (known as Genesis[®] from Monsanto) is the chemical compound that has been registered and approved for field testing in USA since 1997. The main advantages of Croisor[®]100 and Genesis over other gametocides is the relatively lower genotype specificity, allowing a broader choice of parental combinations [72]. These factors reduce the cost and improve the effectiveness of hybrid seed production. Recently, a new broad-spectrum pyridazine compound known as SQ-1 from Key Laboratory of Crop Heterosis of Shaanxi Province is being deployed in China for large scale production of hybrid wheat seeds. SQ-1 has been proven to induce complete male sterility with little associated risk of affecting important agronomic traits [73-75]. Despite Croisor[®]100, Genesis and SQ-1 being more effective than other gametocides, the major factors limiting the use of CHAs remains the narrow window of application that is

dependent on environmental conditions, dosage optimization, as well as the bio-safety concerns of releasing relatively unsafe chemicals into the environment [72,76].

Genetic sterility control strategies

Genetically, male-sterility can be obtained through nuclear-encoded male sterility (NMS) or cytoplasmic male sterility (CMS). NMS, also known as genic or genetic male sterility, is typically expressed through mutations in nuclear-encoded genes or through the expression of transgenes [77]. In wheat, a handful of spontaneous mutants have been recorded to induce NMS. Depending on the mutation, these can further be divided in dominant and recessive mutants. Dominant mutant loci include *Ms2* (4DS), *Ms3* (5AS) and *Ms4* (4DS) [78-80], while recessive mutant loci described so far include *ms1* (4BS) and *ms5* (3AL) [81,82], with seven allelic variations observed for the *ms1* locus in different cultivars: *ms1a*, *ms1b*, *ms1c*, *ms1d*, *ms1e*, *ms1f* and *ms1g* [83-87]. Due to the interest in their use in hybrid breeding, several studies have attempted cloning and characterization of these loci [88-92], however the ideal mutant for practical use in NMS needs to be non-conditional, monogenic and recessive (for example *ms1* and *ms5*), thus limiting the repertoire of available genes. It is worth noticing that the mutated genes inducing NMS, apart from being male-sterile, do not usually confer any additional phenotype. This introduces the issue of distinguishing male-sterile plants from the sibling male-fertile plants and requires finding a way to remove the male-fertile plants in the field (roguing). This issue has been partly addressed by Cao *et al.* (2009) who successfully combined the *Ms2* locus to the dwarfing locus *Rht10*, effectively allowing identification between dwarf male-sterile and tall male-fertile plants [93,94]. The idea of combining a male-sterility locus to a visual marker was further developed in the XYZ system [95,96]. This system requires a male-sterile recessive locus and a fertility restorer

gene on chromosome 5R from rye (*Secale cereale*) as this chromosome also carries the *hairy peduncle* (*hp*) locus. In the XYZ system, the three wheat lines X, Y and Z are homozygous for the recessive male-sterile locus and contain 2, 1 and 0 doses of the alien rye 5R chromosome, respectively. By exploiting trait segregation in the selfed progeny of each line and limited roguing with *hp* as a visual dominant marker, the system allows production of large quantities of homozygous male-sterile seeds which contain no alien introgression that may affect agronomic traits. An improved strategy of the XYZ was proposed by Zhou *et al.* (2006) which introduced the seed colour marker *Ba* locus from chromosome 4E of *Agropyron elongatum* ssp. *ruthenicum* [97]. Seeds containing the *Ba* locus exhibit a dosage-dependent blue coloration and allow for precocious selection of male sterile seeds compared to the use of an adult plant trait like *hp*. Further improvement came from breeding the blue seed colour marker on the same chromosome as the dominant male sterile locus *Ms2* and the dwarfing gene *Rht10* in both durum and common wheat [98]. Nonetheless, using an introgressed visual marker raises limitations in regard to phenotypic penetrance of the visual marker, as well as the effect of alien chromatin on genome stability during meiosis, thus impacting transmission of the trait.

On the other hand, CMS refers to the inability of the plant to produce viable pollen due to rearrangements or mutations in mitochondrial DNA [99-101]. CMS is the most common form of male sterility observed in nature [102] and can arise spontaneously, as a result of interspecific and intraspecific crosses, or as a result of anthropogenic mutagenesis (for example by ethyl methane sulfonate, X-rays or γ -irradiation) [103]. Historically, CMS lines of alloplasmic wheat have been created by transferring the cytoplasm of a wild wheat or related species into wheat through crossing [71,104,105]. As for NMS, each CMS system can be

counteracted by the action of specific nuclear-encoded genes, termed fertility restorer (*Rf*) genes. However, identifying suitable *Rf* genes for each system and understanding the underlying genetics is not trivial.

In some cases, more than one *Rf* gene is needed to restore fertility. For example, two *Rf* genes termed *Rf6H^{chS}* and *Rf1H^{chS}* from *Hordeum chilense* are required to restore pollen fertility in the CMS msH1 system obtained from using the cytoplasm of *H. chilense* H1 accession to create alloplasmic bread wheat [106,107]. Similarly, a single locus does not restore complete fertility in CMS lines of common wheat with the cytoplasm of *T. timopheevii*. QTL analysis, mapping and haplotype analysis identified up to eight *Rf* genes for the *T. timopheevii* cytoplasm: *Rf1* (1A), *Rf2* (7D), *Rf3* (1B), *Rf4* (6B), *Rf5* (6D), *Rf6* (5D), *Rf7* (7B) and *Rf8* (2D) [104,108-113]. However, identifying restorer genes alone is not sufficient to restore full fertility, as the function of major *Rf* genes can be dependent on additional modifier loci [114,115], thus highlighting the importance of determining the underlying genetic mechanisms of CMS for complete and stable fertility restoration.

Overall, CMS is a complex system and can only be used for hybrid seed production once the CMS mutant and its associated *Rf* gene(s) are identified. Furthermore, the CMS mutant germplasm must not exhibit any negative pleiotropic effects on plant growth and development. A major limitation to the utilisation of this method in wheat is the lack of effective fertility restoration to the F₁. Given the hexaploid nature of wheat many *Rf* loci need to be stacked when breeding the parental lines to ensure complete fertility restoration. Furthermore, maintaining the male-sterile lines for both CMS and NMS requires a complex

breeding scheme with these lines often exhibiting a deleterious effect on plant vigour and yield.

On a positive note, improvement to CMS systems by biotechnological engineering promises to overcome some of these major drawbacks [116,117]. Hybridization by protoplast fusion and plastid genome transformation have been demonstrated to be feasible in tobacco [118] and *Brassica oleracea* [119]. Recently, chromosome engineering in wheat allowed to isolate the *Rf^{multi}* locus to ensure complete male sterility in alloplasmic lines with cytoplasm from *Aegilops kotschyi*, *Ae. uniaristata* and *Ae. mutica*, thus extending the range of possible combinations for CMS [120,121].

Environmental sterility control strategies

Plant sterility can also be environmentally induced. In environment-sensitive genetic male sterility (EGMS), pollen viability can be thermosensitive, photoperiod sensitive or photothermal sensitive if responsive to changes in temperature, day length or both, respectively. Originally reported in pepper (*Capsicum frutescens*) [122], EGMS has now been also described in cereal crops such as maize [123] and rice [124,125]. In wheat, gametogenesis is susceptible to temperatures below 4 °C or higher than 25 °C and to daylengths of more than 14 hours [126-128]. Thermosensitive nuclear genes *wtms1* (2B) [129] and *TaPaO1* [130], as well as photoperiod sensitive genes *wptms1*, *wptms2* [131] and *wptms3* [132] located on chromosomes 2B, 5B, and 1BS, respectively, have already been identified in wheat. Although EGMS does not require a maintainer line nor restorer genes, global application is limited to specific geographical regions with appropriate climatic conditions.

Biological sterility control strategies

Another way that self-pollination could be prevented in wheat would be to employ self-incompatibility (SI). SI refers to mechanisms whereby a plant with fully viable female and male gametes is unable to produce seeds upon self-pollination and is a strategy currently used for hybrid production in *Brassica oleracea* and *B. rapa* [133,134]. In grasses, gametophytic SI is controlled by the multiallelic loci *S* and *Z*, so that pollen compatibility in crosses can vary depending on the genotype and combination of *S* and *Z* alleles that have been crossed [135-137]. SI is quite common in grasses, however, introgressing SI into wheat from a related self-infertile cereal like rye may prove challenging as little is still known about the molecular details of SI systems in grasses. Furthermore, whether SI experiences a breakdown in polyploids remains an open question. Confidently, a recent study did not find a strong association between ploidy and SI at the species or family level [138], indicating that wheat may still show the same level of SI if suitable genes are introgressed from self-infertile diploid and closely related species.

Table 1. Main advantages and disadvantages associated to existing male-sterility systems for hybrid wheat seed production; refer to text for details.

Sterility method	Advantages	Disadvantages
<i>Mechanical emasculation</i>		<ul style="list-style-type: none"> ▪ Labour-intensive and time-consuming ▪ Cannot be automated
<i>Chemical Hybridizing Agents (CHAs)</i>	<ul style="list-style-type: none"> ▪ Relatively easier administration method (spraying) ▪ Broad spectrum, no genotype specificity ▪ Some already in use: Croisor®100, Genesis®, SQ-1 	<ul style="list-style-type: none"> ▪ Variability in genotype response ▪ Application is restricted to favourable environmental conditions ▪ Requires dosage optimization ▪ Bio-safety concerns and phytotoxic effects
<i>Nuclear-encoded Male Sterility (NMS)</i>	<ul style="list-style-type: none"> ▪ Some loci have already been identified and characterised ▪ Do not usually confer additional phenotypes 	<ul style="list-style-type: none"> ▪ Locus preferably needs to be non-conditional, monogenic and recessive ▪ Requires roguing ▪ Requires maintenance of male-sterile lines
<i>Cytoplasmic Male Sterility (CMS)</i>	<ul style="list-style-type: none"> ▪ Some CMS systems and corresponding <i>Rf</i> genes have already been investigated ▪ CMS can be engineered to avoid major drawbacks 	<ul style="list-style-type: none"> ▪ Alloplasmic wheat lines need to be created ▪ May require stacking of multiple <i>Rf</i> genes for full fertility restoration ▪ May exhibit negative effects on plant growth ▪ Requires maintenance of male-sterile lines
<i>Environment-sensitive Genetic Male Sterility (EGMS)</i>	<ul style="list-style-type: none"> ▪ Does not require <i>Rf</i> genes ▪ Does not require maintainer line 	<ul style="list-style-type: none"> ▪ Use is restricted to certain geographical regions
<i>Self-incompatibility (SI)</i>	<ul style="list-style-type: none"> ▪ Could be introgressed from related self-infertile cereals 	<ul style="list-style-type: none"> ▪ SI systems still need to be characterized molecularly
<i>Transgenic systems</i>	<ul style="list-style-type: none"> ▪ Specific – no pleiotropic phenotypes ▪ Versatile ▪ Can be designed to produce transgenic-free hybrid seeds 	<ul style="list-style-type: none"> ▪ Most systems produce transgenic hybrid seeds

Transgenic systems: the benefits of biotechnology

Each of these fertility control systems has associated drawbacks in terms of applicability and cost to hybrid seed production. However, advances in the use of biotechnology provide cheaper alternatives to promote hybridization.

The first transgenic hybridization system, called SeedLink™ from Bayer CropScience, relied on tapetum-specific expression of a cytotoxic bacterial ribonuclease BARNASE from *Bacillus amyloliquifaciens* linked to glufosinate resistance [139,140] (**Figure 4A**). When female plants expressing BARNASE are crossed with a plant carrying a tapetal cell-specific ribonuclease inhibitor BARSTAR, fertility is restored in the F₁ progeny [141]. To avoid residual male-sterility in F₁ plants improvement of the SeedLink™ technology initially involved splitting the BARNASE gene into complementary, non-overlapping fragments located at allelic positions [142,143]. These split-gene systems based on “allelic repulsion” are constantly being improved [144-146] and rely on the fact that only plants co-expressing both fragments show male-sterility, while fertility is restored in F₁ plants due to segregation during meiosis.

A more modern and versatile dual-component system has been proposed by Singh *et al.* (2015) [147] (**Figure 4B**). In this system, the female cassette contains a regulatory component driving tapetum-specific expression of the Arabidopsis *BECLIN1* gene to induce male sterility. On the other hand, the male expression cassette regulates tapetum-specific expression of Arabidopsis *COP1*, a gene that does not affect fertility. These plants are therefore fully fertile. In F₁ plants, *BECLIN1* expression is abolished through specific COP1-mediated degradation of the regulatory component of the female cassette, thus resulting in complete fertility restoration. Promisingly, this system works with any male-sterility inducing gene other than *BECLIN1*. This works by simply substituting *BECLIN1* in the female cassette without the need to affect the male expression cassette.

Numerous other alternative systems to control pollination through metabolic engineering and inducible control systems have been proposed or are under development. The major strategies have been comprehensively reviewed by Kempe and Gils (2011) [76]. Although,

the described biotechnological advances contribute in facilitating hybridization and reducing hybrid seed production costs, all of these strategies generate transgenic hybrid seed. Restrictions to global trade combined with onerous regulatory requirements for release of GM crops have contributed to shifting efforts towards using biotechnology to produce non-transgenic hybrid seeds.

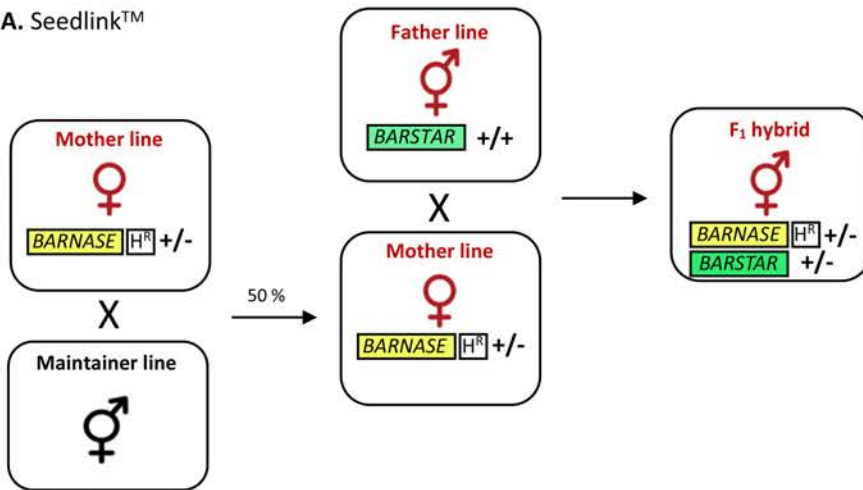
Transgenic plants to produce transgenic-free seeds

One method that uses a transgenic male-sterility system to produce non-transgenic F₁ hybrids is termed Seed Production Technology (SPT, Pioneer Hybrid International, Inc.) (**Figure 4C**) and has been successfully implemented for the production of maize hybrids since 2011. The SPT system requires a maintainer line and a male-sterile female line. The female line is non-transgenic and homozygous recessive for a male-sterility locus. On the other hand, the maintainer line is likewise homozygous recessive for the same male-sterility locus but carries a specifically designed transgenic construct that renders the plant male-fertile. The transgenic construct contains three fundamental elements: a single dominant allele complementing the male-sterility locus, a pollen germination inhibitor and a seed colour marker. The novelty of this method lies in the action of the three elements together. Due to the pollen germination inhibitor, the only viable pollen produced from the transgenic maintainer line will be non-transgenic and carry the male-sterility locus. This pollen is used to cross-fertilize the female plants and the resulting seed (non-transgenic and still homozygous for the male-sterility locus) will be used to grow female lines. In the hybrid seed production field, the female lines will be crossed with a male line able to restore fertility. The F₁ seeds will thus be fully fertile and non-transgenic. Furthermore, self-fertilization of the transgenic maintainer line will result in 50 % non-transgenic seeds and 50 % seeds carrying

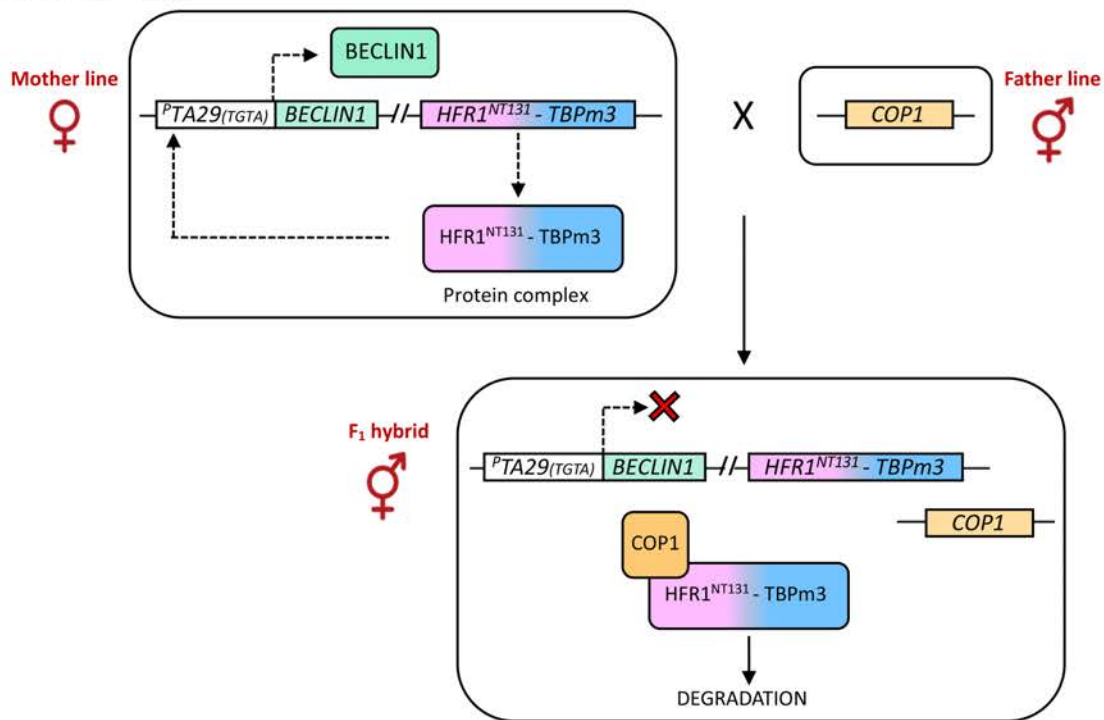
the transgene. Seed selection is possible due to the seed colour marker in the transgenic construct. The transgene event is thus restricted to the maintainer line and serves to produce large amounts of non-transgenic male-sterile seeds for the production of male-sterile female parents.

It would be ideal to adapt the SPT system to wheat. However, despite being successfully deployed in maize, its adaptation for hybrid wheat breeding has been prolonged by the inherent challenge of identifying suitable genes conferring male sterility. Nonetheless, significant advances have recently been made, with all components of the SPT system deemed functional in wheat [88,91]. Originally, the seed colour marker used in the SPT system is the fluorescent protein DsRed originating from the algae *Dicosoma* sp. [148]. Although DsRed is functional and an effective marker in molecular biology, it does come with the drawback of being a non-plant derived sequence and hence contributes to added GM-trait deregulation costs. A more ideal seed colour marker from a GM-trait deregulation perspective would be one that is native to bread wheat or its wild relatives, like the *Ba* locus from *Agropyron elongatum*.

A. Seedlink™



B. BECLIN1 - COP1



C. Seed Production Technology (SPT)

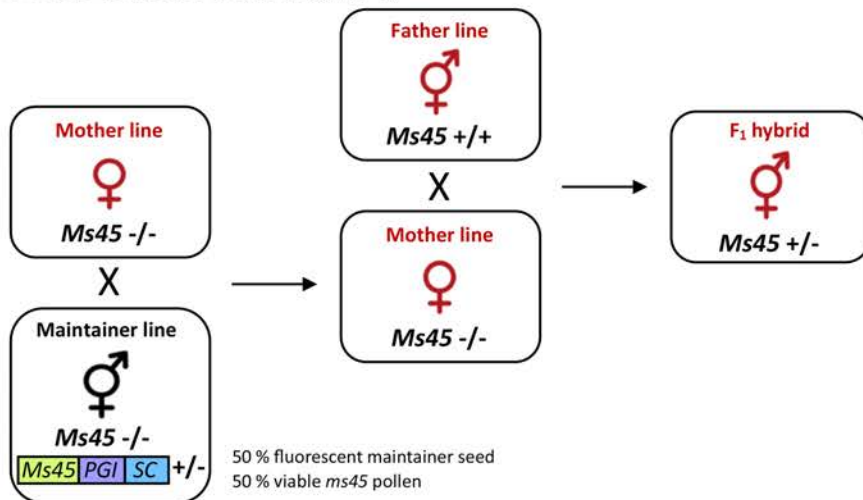


Figure 4. Simplified schematic representation of major existing hybridization systems that utilise transgenes. Red symbols indicate transgenic individuals. **(A)** Seedlink™ gene system; the mother line is made male-sterile by inducing tapetum-specific expression of the bacterial ribonuclease *BARNASE* and maintained by crossing with a fertile plant carrying no transgene. Further mother lines can be selected from progeny seeds due to herbicide resistance (H^R) present on the transgene. Fertility is restored in the F_1 hybrid seeds when the mother line is crossed with a father line expressing the cell-specific ribonuclease inhibitor *BARSTAR*. **(B)** *BECLIN1* - *COP1* system; the female cassette in the mother line contains a regulatory component expressing the conjugated protein $HFR1^{NT131}$ – TBPm3 which binds to the TATA-box mutated promoter ($P^{TA29_{(TGTA)}}$). This promoter controls tapetum-specific expression of the male-sterility gene *BECLIN1*. The father line contains a male cassette driving tapetum-specific expression of *COP1* which has no effect on fertility. In F_1 progeny, *COP1* protein physically interacts with $HFR1^{NT131}$ and degrades it, resulting in abolition of *BECLIN1* expression and thus fertility restoration. **(C)** In the Seed Production Technology (SPT) hybridization system the mother line is homozygous for a recessive male sterile locus (e.g. *Ms45* in maize). This system relies on the transgene present in the maintainer line which contains one copy of a dominant fertility restorer (*Ms45*), a pollen germination inhibitor (PGI) and a seed colour marker (SC). Thanks to the PGI the transgene allows to propagate the male sterile mother lines while confining the transgene to the maintainer line. Crossing of the mother line with a father line homozygous for the dominant allele of the male sterility allows production of non-transgenic and fully fertile F_1 seeds.

The tailoring of a wheat plant – “precision breeding”

One of the biggest challenges in hybrid wheat breeding is ensuring sufficient seed set on the female parental line in the hybrid seed production field. Seed set is restricted due to the inherent structure and development of the wheat inflorescence and the requirement of wind to disperse pollen from the male to the female parent. To effectively maximise cross-pollination, it will be necessary to develop populations with specific and appropriate characteristics for both the male and female parent plants [72,149]. This in turn requires targeted floral modifications (**Figure 5**).

The difficulties of a wheat floret

In wheat, the reproductive unit is the spikelet composed of two bracts, namely glumes, that enclose a variable number of individual florets, usually between two and five [150]. Each floret is in turn composed of two leaf-like structures: an outer lemma and an internal palea that envelope two lodicules, three stamens and one carpel containing a single ovule which will develop into the seed upon fertilization [151]. In wheat, the major determinant impeding cross-pollination is the fact that anthesis – defined as anther extrusion from the floret – usually occurs after the pollen has already been shed inside the floret [152]. Additionally, anthesis is very short and only lasts between 8 minutes to a few hours [11,153]. Floret opening in wheat typically occurs in the morning, succeeded by another round of opening in the afternoon depending on genotype and environmental conditions [11]. Flowering and pollen dispersal are also heavily dependent on environmental factors, including wind speed and direction, temperature and relative air humidity [154-156].

Other factors hindering cross-pollination are the relatively brief stigma receptivity (6 - 13 days, depending on environmental factors) [58,157,158], the reduced pollen viability (0.5 - 3 hours) [159] and low pollen production compared to other cereals. Moreover, pollen weight and shape also play major roles as it has been reported that most of the released wheat pollen falls within three metres of its origin. Consequently, outcrossing rates decrease the farther away from the pollen source and can result in cross-fertilization rates reaching approximately 10 % under optimum conditions while being as low as 0.1 % in unfavourable circumstances [158,160]. These rates are too low and variable for a successful commercial hybrid breeding program.

Some of the wheat floral traits are a likely consequence of domestication as wheat breeding has selected heavily for yield (a product of self-fertility) and counter-selected for outcrossing traits typical of wheat progenitors (*T. monococcum*, *A. speltooides*, *T. tauschii*). Reverting wheat to an obligate outcrosser yet retaining favourable agronomic and quality traits is expected to be challenging and will require the identification and successive stacking of genes that would contribute to cross-pollination.

Characteristics of a male population

Development of the male population should focus on excellent pollen-donating qualities (**Figure 5**). Ideally, male pollinators would produce long spikes containing many spikelets. A good starting point to achieve this is by considering wheat's genetic diversity.

In European wheat varieties, anther extrusion appears to be a quantitative trait influenced by several loci having a small to modest effect. The most significant marker associated with this trait was located in the *Rht-D1* gene, with *Rht-D1a* allele resulting in about 17 % increase in anther extrusion [161]. An analysis of wheat F₂ mapping populations identified multiple QTLs for anther length, spikelet number and spike length at the photoperiod-sensitive *Ppd-D1* locus, with a positive effect from the *Ppd-D1b* allele [162]. Interestingly, the same study also identified a favourable association between the semi-dwarfing *Rht-B1a* allele and large anthers. Another viable option to introduce good pollinator traits is by looking within related species. Addition of rye (cv Imperial) chromosome 4R to wheat (cv Chinese Spring) has been reported to increase suitable traits such as anther length and pollen grain number [163]. Another crucial aspect to consider is synchronizing the flowering time between male and

female lines. In this regard, different flowering patterns could be achieved by manipulating expression and combination of the vernalization (*VRN*) and photoperiodic (*PPD*) genes.

Hybrid seed production can also be facilitated by adjusting the relative height of male and female plants. The male lines would need to be taller than female lines so that pollen can readily spread to the shorter female plant (**Figure 5A**). Plant height can be controlled by exploiting the portfolio of *Rht* genes, introduced and widely distributed in wheat breeding programs during the Green Revolution. Although *Rht-B1* and *Rht-D1* are the two major wheat homeoloci used to reduce plant height, it has been shown that *Rht-D1* also behaves as a major QTL for anther extrusion [164,165], with a negative effect for the height-reducing *Rht-1b* alleles [166]. A recent study highlighted how a novel *Rht24* locus on chromosome 6A is best suited for male plants owing to no pronounced effect on anther extrusion [166]. Predominant use of *Rht24* in the male population with concurrent expression of multiple *Rht* genes in the female population can guarantee relative taller male lines with superior cross-pollination characteristics, but still within the plant height accepted by farmers.

Characteristics of a female population

Apart from relative plant height, other desired qualities for a population of female plants would be optimal pollen-receptive abilities (**Figure 5**). In this context, increasing receptivity, density and length of stigmatic hairs is an interesting research area. This could initially be approached by introgression from durum wheat which generally shows more extended and more crowded stigma hairs into bread wheat. Other ways to increase the chances of successful cross-pollination could come from having supernumerary carpels within the floret

and a more open floret. Steps in understanding the mechanism of flower opening besides lodicule swelling have recently been made [167].

Mutant resources: the joker in the pack

The generation of male and female populations for hybrid breeding entails modifying multiple aspects of the plant structure, particularly the floral architecture. A situation that is challenged by the complexity, size and ploidy of the wheat genome. Being hexaploid (AABBDD), wheat contains triplicate copies (homeologs) of each gene derived from the A, B and D genomes, respectively. Thus, genic redundancy often confounds obtaining mutants with a visible phenotype. In this context, turning to mutant collections of relatively simpler cereals represents a useful resource to exploit. For example, information on flower development in rice is far more advanced and may inform crucial flower development processes in wheat. Barley has a related but simpler and smaller genome constitution than wheat along with numerous genetic mutant resources, facilitating discovery of many genes pertinent to fertility and floral architecture that are also relevant for wheat. Of particular interest to hybrid breeding are mutants whose florets offer an improved or obligate cross-pollination and have the potential to produce more seeds.

Several mutants have been described which appear to possess the double advantages of being male-sterile and of being able to support multiple seed set. In rice, the recessive mutant *dwarf and deformed flower 1-1 (ddf1-1)* shows conversion of lodicules into glume-like organs and transformation of stamens into pistil-like organs [168]. A similar phenotype is observed in the recessive rice mutant *superwoman1 (spw1)* [169], and its maize ortholog *silky1 (si1)* [170]. Likewise, *multiovary (mov)* mutants in barley [171-174] and the mutated

dms locus in wheat cv. Zhoumai 18 [175] also develop flowers exhibiting supernumerary carpels at the expense of stamens. In wheat, an interesting spontaneous mutant line is Three Pistil (TP) [176]. Instead of a single carpel, TP florets contain three fully fertile carpels as well as three stamens, a trait controlled by a single dominant gene *Pis1* located on chromosome 2DL [177,178]. Although TP plants consistently produce increased total grain weight per floret and increased seed set, seed size is generally smaller when seeds simultaneously develop within the same floret [177,178].

The compromise between seed number and seed size observed in wheat TP lines highlights that potential pleiotropic adverse effects need to be considered and avoided before introducing mutants into a hybrid breeding context. For example, DDF1-1 encodes an F-box protein involved in both vegetative growth and floral organ specification, thus all organs except spikelets are significantly reduced in size in the mutant. The additional carpels produced in *spw1* and *si1* florets are sterile, thus unable to produce any seeds. The *dms* locus also affects plant height, male sterility and tiller number. Counter-acting or modulating unwanted effects is possible by studying these mutants and understanding the effects of these mutations on the plant organism.

Other aspects to consider are spike morphology and the arrangement of spikelets along the spike (**Figure 5B**). In the barley *laxatum* (*lax*) mutant the rachis internodes are lengthened. As a result, the spike is longer, spike density is reduced and spikelets are well-spaced [179]; all qualities that would be beneficial in a male parent population to enable florets to open widely and to promote pollen dispersal. Conversely, female plants would require a different spike morphology such as an increased floret aperture and branched inflorescence to raise

the number of florets per unit area. Boosting floret numbers would also maximise the opportunities of successful pollination in a hybrid seed production field.

In this context, the *super opening flower 1 (sof1)* phenotype found in barley would be of interest to hybrid breeding [180]. The *sof1* mutant, originated by γ -irradiation, was found to develop larger and longer lodicules than wild type thus pushing the lemma and palea wider apart and resulting in an overall more open floret structure. Regarding mutants affecting inflorescence branching, the most notable examples come from the tetraploid “Miracle wheat” and corresponding barley orthologue *compositum2 (com2)* which display branched spikes [181]. The “Miracle wheat” phenotype appears to have arisen by a single event mutation in an AP2/ERF transcription factor on chromosome 2AS, selected for during the domestication process, and is associated with an increase in spike weight and grain number. Mutation of the same AP2/ERF transcription factor at the *multi-rowed spike (mrs)* locus in hexaploid wheat gives rise to multirow spikes, in which a cluster of spikelets forms at each rachis node instead of a single spikelet [182]. Another notable example of supernumerary spikelets in wheat is paired spikelets, characterized by the formation of a second complete or rudimentary spikelet at each rachis node, reported to be influenced by relative expression levels of *Ppd-1*, *FLOWERING LOCUS T (TaFT)* and *TEOSINTE BRANCHED1 (TB1)* genes [183,184]. Taken together, these and many other cases highlight the importance of studying floral-related mutants and the need of fundamental research in flower and seed development.

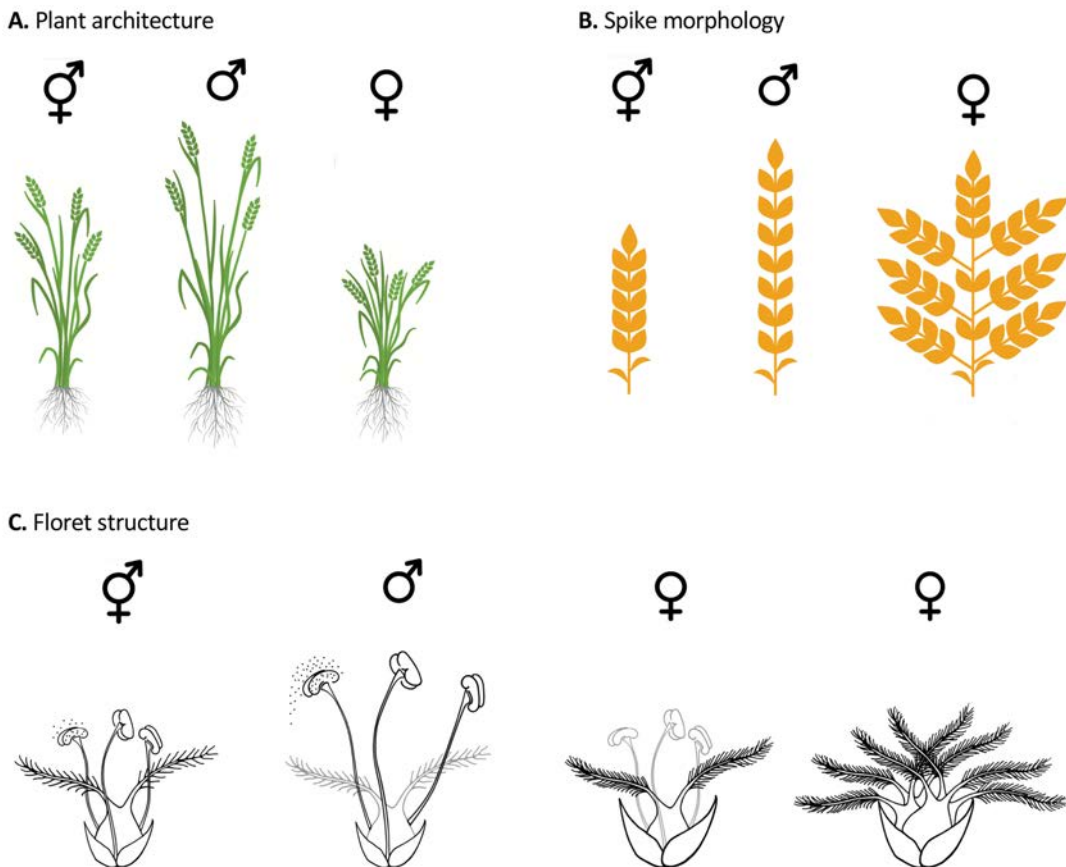


Figure 5. Schematic of male and female wheat parents combining desired characteristics for the **(A)** plant architecture, **(B)** spike morphology and **(C)** floret structure to enhance cross-pollination. **(A)** Plant architecture: the male parent would be taller than the female parent so that pollen can fall directly onto the female plant. **(B)** Spike morphology: the male spike would ideally contain many well-spaced spikelets, while a branched female spike would increase the number of carpels per unit area. At the **(C)** floret structure, the male parent would have optimal pollen-donating characteristics exhibited by increased anther length, anther extrusion, anther size and greater pollen number. On the other hand, the female parent would possess optimal pollen-receptive qualities such as increased density of stigmatic hairs or supernumerary carpels and increased floret aperture due to bigger lodicules.

The future of breeding strategies

The effect on outcrossing from modifications in floret and spike architecture can be maximised with a restructuring of breeding practices. Current hybrid wheat breeding

systems typically involve planting alternate rows of male-sterile and self-fertile pollinator plants. F₁ seeds are only harvested from the rows of male-sterile plants. Optimising outcrossing rates would require reducing the distance between the pollinator and male-sterile plants. This could be achieved by interplanting pollinator plants amongst male-sterile plants, as is routinely done in rye. However, this strategy implies that the harvested product would be composed primarily of hybrid seeds, but still contain a small fraction derived from the selfed pollinator. Thus, the purity of the harvest would be compromised and would require seed mixing to be below 5 % to be compliant with current hybrid seed purity regulations. Consequently, this introduces the need of a seed selection step to distinguish hybrid seeds from pollinator inbred seeds, for example by using a seed colour marker or herbicide tolerance. Even more promising would be the abolition of seed set on the pollinator by breeding female-sterile plants, although this would then require increased breeding efforts in propagating similar lines.

Other aspects to consider for a competitive hybrid breeding strategy is to maximise heterosis. Ideally, this could be done by increasing the genetic diversity of the parents [52], as the amount of genetic variance available for hybrid breeding was reported to contribute the largest effect in the efficiency of hybrid vs line breeding [59]. Specifically, even a moderate increase of heterosis from 10 % to 15 % would be enough to ensure a higher predicted future yield potential for hybrids relative to line breeding for the next 25 years [59].

The genetic divergence of breeding material can be increased by establishing heterotic pools – the grouping of genetically distinct and suitable elite wheat germplasms [185-187] and this

information can then be used to select the most appropriate parental lines. Approaches to establish heterotic pools have been suggested [187,188] and new strategies are constantly being explored and experimentally verified to identify high-yielding heterotic patterns suitable for self-pollinating crops [188].

Heterotic effect can also be predicted with genomic selection (GS) which uses genome-wide molecular markers and a training population to predict phenotypes of complex traits like grain yield using statistical models. GS has already been successfully deployed in wheat [189,190] and maize [191] and has high potential to speed up the breeding process [192]. However, the current accuracy of GS in wheat is still limited by the size of the training population, the relatedness between the training and the testing set and the insufficient models for multi-environment predictions.

Challenges and perspectives

Ultimately, the key to implementing wheat hybrid breeding on a large scale is the design of a facilitated, sustainable and robust hybridization process. This would not only entail finding a suitable fertility control system and investing major efforts in breeding appropriate female and male lines but would also require restructuring current breeding programs.

The study and use of mutants in floral and spike morphology, together with the increasing power of hybrid prediction using genome-wide approaches can contribute new tools towards improving the long-term performance of hybrids relative to conventional line breeding. Already, hybrid wheat offers benefits in terms of yield and yield stability, with the

added opportunity of including tolerance traits to diseases, stresses and herbicides. The greater resilience and adaptability shown by hybrid wheat to marginal land promises to return more reliable and predictable yields in a wider range of environmental conditions. Higher yields will also play a crucial role in lowering hybrid seed rates compared to conventional varieties, offsetting the higher hybrid seed production costs. These attributes have spurred a renewed interest and investments from major seed companies in new generation hybrids, expected to be released in Europe, North America and Australia in the early 2020s.

Acknowledgments

We are grateful to Margaret Pallotta for critical revision of the manuscript.

References

1. **Roser M.** Our World in Data. 2019. <https://ourworldindata.org>
2. **FAOSTAT.** 2019. <http://www.fao.org/faostat/en/#home>
3. **Simkin AJ, López-Calcagno PE, Raines CA.** Feeding the world: improving photosynthetic efficiency for sustainable crop production. *Journal of Experimental Botany.* 2019;70:1119–40.
4. **Han M, Okamoto M, Beatty PH, Rothstein SJ, Good AG.** The genetics of nitrogen use efficiency in crop plants. *Annu. Rev. Genet.* 2015;49:269–89.
5. **Asif MA, Schilling RK, Tilbrook J, Brien C, Dowling K, Rabie H, et al.** Mapping of novel salt tolerance QTL in an Excalibur × Kukri doubled haploid wheat population. *Theor Appl Genet.* 2018;131:2179–96.
6. **Kochian LV, Piñeros MA, Liu J, Magalhaes JV.** Plant adaptation to acid soils: the molecular basis for crop aluminum resistance. *Annu. Rev. Plant Biol.* 2015;66:571–98.
7. **Tricker PJ, ElHabti A, Schmidt J, Fleury D.** The physiological and genetic basis of combined drought and heat tolerance in wheat. *Journal of Experimental Botany.* 2018;69:3195–210.
8. **Fiévet JB, Nidelet T, Dillmann C, de Vienne D.** heterosis is a systemic property emerging from non-linear genotype-phenotype relationships: evidence from *in vitro* genetics and computer simulations. *Front. Genet.* 2018;9:553–26.
9. **Duvick DN.** Biotechnology in the 1930s: the development of hybrid maize. *Nat Rev Genet.* 2001;2:69–74.
10. **Crow JF.** 90 years ago: the beginning of hybrid maize. *Genetics.* 1998;148:923–8.
11. **Virmani SS, Edwards IB.** Current status and future prospects for breeding hybrid rice and wheat. *Advances in Agronomy.* 1983. pp. 145–214.
12. **Longin CFH, Mühleisen J, Maurer HP, Zhang H, Gowda M, Reif JC.** Hybrid breeding in autogamous cereals. *Theor Appl Genet.* 2012;125:1087–96.
13. **Darwin C.** Effects of cross and self fertilization in the vegetable kingdom. 2nd ed. London: Murray; 1900.
14. **Bruce AB.** The mendelian theory of heredity and the augmentation of vigor. *Science.* 1910;32:627–8.
15. **Davenport CB.** Degeneration, albinism and inbreeding. *Science.* 1908;28:454–5.
16. **Jones DF.** Dominance of linked factors as a means of accounting for heterosis. *Genetics.* 1917;2:466–79.
17. **East EM.** Heterosis. *Genetics.* 1936;21:375–97.
18. **Shull GH.** The composition of a field of maize. *J Hered.* 1908.
19. **Crow JF.** Alternative hypotheses of hybrid vigor. *Genetics.* 1948;33:477–87.
20. **Hull FH.** Recurrent selection for specific combining ability in corn. *Journal of the American Society of Agronomy.* 1945;37:134–45.
21. **Powers L.** An expansion of Jones's theory for the explanation of heterosis. *American Society of Naturalists.* 1944;78:275–280
22. **Richey FD.** Mock-dominance and hybrid vigor. *Science.* 1942;96:280–1.
23. **Williams W.** Heterosis and the genetics of complex characters. *Nature.* 1959;184:527–30.
24. **Li L, Lu K, Chen Z, Mu T, Hu Z, Li X.** Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics.* 2008;180:1725–42.
25. **Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, et al.** Overdominant quantitative trait loci for yield and fitness in tomato. *Proc Natl Acad Sci USA.* 2006;103:12981–6.
26. **Xiao J, Li J, Yuan L, Tanksley SD.** Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics.* 1995;140:745–54.
27. **Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R, et al.** Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics.* 2001;158:1737–53.
28. **Luo LJ, Li ZK, Mei HW, Shu QY, Tabien R, Zhong DB, et al.** Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics.* 2001;158:1755–71.
29. **Schön CC, Dhillon BS, Utz HF, Melchinger AE.** High congruency of QTL positions for heterosis of grain yield in three crosses of maize. *Theor Appl Genet.* 2009;120:321–32.
30. **Larièpe A, Mangin B, Jasson S, Combes V, Dumas F, Jamin P, et al.** The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among

- traits of agronomical interest in maize (*Zea mays* L.). *Genetics*. 2012;190:795–811.
31. **Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES.** Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*. 1992;132:823–39.
32. **Osborn TC, Chris Pires J, Birchler JA, Auger DL, Jeffery Chen Z, Lee H-S, et al.** Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics*. 2003;19:141–7.
33. **Song R, Messing J.** Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc Natl Acad Sci USA*. 2003;100:9055–60.
34. **Kawanabe T, Ishikura S, Miyaji N, Sasaki T, Wu LM, Itabashi E, et al.** Role of DNA methylation in hybrid vigor in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 2016;113:E6704–11.
35. **Zhang Q, Li Y, Xu T, Srivastava AK, Wang D, Zeng L, et al.** The chromatin remodeler DDM1 promotes hybrid vigor by regulating salicylic acid metabolism. *Nature*. 2016;2:1–12.
36. **Zhao X, Chai Y, Liu B.** Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Science*. 2007;172:930–8.
37. **Nakamura S, Hosaka K.** DNA methylation in diploid inbred lines of potatoes and its possible role in the regulation of heterosis. *Theor Appl Genet*. 2009;120:205–14.
38. **Qi X, Li ZH, Jiang LL, Yu XM, Ngezahayo F, Liu B.** Grain-yield heterosis in *Zea mays* L. shows positive correlation with parental difference in CHG methylation. *Crop Science*. 2010;50:2338–10.
39. **Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, et al.** Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009;461:1130–4.
40. **Chen ZJ, Ni Z.** Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays*. 2006;28:240–52.
41. **Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES.** Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci*. 2012;109:7109–14.
42. **Saeki N, Kawanabe T, Ying H, Shimizu M, Kojima M, Abe H, et al.** Molecular and cellular characteristics of hybrid vigor in a commercial hybrid of Chinese cabbage. *BMC Plant Biol*. 2016;16:45–15.
43. **Zhu A, Greaves IK, Liu P-C, Wu L, Dennis ES, Peacock WJ.** Early changes of gene activity in developing seedlings of *Arabidopsis* hybrids relative to parents may contribute to hybrid vigor. *Plant Journal*. 2016;88:597–607.
44. **Song G-S, Zhai H-L, Peng Y-G, Zhang L, Wei G, Chen X-Y, et al.** Comparative transcriptional profiling and preliminary study on heterosis mechanism of super-hybrid rice. *Molecular Plant*. 2010;3:1012–25.
45. **Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, et al.** Altered circadian rhythms regulate growth vigor in hybrids and allopolyploids. *Nature*. 2009;457:327–31.
46. **Ng DW-K, Lu J, Chen ZJ.** Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Current Opinion in Plant Biology*. 2012;15:154–61.
47. **Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, et al.** Repeat associated small RNAs vary among parents and following hybridization in maize. *Proc. Natl. Acad. Sci*. 2012;109:10444–9.
48. **Shen Y, Sun S, Hua S, Shen E, Ye C-Y, Cai D, et al.** Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *Plant Journal*. 2017;91:874–93.
49. **Lippman ZB, Zamir D.** Heterosis: revisiting the magic. *Trends in Genetics*. 2007;23:60–6.
50. **Jiang Y, Schmidt RH, Zhao Y, Reif JC.** A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nature Genetics*. 2017;49:1741–6.
51. **Birchler JA, Auger DL, Riddle NC.** In search of the molecular basis of heterosis. *Plant Cell*. 2003;15:2236–9.
52. **Coors JG, Pandey S.** The Genetics and Exploitation of Heterosis in Crops. American Society of Agronomy. 1999.
53. **Duvik D.** Heterosis: feeding people and protecting natural resources. The Genetics and Exploitation of Heterosis in Crops. International Symposium on the genetics and exploitation of heterosis in crop. CIMMYT, Mexico. 1997.
54. **Chen ZJ.** Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*. 2010;15:57–71.
55. **Janick J.** Hybrid rice: genetics, breeding, and seed production. Janick J, editor. *Plant Breeding Reviews*. John Wiley & Sons; 2010. pp. 15–58.
56. **Jiming Li YXLY.** Hybrid rice technology development: Ensuring China's food security. *Intl Food Policy Res Inst*; 2009.
57. **Freeman GF.** The heredity of quantitative characters in wheat. *Genetics*. 1919;4:1–93.
58. **Pickett AA.** Hybrid wheat: results and problems. *Fortschritte der Pflanzenzuchtung*; 1993.

59. Longin CFH, Reif JC, Würschum T. Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. *Theor Appl Genet.* 2014;127:1635–41.
60. Mühleisen J, Piepho H-P, Maurer HP, Longin CFH, Reif JC. Yield stability of hybrids versus lines in wheat, barley, and triticale. *Theor Appl Genet.* 2014;127:309–16.
61. Longin CFH, Gowda M, Mühleisen J, Ebmeyer E, Kazman E, Schachschneider R, *et al.* Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. *Theor Appl Genet.* 2013;126:2791–801.
62. Thorwarth P, Piepho HP, Zhao Y, Ebmeyer E, Schacht J, Schachschneider R, *et al.* Higher grain yield and higher grain protein deviation underline the potential of hybrid wheat for a sustainable agriculture. *Plant Breed.* 2018;137:326–37.
63. Shewry PR, Hey SJ. The contribution of wheat to human diet and health. *Food Energy Secur.* 2015;4:178–202.
64. Simmonds DH. Wheat and wheat quality in Australia. CSIRO Publishing; 1989. pp. 1–299.
65. Glithero NJ, Wilson P, Ramsden SJ. Straw use and availability for second generation biofuels in England. *Biomass Bioenergy.* 2013;55:311–21.
66. Karlsson H, Ahlgren S, Sandgren M, Passoth V, Wallberg O, Hansson P-A. A systems analysis of biodiesel production from wheat straw using oleaginous yeast: process design, mass and energy balances. *Biotechnol Biofuels.* 2016;9:229–13.
67. Kaparaju P, Serrano M, Thomsen AB, Kongjan P, Angelidaki I. Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. *Bioresour. Technol.* 2009;100:2562–8.
68. Singh SK, Chatrath R, Mishra B. Perspective of hybrid wheat research: A review. *Indian Journal of Agricultural Sciences.* 2010;80:1013–27.
69. Wells DG, Caffey HR. Scissor Emasculation of Wheat and Barley. *Experimental Agriculture.* 1994;30:110–0.
70. Parodi PC, de los Angeles Gaju M. Male sterility induced by the chemical hybridizing agent clofencet on wheat, *Triticum aestivum* and *T. turgidum* var. *durum*. *Ciencia e investigación agraria.* 2009;36:267–76.
71. Adugna A, Nanda GS, Singh K, Bains NS. A comparison of cytoplasmic and chemically-induced male sterility systems for hybrid seed production in wheat (*Triticum aestivum* L.). *Euphytica.* 2004;135:297–304.
72. Whitford R, Fleury D, Reif JC, Garcia M, Okada T, Korzun V, *et al.* Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *Journal of Experimental Botany.* 2013;64:5411–28.
73. Wang S, Zhang Y, Song Q, Fang Z, Chen Z, Zhang Y, *et al.* Mitochondrial dysfunction causes oxidative stress and tapetal apoptosis in chemical hybridization reagent-induced male sterility in wheat. *Front. Plant Sci.* 2017;8:2217.
74. Song Y, Wang J, Zhang P, Zhang G, Zhang L, Zhao X, *et al.* Cytochemical investigation at different microsporogenesis phases of male sterility in wheat, as induced by the chemical hybridising agent SQ-1. *Crop Pasture Sci.* CSIRO Publishing. 2014;65:868–77.
75. Wang S, Zhang G, Song Q, Zhang Y, Li Z, Guo J, *et al.* Abnormal development of tapetum and microspores induced by chemical hybridization agent SQ-1 in wheat. *PLoS ONE.* 2015;10:e0119557.
76. Kempe K, Gils M. Pollination control technologies for hybrid breeding. *Mol Breeding.* 2011;27:417–37.
77. Chase CD, Ribarits A, Heberle-Bors E. Male Sterility. Pua EC, Davey MR, editors. *Plant Developmental Biology - Biotechnological Perspectives: Volume 1.* Springer Berlin. 2010. pp. 437–57.
78. Bing-Hua L, Jing-Yang D. A dominant gene for male sterility in wheat. *Plant Breed.* 1986;97:204–9.
79. Maan SS, Carlson KM, Williams ND, Yang T. Chromosomal arm location and gene-centromere distance of a dominant gene for male sterility in wheat. *Crop Science.* 1987;27:494.
80. Maan SS, Kianian SF. Third dominant male sterility gene in common wheat. *Wheat Information Service.* 2001.
81. Endo TR, Mukai Y, Yamamoto M, Gill BS. Physical mapping of a male-fertility gene of common wheat. *Jpn J Genet.* 1991;66:291–5.
82. Klindworth DL, Williams ND, Maan SS. Chromosomal location of genetic male sterility genes in four mutants of hexaploid wheat. *Crop Science.* 2002;42:1447.
83. Suneson CA. Use of Pugsley's sterile wheat in cross breeding. *Crop Science.* 1962;2:534.
84. Fossati A, Ingold M. A male sterile mutant in *Triticum aestivum*. *Wheat Information Service.* 1970;8–10.
85. Driscoll CJ. Induction and use of the “Cornerstone” male-sterility in wheat. 1978. pp. 499–502.
86. Sasakuma T, Maan SS, Williams ND. EMS-induced male-sterile mutants in euplasmic and alloplasmic common wheat. *Crop Science.* 1978;18:850–3.

87. **Zhou K, Wang S, Feng Y, Ji W, Wang G.** A new male sterile mutant LZ in wheat (*Triticum aestivum* L.). *Euphytica*. 2007;159:403–10.
88. **Tucker EJ, Baumann U, Kouidri A, Suchecki R, Baes M, Garcia M, et al.** Molecular identification of the wheat male fertility gene *Ms1* and its prospects for hybrid breeding. *Nature Communications*. 2017;8:869–810.
89. **Kouidri A, Baumann U, Okada T, Baes M, Tucker EJ, Whitford R.** Wheat *TaMs1* is a glycosylphosphatidylinositol-anchored lipid transfer protein necessary for pollen development. *BMC Plant Biol*. 2018;18:332–13.
90. **Wang Z, Li J, Chen S, Heng Y, Chen Z, Yang J, et al.** Poaceae-specific *MS1* encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl. Acad. Sci*. 2017;114:12614–9.
91. **Pallotta MA, Warner P, Kouidri A, Tucker EJ, Baes M, Suchecki R, et al.** Wheat *ms5* male-sterility is induced by recessive homoeologous A and D genome non-specific lipid transfer proteins. *Plant Journal*. 2019;99:673–85.
92. **Ni F, Qi J, Hao Q, Lyu B, Luo M-C, Wang Y, et al.** Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species. *Nature Communications*. 2017;8:15121.
93. **Cao W, Somers DJ, Fedak G.** A molecular marker closely linked to the region of *Rht-D1c* and *Ms2* genes in common wheat (*Triticum aestivum*). *Genome*. 2009;52:95–9.
94. **Yang L, Ben Hui Liu, Zhai HQ, Wang SH, Liu HW, Zhou Y, et al.** Dwarf male-sterile wheat: a revolutionary breeding approach to wheat. *Joint FAO/IAEA Programme*. 2009.
95. **Driscoll CJ.** XYZ System of producing hybrid wheat. *Crop Science*. 1972;1–2.
96. **Driscoll CJ.** Modified XYZ system of producing hybrid wheat. *Crop Science*. 1985;25:1115–6.
97. **Zhou K, Wang S, Feng Y, Liu Z, Wang G.** The *4E-ms* system of producing hybrid wheat. *Crop Science*. 2006;46:250–5.
98. **Tian N, Liu ZQ.** Development of dominant nuclear male-sterile lines with a blue seed marker in durum and common wheat. *Plant Breed*. 2001;120:79–81.
99. **Chase CD, Gabay-Laughnan S.** Cytoplasmic male sterility and fertility restoration by nuclear genes. Daniell H, Chase C, editors. *Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and Mitochondria*. Springer Netherlands; 2004. pp. 593–621.
100. **Hanson MR, Bentolila S.** Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell*. 2004;16:S154–69.
101. **Horn R.** Recombination: cytoplasmic male sterility and fertility restoration in higher plants. *Genetics*. 2006;67:31–52.
102. **Budar F, Pelletier G.** Male sterility in plants: occurrence, determinism, significance and use. *C. R. Acad. Sci*. 2001;324:543–50.
103. **Kaul MLH.** Male sterility in higher plants. 10 ed. Springer-Verlag; 1988.
104. **Mukai Y, Tsunewaki K.** Basic studies on hybrid wheat breeding: VIII. A new male sterility-fertility restoration system in common wheat utilizing the cytoplasm of *Aegilops kotschyi* and *Ae. variabilis*. *Theor Appl Genet*. 1979;54:153–60.
105. **Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A.** Male fertility restoration of wheat in *Hordeum chilense* cytoplasm is associated with 6H^{chS} chromosome addition. *Aust. J. Agric. Res.* CSIRO Publishing. 2008;59:206–8.
106. **Castillo A, Atienza SG, Martín AC.** Fertility of CMS wheat is restored by two *Rf* loci located on a recombined acrocentric chromosome. *Journal of Experimental Botany*. 2014;65:6667–77.
107. **Castillo A, Rodríguez-Suárez C, Martín AC, Pistón F.** Contribution of chromosomes 1H^{chS} and 6H^{chS} to fertility restoration in the wheat *msH1* CMS system under different environmental conditions. Budak H, editor. *PLoS ONE*. 2015;10:1–11.
108. **Yen F-S, Evans LE, Larter EN.** Monosomic analysis of fertility restoration in three restorer lines of wheat. *Canadian Journal of Genetics and Cytology*. 1969;11:531–46.
109. **Bahl PN, Maan SS.** Chromosomal location of male fertility restoring genes in six lines of common wheat. *Crop Science*. 1973;13:317–20.
110. **Wilson P, Driscoll CJ.** Hybrid wheat. *Heterosis*. Springer; 1983. pp. 94–123.
111. **Maan SS, Luchen KA, Bravo JM.** Genetic analyses of male-fertility restoration in wheat. I. Chromosomal location of *Rf* genes. *Crop Science*. 1984;24:17–20.
112. **Du H, Maan SS, Hammond JJ.** Genetic analyses of male-fertility restoration in wheat: III. Effects of aneuploidy. *Crop Science*. 1991;31:319–22.
113. **Sinha P, Tomar SMS, Vinod, Singh VK, Balyan HS.** Genetic analysis and molecular mapping of a new fertility restorer gene *Rf8* for *Triticum timopheevi* cytoplasm in wheat (*Triticum aestivum* L.) using SSR markers. *Genetica*. 2013;141:431–41.

114. **Würschum T, Leiser WL, Weissmann S, Maurer HP.** Genetic architecture of male fertility restoration of *Triticum timopheevii* cytoplasm and fine-mapping of the major restorer locus *Rf3* on chromosome 1B. *Theor Appl Genet.* 2017;130:1253–66.
115. **Geyer M, Albrecht T, Hartl L, Mohler V.** Exploring the genetics of fertility restoration controlled by *Rf1* in common wheat (*Triticum aestivum* L.) using high-density linkage maps. *Mol Genet Genomics.* 2018;293:451–62.
116. **Pelletier G, Budar F.** The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. *Curr. Opin. Biotechnol.* 2007;18:121–5.
117. **Davey MR, Anthony P, Power JB, Lowe KC.** Plant protoplasts: status and biotechnological perspectives. *Biotechnology Advances.* 2005;23:131–71.
118. **Ruiz ON, Daniell H.** Engineering cytoplasmic male sterility via the chloroplast genome by expression of {beta}-ketothiolase. *Plant Physiology.* 2005;138:1232–46.
119. **Havey MJ.** The use of cytoplasmic male sterility for hybrid seed production. Daniell H, Chase CD, editors. *Molecular Biology and Biotechnology of Plant Organelles.* Springer Netherlands; 2004. pp. 623–34.
120. **Lukaszewski AJ.** Chromosomes 1BS and 1RS for control of male fertility in wheats and triticales with cytoplasm of *Aegilops kotschyi*, *Ae. mutica* and *Ae. uniaristata*. *Theor Appl Genet.* 2017;130:2521–6.
121. **Hohn CE, Lukaszewski AJ.** Engineering the 1BS chromosome arm in wheat to remove the *Rf^(multi)* locus restoring male fertility in cytoplasm of *Aegilops kotschyi*, *Ae. uniaristata* and *Ae. mutica*. *Theor Appl Genet.* 2016;129:1769–74.
122. **Martin JA, Crawford JH.** Several types of sterility in *Capsicum frutescens*. *Association of Southern Agricultural Workers;* 1950.
123. **Tang JH, Fu ZY, Hu YM, Li JS, Sun LL, Ji HQ.** Genetic analyses and mapping of a new thermo-sensitive genic male sterile gene in maize. *Theor Appl Genet.* 2006;113:11–5.
124. **Shi Y, Zhao S, Yao J.** Premature tapetum degeneration: a major cause of abortive pollen development in photoperiod sensitive genic male sterility in rice. *J. Integr. Plant Biol.* 2009;51:774–81.
125. **Ku S, Yoon H, Suh HS, Chung Y-Y.** Male-sterility of thermosensitive genic male-sterile rice is associated with premature programmed cell death of the tapetum. *Planta.* 2003;217:559–65.
126. **Murai K.** Factors responsible for levels of male sterility in photoperiod-sensitive cytoplasmic male sterile (PCMS) wheat lines. *Euphytica.* 2001;117:111–6.
127. **Tepliakov BI, Maksimenko VP, Chekurov VM.** The influence of decreased temperatures on meiotic disorders in spring wheat. *TSitologiya i genetika.* 1974;8:406–8.
128. **Campbell CA, Davidson HR.** Effect of temperature, nitrogen fertilization and moisture stress on growth, assimilate distribution and moisture use by Manitou spring wheat. *Canadian Journal of Plant Science.* 1979;59:603–26.
129. **Xing QH, Ru ZG, Zhou CJ, Xue X, Liang CY, Yang DE, et al.** Genetic analysis, molecular tagging and mapping of the thermo-sensitive genic male-sterile gene *wtns1* in wheat. *Theor Appl Genet.* 2003;107:1500–4.
130. **Yuan G, Wang Y, Yuan S, Wang P, Duan W, Bai J, et al.** Functional analysis of wheat *TaPaO1* gene conferring pollen sterility under low temperature. *Journal of Plant Biology.* 2018;61:25–32.
131. **Guo RX, Sun DF, Tan ZB, Rong DF, Li CD.** Two recessive genes controlling thermophotoperiod-sensitive male sterility in wheat. *Theor Appl Genet.* 2006;112:1271–6.
132. **Chen X-D, Sun D-F, Rong D-F, Peng J-H, Li C-D.** A recessive gene controlling male sterility sensitive to short daylength/low temperature in wheat (*Triticum aestivum* L.). *J. Zhejiang Univ. Sci. B.* 2011;12:943–50.
133. **Goring DR, Rothstein SJ.** The S-locus receptor kinase gene in a self-incompatible *Brassica napus* line encodes a functional serine/threonine kinase. *Plant Cell.* 1992;4:1273–81.
134. **Rahman MH.** Resynthesis of *Brassica napus* L. for self-incompatibility: self-incompatibility reaction, inheritance and breeding potential. *Plant Breed.* 2005;124:13–9.
135. **de Nettancourt D.** Incompatibility and Incongruity in wild and cultivated plants. Springer; 2001.
136. **Kaothien-Nakayama P, Isogai A, Takayama S.** Self-incompatibility systems in flowering plants. *Plant Developmental Biology - Biotechnological Perspectives: Volume 1.* Springer Berlin; 2010. pp. 459–85.
137. **Langridge P, Baumann U.** Self-incompatibility in the grasses. Franklin-Tong VE, editor. *Self-Incompatibility in Flowering Plants.* Springer; 2008. pp. 275–87.
138. **Mable BK.** Polyploidy and self-compatibility: is there an association? *New Phytol.* 2004;162:803–11.
139. **Paddon CJ, Hartley RW.** Cloning, sequencing and transcription of an inactivated copy of *Bacillus amyloliquefaciens* extracellular ribonuclease (barnase). *Gene.* 1986;40:231–9.

140. **Mariani C, De Beuckeleer M, Truettner J, Leemans J, Goldberg RB.** Induction of male sterility in plants by a chimaeric ribonuclease gene. *Nature*. 1990;347:737.
141. **Mariani C, Gossele V, De Beuckeleer M, De Block M, Goldberg RB, De Greef W, et al.** A chimaeric ribonuclease-inhibitor gene restores fertility to male sterile plants. *Nature*. 1992;357:384.
142. **Burgess DG, Ralston EJ, Hanson WG, Heckert M, Ho M, Jenq T, et al.** A novel, two-component system for cell lethality and its use in engineering nuclear male-sterility in plants. *Plant Journal*. 2002;31:113–25.
143. **Gils M, Marillonnet S, Werner S, Grützner R, Giritch A, Engler C, et al.** A novel hybrid seed system for plants. *Plant Biotechnology Journal*. 2008;6:226–35.
144. **Kempe K, Rubtsova M, Gils M.** Intein-mediated protein assembly in transgenic wheat: production of active barnase and acetolactate synthase from split genes. *Plant Biotechnology Journal*. 2009;7:283–97.
145. **Kempe K, Rubtsova M, Riewe D, Gils M.** The production of male-sterile wheat plants through split barnase expression is promoted by the insertion of introns and flexible peptide linkers. *Transgenic Res*. 2013;22:1089–105.
146. **Kempe K, Rubtsova M, Gils M.** Split-gene system for hybrid wheat seed production. *Proc. Natl. Acad. Sci*. 2014;111:9097–102.
147. **Singh SP, Singh SP, Pandey T, Singh RR, Sawant SV.** A novel male sterility-fertility restoration system in plants for hybrid seed production. *Sci. Rep*. 2015;5:11274.
148. **Jach G, Binot E, Frings S, Luxa K, Schell J.** Use of red fluorescent protein from *Discosoma* sp. (dsRED) as a reporter for plant gene expression. *Plant Journal*. 2001;28:483–91.
149. **Mette MF, Gils M, Longin CFH, Reif JC.** Hybrid breeding in wheat. 12 ed. Oghara Y, Takumi S, Handa H, editors. Tokyo: Springer Open; 2015. pp. 1–421.
150. **Lersten NR.** Morphology and anatomy of the wheat plant. *Wheat and Wheat Improvement*; 1987.
151. **Anderson WK, Garlinge JR.** The wheat book: principles and practice. 2000.
152. **Frankel R, Galun I.** Definitions of sex types in flowering plants. Pollination mechanisms, reproduction, and plant breeding. 1977;11.
153. **De Vries AP.** Flowering biology of wheat, particularly in view of hybrid seed production — A review. *Euphytica*. 1971;20:152–70.
154. **Sun A-Q, Zhang C-Q, Wu C-L, Gao Q-R.** Pollen flow of wheat under natural conditions in the Huanghuai River Wheat Region, China. *GM Crops & Food*. 2015;6:135–49.
155. **Jackson S, Lyford M.** Pollen dispersal models in quaternary plant ecology: assumptions, parameters, and prescriptions. *The Botanical Review*. 1999;65:39–75.
156. **Song Z, Lu B-R, Chen J.** Pollen flow of cultivated rice measured under experimental conditions. *Biodiversity and Conservation*. 2004;13:579–90.
157. **Kirby E.** Botany of the wheat plant. *Bread Wheat: Improvements and Production*. 2002.
158. **Office of the Gene Technology Regulator.** The Biology of *Triticum aestivum* L. em Thell (Bread Wheat). 2nd ed. Feb, 2008.
159. **D'Souza L.** Studies on the suitability of wheat as pollen donor for cross pollination, compared with rye, *Triticale* and *Secalotricum*. *Zeitschrift für Pflanzenzuchtung*. 1970;63:246–69.
160. **Jain SK.** Population structure and the effects of breeding system. Frankel OH, Hawkes JG, editors. *Crop genetic resources for today and tomorrow*. Cambridge University Press. 1975.
161. **Muqaddasi QH, Pillen K, Plieske J, Ganai MW, Röder MS.** Genetic and physical mapping of anther extrusion in elite European winter wheat. Raman H, editor. *PLoS ONE*. 2017;12:e0187744–17.
162. **Okada T, Jayasinghe JEARM, Eckermann P, Watson-Haigh NS, Warner P, Hendrikse Y, et al.** Effects of *Rht-B1* and *Ppd-D1* loci on pollinator traits in wheat. *Theor Appl Genet*. 2019;132:1965–79.
163. **Nguyen V, Fleury D, Timmins A, Laga H, Hayden M, Mather D, et al.** Addition of rye chromosome 4R to wheat increases anther length and pollen grain number. *Theor Appl Genet*. 2015;128:953–64.
164. **Muqaddasi QH, Lohwasser U, Nagel M, Börner A, Pillen K, Röder MS.** Genome-wide association mapping of anther extrusion in hexaploid spring wheat. *PLoS ONE*. 2016;11:e0155494.
165. **Boeven PHG, Longin CFH, Leiser WL, Kollers S, Ebmeyer E, Würschum T.** Genetic architecture of male floral traits required for hybrid wheat breeding. *Theor Appl Genet*. 2016;129:2343–57.
166. **Würschum T, Liu G, Boeven PHG, Longin CFH, Mirdita V, Kazman E, et al.** Exploiting the *Rht* portfolio for hybrid wheat breeding. *Theor Appl Genet*. 2018;131:1433–42.
167. **Okada T, Jayasinghe JEARM, Nansamba M, Baes M, Warner P, Kouidri A, et al.** Unfertilized ovary pushes wheat flower open for cross-pollination. *Journal of Experimental Botany*. 2018;69:399–412.

168. **Duan Y, Li S, Chen Z, Zheng L, Diao Z, Zhou Y, et al.** *Dwarf and deformed flower 1*, encoding an F-box protein, is critical for vegetative and floral development in rice (*Oryza sativa* L.). *Plant Journal*. 2012;72:829–42.
169. **Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, et al.** *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development*. 2003;130:705–18.
170. **Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ.** Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell*. 2000;5:569–79.
171. **Barley Genetics Newsletter.** 2013:pp. 48–223.
172. **Gregory FJ, Purvis ON.** Abnormal flower development in barley involving sex reversal. *Nature*. 1947;160:221–2.
173. **Kamra OP, Nilan RA.** Multi-ovary in barley: floral anatomy and embryo-sac development. *J Hered*. 1959;50:159–66.
174. **Moh CC, Nilan RA.** Multi-ovary in barley: a mutant induced by atomic bomb irradiation. *J Hered*. 1953;44:183–4.
175. **Duan Z-B, Shen C-C, Li Q-Y, Lü G-Z, Ni Y-J, Yu D-Y, et al.** Identification of a novel male sterile wheat mutant *dms* conferring dwarf status and multi-pistils. *Journal of Integrative Agriculture*. 2015;14:1706–14.
176. **Peng Z-S.** A new mutation in wheat producing three pistils in a floret. *Journal of Agronomy and Crop Science*. 2003;189:270–2.
177. **Peng Z-S, Yang J, Wei S-H, Zeng J-H.** Characterization of the common wheat (*Triticum aestivum* L.) mutation line producing three pistils in a floret. *Hereditas*. 2004;141:15–8.
178. **Peng Z-S, Martinek P, Kosuge K, Kuboyama T, Watanabe N.** Genetic mapping of a mutant gene producing three pistils per floret in common wheat. *J. Appl. Genet*. 2008;49:135–9.
179. **Larsson HEB.** Morphological analysis of *laxatum* barley mutants. *Hereditas*. 1985;103:239–53.
180. **Hamada Y, Anwar N, Sakuma S, Ning S, Pourkheirandish M, Koba T, et al.** *Super-open flowering 1* mutant generated by chronic irradiation by gamma ray to wild barley. *International Barley Mutant Workshop*. Dundee, Scotland; 2018.
181. **Poursarebani N, Seidensticker T, Koppolu R, Trautewig C, Gawroński P, Bini F, et al.** The genetic basis of composite spike form in barley and 'Miracle-Wheat'. *Genetics*. 2015;201:155–65.
182. **Dobrovolskaya O, Pont C, Sibout R, Martinek P, Badaeva E, Murat F, et al.** *FRIZZY PANICLE* drives supernumerary spikelets in bread wheat. *Plant Physiology*. 2015;167:189–99.
183. **Boden SA, Cavanagh C, Cullis BR, Ramm K, Greenwood J, Jean Finnegan E, et al.** *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants*. 2015;1:14016.
184. **Dixon LE, Greenwood JR, Bencivenga S, Zhang P, Cockram J, Mellers G, et al.** *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread wheat (*Triticum aestivum*). *Plant Cell*. 2018;30:563–81.
185. **Koekemoer FP, Van Eeden E, Bonjean AP, Angus WJ.** An overview of hybrid wheat production in South Africa and review of current worldwide wheat hybrid developments. *The world wheat book—a history of wheat breeding*. Lavoisier Publishing Paris; 2011;2:907–50.
186. **Reif JC, Hallauer AR, Melchinger AE.** Heterosis and heterotic patterns in maize. *Maydica*. 2005;50:1–10.
187. **Boeven PHG, Longin CFH, Würschum T.** A unified framework for hybrid breeding and the establishment of heterotic groups in wheat. *Theor Appl Genet*. 2016;129:1231–45.
188. **Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, et al.** Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci*. 2015;112:15624–9.
189. **Zhao Y, Zeng J, Fernando R, Reif JC.** Genomic prediction of hybrid wheat performance. *Crop Science*. 2013;53:802–9.
190. **Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC.** Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity*. 2014;112:638–45.
191. **Massman JM, Gordillo A, Lorenzana RE, Bernardo R.** Genomewide predictions from maize single-cross data. *Theor Appl Genet*. 2013;126:13–22.
192. **Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME.** plant breeding with genomic selection: gain per unit time and cost. *Crop Science*. 2010;50:1681–90.

Chapter 2



The ABC of flower development

Flowers occupy an important part of our lives. Throughout history flowers have been a symbolic object and thus occupy a distinct role in human celebrations and rituals, from worship to special occasions, and from birthdays to funerals. Flowers are also used as a source of food and for medicinal purposes. For example, the Chinese hibiscus (*Hibiscus rosa-sinensis*), purple coneflower (*Echinacea purpurea*) and marigold flowers (*Calendula officinalis*) are widely used to support the immune system, digestive system, wound healing or to supplement levels of antioxidants in diets [1-3]. Although flowers are highly regarded as decorative for their aesthetic beauty or used as celebratory symbols to commemorate or donate to loved ones, they govern a crucial role in the lifecycle of flowering plants (Angiosperms). The principal purpose of the flower is to ensure reproduction by providing a mechanism for the union of the male and female gametes. The formation of a flower is one of the distinguishing characteristic of Angiosperms, which are currently the most diverse group of land plants and this diversity is reflected in the vast array of floral structures.

It's a matter of class: the genes behind flower development

Most of our knowledge of the molecular mechanisms governing flower development has been achieved thanks to pioneering studies in the model dicotyledonous plants *Arabidopsis thaliana* and *Antirrhinum majus*. In *Arabidopsis thaliana*, flowers are composed of structures that are organised in a very specific pattern. Moving towards the centre of the flower, these structures include: four green sepals, four white petals of equal size and shape found in an alternate position with respect to the sepals, six uneven stamens and nectaries at their base,

and a gynoecium formed by two fused carpels [4]. Sepals and petals together form the perianth which serves to protect the flower from the unpredictable environment and to attract pollinators. On the other hand, stamens and carpels form the true reproductive units of the flower, whose function is to produce viable gametes for sexual reproduction. In 1991, through characterization of floral homeotic mutants, Coen and Meyerowitz (1991) [5] proposed a simple and elegant genetic model to explain flower development. This model is called the ABC model and postulates that the flower is organized in four concentric whorls, and that the floral organs that develop within each whorl are specified by the action of distinctive classes of genes called the A-, B- and C-class genes. Each gene class is expressed in two adjacent whorls: A-class genes are expressed only in the first and second whorl, B-class genes function in the second and third whorl, whilst C-class genes are confined to the third and fourth whorl (**Figure 1**). An important aspect of the ABC model is that the A and C functions are mutually antagonistic. Each whorl, therefore, is defined by the expression of different combinations of genes.

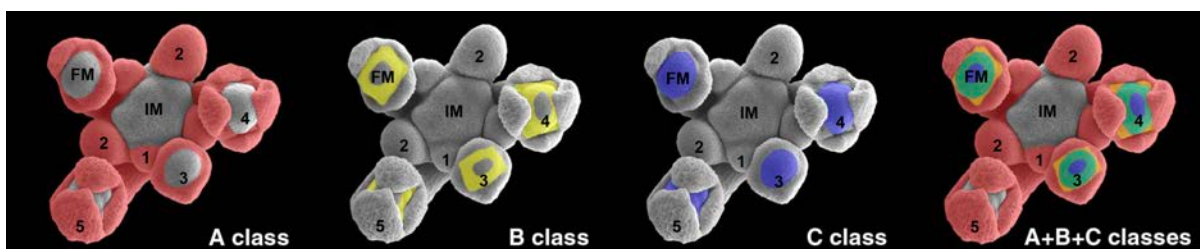


Figure 1. Expression domain of the A- (red), B- (yellow) and C-class (blue) genes during early flower development in *Arabidopsis thaliana*. In the merged image, orange signifies expression overlap between A- and B- class genes, whilst green represent expression overlap between B- and C-class genes. Numbers indicate flowers at different developmental stages, with 1 being the youngest. Inflorescence meristem (IM) and floral meristem (FM) are indicated. Figure adapted from Alvarez-Buylla *et al.* (2010) [4].

Consequently, for each whorl, it is the distinctive interaction between expressed genes that defines the identity of the floral organs that will develop. Specifically in *Arabidopsis*, A-class genes function to specify formation of sepals in the first whorl, the combination of A-class and B-class genes determines petal initiation in the second whorl, in the third whorl B- and C-class genes together give rise to stamens, while C-class genes lead to the development of carpels in the innermost fourth whorl.

In 1995 the ABC model was expanded to also include the D-class genes. Genes belonging to the D-class family work in combination with C-class genes to specify the ovules within the carpel. These genes were initially discovered in *Petunia hybrida* [6], but orthologues were soon found also in *Arabidopsis thaliana* [7,8]. The most current model was only completed in 2000 with the discovery of E-class genes. In *Arabidopsis thaliana* where they were first identified, the E-class genes are expressed in all four whorls and mediate the interactions between the floral organ identity proteins. The E-class members therefore function primarily in establishing the context in which the floral organ identity genes can act [9,10].

Monocotyledonous plants exhibit a variety of floral structures that can differ greatly from the botanically perfect flower of *Arabidopsis*. A notable difference lies in the highly modified floral structure of *Poaceae*, commonly known as grasses. Although the flower structure of grasses is quite different to that of dicotyledons, it still maintains an organization in four concentric whorls. These incorporate bracts called lemma and palea in whorl 1, two lodicules in whorl 2, the androecium comprising stamens in the third whorl and the gynoecium (the pistil) in the fourth whorl. As seen for *Arabidopsis*, the perianth of the *Poaceae* functions to protect the floret and aid gamete encounter. The lodicules are located in whorl 2 and are

positioned at the base of the floret where they swell at flowering. The swollen lodicules physically push the lemma aside, thus allowing opening of the floret which helps anther extrusion and maximises the chances of pollination and fertilization of the ovule [11]. Lodicules in monocotyledon plants are considered to be equivalent to petals in dicotyledonous flowers [12]. Indeed, thanks to phylogenetic analyses and characterization of floral mutants in other species, it was shown that the ABCDE model is widely conserved and valid for all Angiosperms. Recently, the model for flower development was reviewed and a more comprehensive and completed model termed (A)BC was proposed [13]. The revised model incorporates and unifies findings from over 20 years of studies and defines a new function to A-class genes, in which they act to establish the floral meristem identity and set the floral context to enable the B- and C- functions to fulfil their regulatory role over floral organ identity. In this regard, the (A)-function therefore partly overlaps with the function of E-class genes.

Although the ABCDE model is a fundamental component for the reproductive success of Angiosperms, it is still a framework that allows for flexibility and variation. Examples can be found in plants belonging to the genus *Tulipa* [14], *Lilium*, *Helleborus* and *Magnoliidae* in which the B function has expanded from the second and third whorls to the first whorl and the perianth is consequently composed of organs called tepals which can have a petaloid appearance, but show no distinguishable morphological difference between sepals and petals.

However, the most striking variation can be found in the plant *Lacandonia schismatica* (*Triuridaceae*), a rainforest monocotyledon discovered in Mexico, in which the stamens are

located centrally to the flower and are surrounded by the gynoecium formed by 60 - 80 carpels, an extremely rare characteristic that is otherwise known only for the genus of early-divergent Angiosperms *Trithuria* [15,16]. In these cases, the ABCDE model has effectively shifted to ACBDE.

These species-specific variations in the ABCDE model are thought to have arisen mainly from duplications of entire gene families followed by non-functionalization, neo-functionalization, and sub-functionalization of the duplicated genes [17] and is believed to have been a key process during flower evolution. Nonetheless, these observations highlight the fact that despite the great diversity in the size, symmetry, colour, structure and number of organs, all Angiosperms share a basic molecular plan for flower development. It is therefore theoretically possible to create flowers with any particular organ in any particular whorl, simply by altering the expression of the ABCDE genes [18] (**Figure 2**).

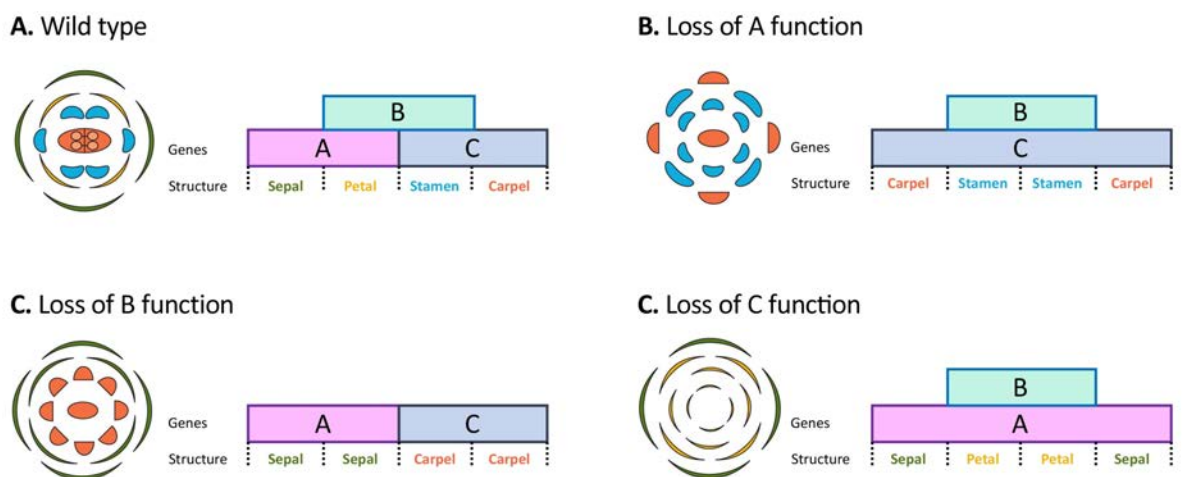


Figure 2. (A) Wild-type flower and homeotic conversion of floral organs in mutants of the (B) A-class, (C) B-class and (D) C-class genes based on the ABCDE model. Figure adapted from Taiz and Zeiger (2010) [18].

Living in a MADS world

The majority of the genes belonging to the ABCDE model, with the exception of genes such as *APETALA2* (*AP2*, *Arabidopsis*) and *DROOPING LEAF* (*DL*, *Oryza sativa*), are transcription factors belonging to the MADS-box protein family. The acronym MADS [19] derives from the initials of the four founding members of this gene family, namely: *MINICHROMOSOME MAINTENANCE1* (*MCM1*, *Saccharomyces cerevisiae*) [20], *AGAMOUS* (*AG*, *Arabidopsis thaliana*) [21], *DEFICIENS* (*DEF*, *Antirrhinum majus*) [22], and *SERUM RESPONSE FACTOR* (*SRF*, *Homo sapiens*) [23]. In plants, MADS-box transcription factors are key regulators of many crucial developmental processes ranging from root, flower, fruit to seed development. For this reason, MADS-box proteins have long been considered master regulators because altering their expression is sufficient to disrupt or trigger an entire developmental program outside of its normal context.

Sequence conservation and protein structure suggest that a gene duplication event that occurred more than a billion years ago prior to the divergence of plants and animals originated two main lineages of MADS-box genes, termed Type I and Type II [24] (**Figure 3**). Type I (SRF-like) genes are found in animals, fungi and plants [25]. Studies conducted in animals have shown that Type I MADS-box genes are involved in the response to growth factors [23]. In plants, the function of Type I MADS-box genes is still largely uncharacterized [26]. However, studies are now starting to unravel the function of these genes and have identified a key regulatory role in plant reproduction, especially in regard to controlling female gametophyte, embryo, and endosperm development [27,28], with an overarching effect on controlling genome dosage and post-zygotic compatibility [29,30].

Type II genes include MEF2-like genes from yeast, fungi and animals, where they play an important role in muscle development in the latter [31], as well as the exclusively plant-specific MIKC-type genes whose name is derived from the protein's characteristic structure in four domains (**Figure 3**). The MIKC-type MADS-box genes consist of a MADS-box domain (M), an intervening domain (I), a K-box (K), and a C-terminal domain (C) [25]. MIKC-type MADS-box proteins can further be divided into MIKC^C-type and the MIKC*-type based on the length and intron/exon structure of the I- and K-domains [32]. Specifically, all MADS-box proteins involved in flower development belong to the MIKC^C-type category. Intriguingly, the MIKC^C-type genes act during all steps of flower development: from the floral transition to ovule and fruit development.

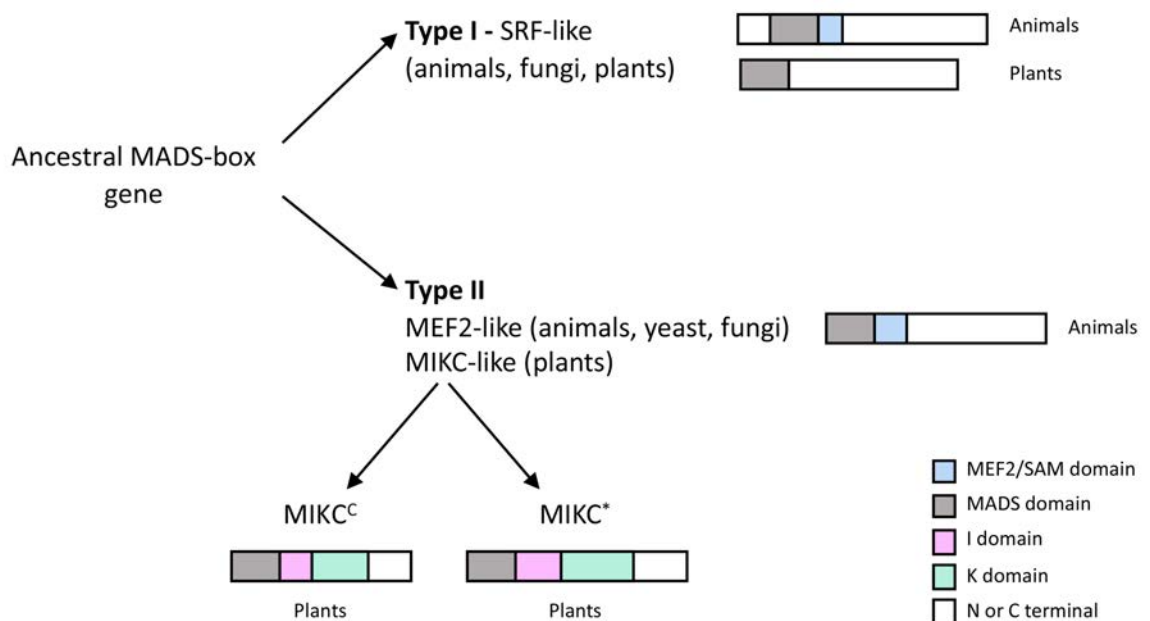


Figure 3. Classification and domain structure of Type I and Type II MADS-box genes. Figure adapted from Alvarez-Buylla *et al.* (2000) [24]; Theissen and Saedler H. (1996) [25].

The MADS domain of approximately 180 base pairs is the most conserved region of the entire protein sequence. Once translated, this domain folds into an α -helix that is required for DNA

binding and two antiparallel β -strands which are involved in dimerization [33] (**Figure 4**). The MADS domain is thus able to recognize and bind a specific DNA consensus sequence called CArG box [CC(A/T)₆GG] usually located in the promoter region of target genes [34,35]. In contrast, the more variable I-domain is required to determine the selective formation of DNA-binding dimers between specific MADS-box proteins [36]. The K-domain is characterized by the presence of three amphipatic α -helices (K1 – K3) which form a second interaction surface to mediate the interactions between MADS-box proteins [37,38]. In *Arabidopsis*, part of the K-domain also acts as a determinant of dimer specificity for B-class genes, together with the I-domain [35,39] (**Figure 4**). Finally, the C-terminal domain can contain specific sequence motifs, especially in the case of B-class genes, but is generally the least conserved domain of MADS-box proteins. Different studies have shown that this domain can function as a trans-activation domain [40-42]. In other cases, the C-terminal domain seems to have a contributing effect in the formation of multimeric complexes of MADS-box proteins [41,43].

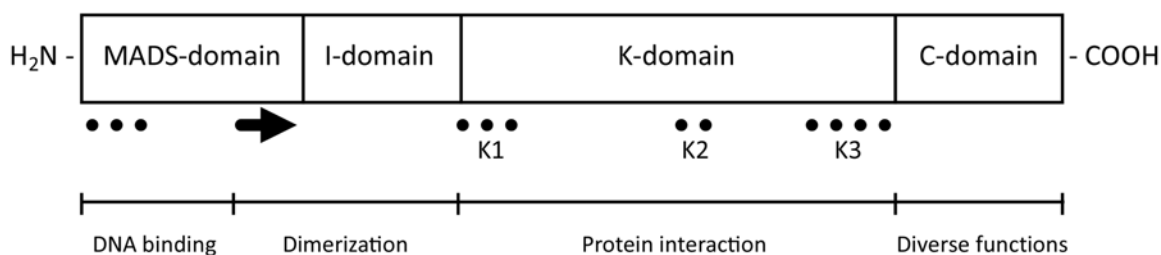


Figure 4. Schematic representation showing structure of MIKC-type MADS proteins and general function for each domain. The conserved secondary elements are shown using dashed lines for α -helices and arrow for β -strand. Figure adapted from Kaufmann *et al.* (2005) [44].

The ability of MADS-box proteins to form higher order complexes led to the postulation of the floral quartet model. In this model, dimers of MADS-box proteins can interact together

to form tetrameric complexes [45]. By binding to two different CA_nG sites, the multimeric transcription factor complex can form DNA looping which brings together distal promoter regions (**Figure 5**). Compelling support for the floral quartet model has recently been provided in *Arabidopsis thaliana* using multiple techniques such as bandshift assays [46] and affinity purification coupled with mass spectrometry [47]. These have allowed to determine the exact identity and stoichiometry of MADS-box proteins forming the complexes. It was thus observed that the floral tetramers also allow recruitment of transcriptional co-factors, chromatin remodelling proteins and other transcription factors to activate or repress the expression of downstream genes [47].

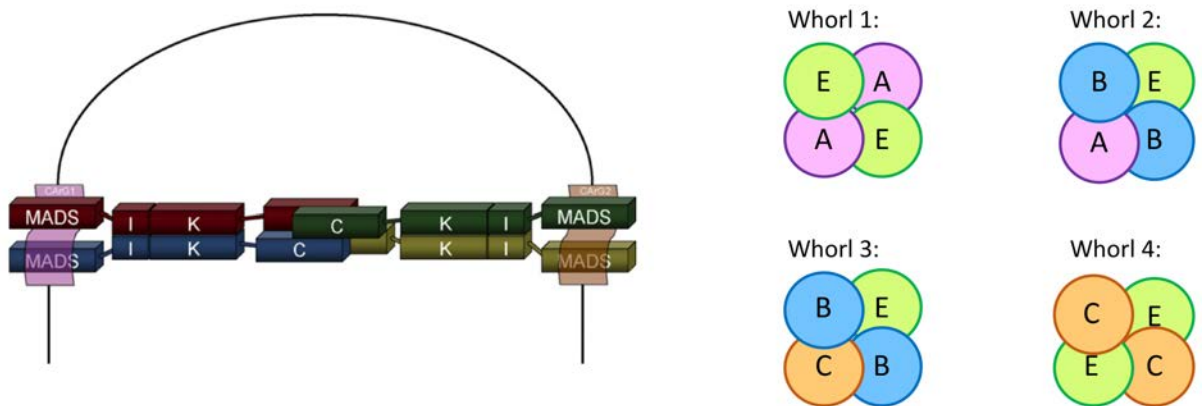


Figure 5. The floral quartet model; dimers of MADS-box protein interact via the C terminals to form tetrameric complexes. This interaction brings together distal CA_nG motifs and promoter elements and creates DNA looping. Composition of the floral quartets depends on the floral whorl. Figure adapted from Causier *et al.* (2010) [13].

The intimate relationship of B-class genes

The amazing variety in size, colour and structure of modern flowers is the result of species-specific fine-tuning of the ABCDE model in order to adapt to the environment and maximise pollination and seed dispersal. However, given their role in determining the reproductive

organs, it is perhaps not surprising that the functions of B- and C-class genes remain the most conserved even between distantly related species. Indeed, correct development of reproductive organs ensures successful reproduction and consequently species' survival.

Within the lineage that led to the Angiosperms, a duplication event occurred ~200 – 300 million years ago of the ancestral B gene gave rise to two clades, termed GLO-like (GLOBOSA-like) and DEF-like (DEFICIENS-like) [48,49]. Evidence of this event can be found in extant Gymnosperms which possess gene subfamilies that are ancestral to the DEF- and GLO-like genes [48], whereas *Amborella trichopoda*, accepted as the most basal lineage in the clade of the Angiosperms, has both DEF-like and GLO-like genes. This indicates that the duplication event of these genes occurred after the split of the Angiosperms from the lineage that led to the extant Gymnosperms [49,50]. The name of these clades derives from the two B-class genes present in *Antirrhinum majus* as they were the first B-function genes to be molecularly characterized [22,51].

In *Antirrhinum*, both *DEF* and *GLO* genes show a similar expression pattern very early during floral development, specifically in the petal (whorl 2) and stamen (whorl 3) primordia and expression continues in these organs as they differentiate [51,52]. In contrast to what is expected from the ABCDE model, *DEF* expression has also been detected in developing carpels, whereas *GLO* levels remain low in this tissue. Similar to *Antirrhinum*, only two B-class genes have been identified in *Arabidopsis*. These genes, termed *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), belong to the DEF-like and GLO-like clades, respectively [53,54]. Interestingly, *AP3* and *PI* show comparable spatial and temporal expression to their *Antirrhinum* orthologues, and mutants in *Antirrhinum* and *Arabidopsis* B-class genes (*def*, *glo*, *ap3* and *pi*) result in a similar phenotype [51-54]. Namely, the mutants exhibit a homeotic

conversion of petals into sepals in the second whorl and stamens into carpels in the third whorl, consistent with the ABC model.

Unlike *Antirrhinum* and *Arabidopsis* that only include one member for each B-class clade, other plant species have undergone lineage-specific duplications followed by neo- and sub-functionalization. For example, in orchids two rounds of gene duplications occurred in the DEF-like clade during early orchid evolution, while retaining a single GLO-like gene. It is believed that the combinatorial interaction of the resulting four DEF-like genes underlies the unique and innovative flower structure of orchids [55]. Within the grasses, duplications have mainly occurred in the GLO-like clade. In rice, there are two GLO-like genes: *OsMADS2* and *OsMADS4* [56]. These genes are mainly expressed in lodicules, stamens, and carpels [56,57]. It has been observed that suppression of *OsMADS2* by RNAi results in a homeotic change of lodicules into bract-like structures, while stamens still develop normally [58]. In contrast, RNAi suppression of *OsMADS4* does not induce any apparent alterations to either lodicules or stamens [59]. Interestingly, the simultaneous loss-of-function in both *OsMADS2* and *OsMADS4* results in conversion of lodicules to palea-like organs and stamens to carpel-like organs [59], suggesting that both genes have a redundant function in stamen formation. These results hint to a sub-functionalization of rice GLO-like genes, as *OsMADS2* seems to play a more important role than *OsMADS4* in lodicule specification. *OsMADS16* (also known as *SUPERWOMAN1 – SPW1*) is the only DEF-like gene in the rice genome [60,61]. Loss of function of *OsMADS16* causes a phenotype in the second and third whorl with conversion of lodicules and stamens into palea-like and carpel-like organs, respectively [61]. Similarly in maize, *SILKY1* – the only DEF-like gene, is required for the normal development of lodicules and stamens [62]. However, sequence and expression data of the three maize GLO-like

genes: *ZMM16*, *ZMM18* and *ZMM29* suggest a functional diversification of GLO-like genes in grasses [63]. *ZMM18/29* mRNAs are expressed in male and female inflorescences and in developing kernels, while even though *ZMM16* is expressed in the same tissues it is also weakly expressed in different vegetative organs [63].

In almost all core eudicots studied so far, DEF- and GLO-like proteins are obligate hetero-dimers (they form functional dimers exclusively with each other). This is distinct from other MADS-domain proteins that are capable of forming either homo-dimers or hetero-dimers [36]. Furthermore, DEF-GLO hetero-dimers initiate a positive autoregulatory feedback loop, in which they activate their own expression, a genetic regulation which is quite rare outside of flower development [52,54,64,65]. Interestingly, in the extant gymnosperm *Gnetum gnemon* the class-B genes (ancestral to both DEF- and GLO-like genes) are only able to bind DNA as homo-dimers in *in vitro* studies [66], while in the Gymnosperm *Picea abies* the class-B genes are able to combine in both homo-dimeric and hetero-dimeric complexes [67]. These observations have led to the hypothesis that the obligate DEF-GLO hetero-dimerization observed in core eudicots evolved several times independently from an ancestral homo-dimerization state after the separation of the clades that led to DEF- and GLO-like genes [66,68]. This shift is thought to have conferred a distinct selective advantage by ensuring a stricter control on the spatio-temporal activation of B-class genes and increasing robustness of the system against unwanted deactivation by chance [65].

In contrast to the core eudicots, a complete array of interactions has been observed in monocotyledonous plants. In the monocot plant *Joinvillea*, GLO-like proteins can bind DNA as a homo-dimer [69], in *Lilium* hetero-dimerization occurs between DEF- and GLO- proteins

and homo-dimerization occurs among GLO proteins [70], while in other species like rice DEF and GLO-like proteins are obligate hetero-dimers. These findings suggest that obligate hetero-dimerization is the more likely ancestral state in the grasses, but that homo-dimerization has recently re-emerged. It also uncovers an unexpected evolutionary lability in B-class MADS-box interactions. Indeed, it has been shown that a single amino acid change is enough to switch the maize GLO-like protein *ZMM16* between homo-dimerization and obligate hetero-dimerization [69]. These observations highlight once more the vast plasticity of MADS-box protein-protein interactions.

Expanding our view: flower development in wheat and barley

While most studies on flower development initially focused on model species such as *Antirrhinum*, *Arabidopsis* and rice, attention is now being turned to a wider variety of plant species. These include economically important crops such as barley and wheat where modified flower structure might be engineered to support novel breeding strategies.

Wheat is an allohexaploid grass with the genome constitution AABBDD. It is a result of the cross between different ancestral diploid species, each contributing a different genome. In particular, the A genome is thought to be derived from *Triticum urartu*, the B genome from *Aegilops speltoides*, and the D genome from *Ae. tauschii* [71]. Consequently, allohexaploid wheat can contain up to three copies of each gene (homeologs) derived from the A, B and D genomes, respectively. This genomic composition makes it hard to obtain mutants with a visible phenotype. On the other hand, it raises intriguing questions about the effect of polyploidization on gene regulatory networks, especially in regard to genetic redundancy, functionalization and possible genomic expression dominance — the effect whereby a

particular sub-genome manifests a differential control of morphological traits, as observed in allotetraploid cotton [72,73]. Some MADS-box genes involved in flower development have already been identified within the wheat genome, and indeed a homeolog-specific expression pattern has been detected for some [74].

On the other hand, cultivated barley derives from domestication of a diploid grass species known as wild barley (*H. spontaneum*). Similarly to wheat, barley represents an important source of calories in human nutrition, is largely used as malt in beverages and food processing and as source of animal feed [75]. The dual role of barley in both human and animal consumption, make it the fourth most important cereal crop in the world after wheat, maize, and rice. Even though the barley genome is still quite large, it is nonetheless smaller and less complex than that of wheat. Given the shared economic importance and the close evolutionary relatedness of wheat and barley, the latter has regularly been considered a model crop species for plant breeding due to its relatively simpler genetics.

In barley and wheat, most of the orthologs of the A-, B-, C-, D- and E-classes have been identified from phylogenetic and sequence analyses [76,77]. Despite this, temporal and spatial expression data for most of the MADS-box genes are still unavailable. Moreover, the lack of confirmed knockout mutants is a limiting factor that represents a major knowledge gap. However, in recent times, mutant resources for both species are expanding and becoming more accessible [78-82]. Additionally, spontaneously occurring mutants also act as potentially useful resources. Of particular interest for hybrid breeding are mutants whose florets offer obligate cross-pollination and have the potential to produce additional seeds. In wheat, an example of such a mutant line is Three Pistil (TP) [83]. Instead of a single carpel as

normally found, TP florets contain three stamens and three fully fertile carpels. Overall, TP plants produce an increased total grain weight per floret; however, the embryos of the two lateral seeds do not mature properly compared to the central seed. Additionally, seed size is generally smaller, probably due to increased competition for nutrients [84,85]. The TP trait was determined to be controlled by a single dominant gene termed *Pis1* located on long arm of chromosome 2D [84,85]. An even more striking phenotype in some wheat lines is pistillody, whereby the stamens are converted fully or partially into pistil-like structures. In wheat, this phenotype is mostly obtained by nuclear-cytoplasm interactions when the cytoplasm of wheat is substituted with the cytoplasm of a wild relative species [86]. Studies report that pistillody in cytoplasmic substitution wheat lines is associated with alterations in the expression pattern of B-class MADS-box genes [87]. Interestingly, an occurrence of pistillody was also recorded in barley for the multiovary (*mov*) mutants at the beginning of the 1950's [88-90]. Since then, little work has been done to understand the genetics behind the pistillody trait in barley.

Perspective

This Chapter provides background information regarding flower development and MADS-box genes in flowering plants. Despite the extensive knowledge which is already available, opportunities exist to (1) examine the function of MADS box genes in barley and wheat, and (2) apply this information to control flower structure in a way that might be suitable for downstream applications in hybrid breeding. These opportunities will be considered in the following Results chapters.

References

1. **Mishra N, Tandon VL, Gupta R.** Immunomodulation by *Hibiscus rosa-sinensis*: effect on the humoral and cellular immune response of *Mus musculus*. *Pak. J. Biol. Sci.* 2012;15:277–83.
2. **Sharifi-Rad M, Mnayer D, Morais-Braga MFB, Carneiro JNP, Bezerra CF, Coutinho HDM, et al.** *Echinacea* plants as antioxidant and antibacterial agents: from traditional medicine to biotechnological applications. *Phytother Res.* 2018;32:1653–63.
3. **Givol O, Kornhaber R, Visentin D, Cleary M, Haik J, Harats M.** A systematic review of *Calendula officinalis* extract for wound healing. *Wound Repair Regen.* 2019.
4. **Alvarez-Buylla ER, Benítez M, Corvera-Poiré A, Chaos Cador Á, de Folter S, Gamboa de Buen A, et al.** Flower development. *The Arabidopsis Book.* 2010;8:e0127.
5. **Coen ES, Meyerowitz EM.** The war of the whorls: genetic interactions controlling flower development. *Nature.* 1991;353:31–7.
6. **Colombo L, Franken J, Koetje E, van Went J, Dons HJ, Angenent GC, et al.** The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell.* 1995;7:1859–68.
7. **Pinyopich A, Ditta GS, Savidge B, Liljegen SJ, Baumann E, Wisman E, et al.** Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature.* 2003;424:85–8.
8. **Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, et al.** MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell.* 2003;15:2603–11.
9. **Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF.** B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature.* 2000;405:200–3.
10. **Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF.** The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 2004;14:1935–40.
11. **Hoshikawa K.** The growing rice plant: an anatomical monograph. *Nosan Gyoson Bunka.* Nobunkyo; 1989;199–205.
12. **Yoshida H.** Is the lodicule a petal: molecular evidence? *Plant Sci.* 2012;184:121–8.
13. **Causier B, Schwarz-Sommer Z, Davies B.** Floral organ identity: 20 years of ABCs. *Seminars in Cell and Developmental Biology.* 2010;21:73–9.
14. **Kanno A, Saeki H, Kameya T, Saedler H, Theissen G.** Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* 2003;52:831–41.
15. **Rudall PJ, Remizowa MV, Prenner G, Prychid CJ, Tuckett RE, Sokoloff DD.** Nonflowers near the base of extant Angiosperms? Spatiotemporal arrangement of organs in reproductive units of *Hydatellaceae* and its bearing on the origin of the flower. *American Journal of Botany.* 2009;96:67–82.
16. **Garay-Arroyo A, Piñeyro-Nelson A, García-Ponce B, Sánchez M de LP, Alvarez-Buylla ER.** When ABC becomes ACB. *Journal of Experimental Botany.* 2012;63:2377–95.
17. **Lynch M, Conery JS.** The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290:1151–5.
18. **Taiz L, Zeiger E.** *Plant Physiology.* Sinauer Associates; 2010.
19. **Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H.** Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science.* 1990;250:931–6.
20. **Passmore S, Maine GT, Elble R, Christ C, Tye BK.** *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT α cells. *Journal of Molecular Biology.* 1988;204:593–606.
21. **Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM.** The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature.* 1990;346:35–9.
22. **Sommer H, Beltrán JP, Huijser P, Pape H, Lönnig WE, Saedler H, et al.** *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *The EMBO Journal.* 1990;9:605–13.
23. **Norman C, Runswick M, Pollock R, Treisman R.** Isolation and properties of cDNA clones encoding *SRF*, a transcription factor that binds to the c-fos serum response element. *Cell.* 1988;55:989–1003.
24. **Alvarez-Buylla ER, Pelaz S, Liljegen SJ, Gold SE, Burgeff C, Ditta GS, et al.** An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA.* 2000;97:5328–33.
25. **Theissen G, Kim JT, Saedler H.** Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution.* 1996;43:484–516.

26. **De Bodt S, Raes J, Van de Peer Y, Theissen G.** And then there were many: MADS goes genomic. *Trends in Plant Science*. 2003;8:475–83.
27. **Portereiko MF, Lloyd A, Steffen JG, Punwani JA, Otsuga D, Drews GN.** *AGL80* is required for central cell and endosperm development in *Arabidopsis*. *Plant Cell*. 2006;18:1862–72.
28. **Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM.** The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*. 2011;23:865–72.
29. **Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, Salazar A, et al.** Transcriptional profiles underlying parent-of-origin effects in seeds of *Arabidopsis thaliana*. *BMC Plant Biol*. 2010;10:72–22.
30. **Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, et al.** Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. *PLoS Genet*. 2011;7:e1001303.
31. **Yu YT, Breitbart RE, Smoot LB, Lee Y, Mahdavi V, Nadal-Ginard B.** Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. *Genes & Development*. 1992;6:1783–98.
32. **Becker A, Theissen G.** The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution*. 2003;29:464–89.
33. **Huang K, Louis JM, Donaldson L, Lim FL, Sharrocks AD, Clore GM.** Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *The EMBO Journal*. 2000;19:2615–28.
34. **Hayes TE, Sengupta P, Cochran BH.** The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes & Development*. 1988;2:1713–22.
35. **Riechmann JL, Wang M, Meyerowitz EM.** DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Research*. 1996;24:3134–41.
36. **Riechmann JL, Krizek BA, Meyerowitz EM.** Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc Natl Acad Sci USA*. 1996;93:4793–8.
37. **Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H, Sommer H.** Multiple interactions amongst floral homeotic MADS box proteins. *The EMBO Journal*. 1996;15:4330–43.
38. **Fan HY, Hu Y, Tudor M, Ma H.** Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant Journal*. 1997;12:999–1010.
39. **Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H.** DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. *Plant Cell*. 1996;8:81–94.
40. **Cho S, Jang S, Chae S, Chung KM, Moon YH, An G, et al.** Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. *Plant Mol. Biol*. 1999;40:419–29.
41. **Honma T, Goto K.** Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*. 2001;409:525–9.
42. **de Folter S, Immink RGH, Kieffer M, Parenicová L, Henz SR, Weigel D, et al.** Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell*. 2005;17:1424–33.
43. **Egea-Cortines M, Saedler H, Sommer H.** Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *The EMBO Journal*. 1999;18:5370–9.
44. **Kaufmann K, Melzer R, Theissen G.** MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*. 2005;347:183–98.
45. **Theissen G.** Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology*. 2001;4:75–85.
46. **Melzer R, Theissen G.** Reconstitution of ‘floral quartets’ *in vitro* involving class B and class E floral homeotic proteins. *Nucleic Acids Research*. 2009;37:2723–36.
47. **Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, et al.** Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci*. 2012;109:1560–5.
48. **Winter K-U, Saedler H, Theissen G.** On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *Plant Journal*. 2002;31:457–75.
49. **Kim S, Yoo M-J, Albert VA, Farris JS, Soltis PS, Soltis DE.** Phylogeny and diversification of B-function MADS-box genes in Angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany*. 2004;91:2102–18.

50. **Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M.** Phylogeny and divergence of basal Angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *J Plant Res.* 2004;117:229–44.
51. **Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, et al.** *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *The EMBO Journal.* 1992;11:4693–704.
52. **Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, et al.** Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal.* 1992;11:251–63.
53. **Jack T, Brockman LL, Meyerowitz EM.** The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell.* 1992;68:683–97.
54. **Goto K, Meyerowitz EM.** Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Development.* 1994;8:1548–60.
55. **Mondragón-Palomino M, Theissen G.** MADS about the evolution of orchid flowers. *Trends in Plant Science.* 2008;13:51–9.
56. **Chung Y-Y, Kim S-R, Kang H-G, Noh Y-S, Park MC, Finkel D, et al.** Characterization of two rice MADS box genes homologous to *GLOBOSA*. *Plant Science.* 1995;109:45–56.
57. **Kyozuka J, Kobayashi T, Morita M, Shimamoto K.** Spatially and temporally regulated expression of rice MADS box genes with similarity to *Arabidopsis* class A, B and C genes. *Plant and Cell Physiology.* 2000;41:710–8.
58. **Prasad K, Vijayraghavan U.** Double-stranded RNA interference of a rice *PI/GLO* paralog, *OsMADS2*, uncovers its second-whorl-specific function in floral organ patterning. *Genetics.* 2003;165:2301–5.
59. **Yao S-G, Ohmori S, Kimizu M, Yoshida H.** Unequal genetic redundancy of rice *PISTILLATA* orthologs, *OsMADS2* and *OsMADS4*, in lodicule and stamen development. *Plant and Cell Physiology.* 2008;49:853–7.
60. **Moon YH, Jung JY, Kang HG, An G.** Identification of a rice *APETALA3* homologue by yeast two-hybrid screening. *Plant Mol. Biol.* 1999;40:167–77.
61. **Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, et al.** *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development.* 2003;130:705–18.
62. **Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ.** Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell.* 2000;5:569–79.
63. **Münster T, Wingen LU, Faigl W, Werth S, Saedler H, Theissen G.** Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene.* 2001;262:1–13.
64. **McGonigle B, Bouhidel K, Irish VF.** Nuclear localization of the *Arabidopsis* *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes & Development.* 1996;10:1812–21.
65. **Lenser T, Theissen G, Dittrich P.** Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Comput. Biol.* 2009;5:e1000264.
66. **Winter K-U, Weiser C, Kaufmann K, Bohne A, Kirchner C, Kanno A, et al.** Evolution of class B floral homeotic proteins: obligate hetero-dimerization originated from homo-dimerization. *Mol Biol Evol.* 2002;19:587–96.
67. **Sundström J, Engström P.** Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant Journal.* 2002;31:161–9.
68. **Smaczniak C, Immink RGH, Angenent GC, Kaufmann K.** Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development.* 2012;139:3081–98.
69. **Bartlett M, Thompson B, Brabazon H, Del Gizzi R, Zhang T, Whipple C.** Evolutionary dynamics of floral homeotic transcription factor protein-protein interactions. *Mol Biol Evol.* 2016;33:1486–501.
70. **Tzeng TY, Yang CH.** A MADS box gene from lily (*Lilium longiflorum*) is sufficient to generate dominant negative mutation by interacting with *PISTILLATA (PI)* in *Arabidopsis thaliana*. *Plant and Cell Physiology.* 2001;42:1156–68.
71. **Feldman M.** The origin of cultivated wheat. *The World Wheat Book.* Paris: Lavoisier Publishing; 2001.
72. **Rapp RA, Udall JA, Wendel JF.** Genomic expression dominance in allopolyploids. *BMC Biology.* 2009;7:18.
73. **Flagel LE, Wendel JF.** Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 2010;186:184–93.
74. **Tanaka M, Tanaka H, Shitsukawa N, Kitagawa S, Takumi S, Murai K.** Homoeologous copy-specific expression patterns of MADS-box genes for floral formation in allopolyploid wheat. *Genes Genet. Syst.* 2016;90:217–29.

75. **Akar T, Avci M, Dusunceli F.** Barley: post-harvest operations. The Central Research Institute for Field Crops, Turkey (FAO). 2004.
76. **Schmitz J, Franzen R, Ngyuen TH, Garcia-Maroto F, Pozzi C, Salamini F, et al.** Cloning, mapping and expression analysis of barley MADS-box genes. *Plant Mol. Biol.* 2000;42:899–913.
77. **Callens C, Tucker MR, Zhang D, Wilson ZA.** Dissecting the role of MADS-box genes in monocot floral development and diversity. *Journal of Experimental Botany.* 2018;69:2435–59.
78. **Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, et al.** Genetic dissection of barley morphology and development. *Plant Physiology.* 2011;155:617–27.
79. **Caldwell DG, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R.** A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). *Plant Journal.* 2004;40:143–50.
80. **Talamè V, Bovina R, Sanguineti MC, Tuberosa R, Lundqvist U, Salvi S.** TILLMore, a resource for the discovery of chemically induced mutants in barley. *Plant Biotechnology Journal.* 2008;6:477–85.
81. **Szurman-Zubrzycka ME, Zbieszczek J, Marzec M, Jelonek J, Chmielewska B, Kurowska MM, et al.** HorTILLUS - A rich and renewable source of induced mutations for forward/reverse genetics and pre-breeding programs in barley (*Hordeum vulgare* L.). *Front. Plant Sci.* 2018;9:216.
82. **Schreiber M, Barakate A, Uzrek N, Macaulay M, Sourdille A, Morris J, et al.** A highly mutagenised barley (cv. Golden Promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods.* 2019;1–14.
83. **Peng Z-S.** A new mutation in wheat producing three pistils in a floret. *Journal of Agronomy and Crop Science.* 2003;189:270–2.
84. **Peng Z-S, Yang J, Wei S-H, Zeng J-H.** Characterization of the common wheat (*Triticum aestivum* L.) mutation line producing three pistils in a floret. *Hereditas.* 2004;141:15–8.
85. **Peng Z-S, Martinek P, Kosuge K, Kuboyama T, Watanabe N.** Genetic mapping of a mutant gene producing three pistils per floret in common wheat. *J. Appl. Genet.* 2008;49:135–9.
86. **Murai K, Takumi S, Koga H, Ogihara Y.** Pistillody, homeotic transformation of stamens into pistil-like structures, caused by nuclear-cytoplasm interaction in wheat. *Plant Journal.* 2002;29:169–81.
87. **Hama E, Takumi S, Ogihara Y, Murai K.** Pistillody is caused by alterations to the class-B MADS-box gene expression pattern in alloplasmic wheats. *Planta.* 2004;218:712–20.
88. **Gregory FJ, Purvis ON.** Abnormal flower development in barley involving sex reversal. *Nature.* 1947;160:221–2.
89. **Moh CC, Nilan RA.** Multi-ovary in barley: a mutant induced by atomic bomb irradiation. *J Hered.* 1953;44:183–4.
90. **Kamra OP, Nilan RA.** Multi-ovary in barley floral anatomy and embryo-sac development. *J Hered.* 1959;50:159–66.

Proposed study



Modification of floral architecture can lead to the development of fertility control systems, which are an important component of hybrid breeding. One way to modify the floral architecture could be to exploit the advanced genetic knowledge of flower development from diverse species. This could be utilised, for example, to modulate the function of MADS-box transcription factors, which work together to generate the vast floral diversity we observe around us. Alternatively, mutant resources can be screened to identify species-specific components that influence fertility. In this context, pistillody mutants in barley represent a promising starting point as they are inherently male-sterile and have the potential to produce multiple seeds.

Given the above premises, this work focuses on the characterization of three barley pistillody mutants: *multiovary1 (mov1)*, *mov2* and *mov5*. Specifically, the present work aims to:

- Identify the genes responsible for the multiovary phenotype in 3 barley *mov* mutants
- Understand the interactions between the candidate genes of the 3 *mov* mutants
- Establish how the candidate genes fit in the known floral developmental network

The work also discusses the implications of transferring the knowledge created on *mov* mutants in wheat for use in a hybrid breeding scenario.

Chapter 3



Statement of Authorship

Title of Paper	The B-class MADS box gene <i>HvMADS16</i> underlies barley multiovary mutant <i>mov1</i>
Publication Status	Unpublished and unsubmitted work written in manuscript style.

By signing the Statement of Authorship, each author certifies that:

- i. each author's contribution to the manuscript is accurate; and
- ii. permission is granted to include the manuscript in the thesis

Principal Author

Name of Principal Author	Caterina Selva		
Contribution to the Paper	Designed and performed the experiments. Analysed and interpreted the results. Wrote the manuscript.		
Percentage of Contribution	70 %		
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	17/09/2019

Co-author Contributions

Name of Co-author	Dr. Xiujuan Yang		
Contribution to the Paper	Optimised the automated <i>in situ</i> hybridization protocol. Provided protocols and plasmids for cloning of CRISPR constructs and BiFC. I hereby certify that the Statement of		
Signature		Date	3/9/2019

Name of Co-author	Dr. Neil Shirley		
	I hereby certify that the Statement of Authorship is accurate.		

Contribution to the Paper	Supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of authorship is accurate.		
Signature		Date	7/09/2019

Name of Co-author	Dr. Ute Baumann		
	I hereby certify that the Statement of authorship is accurate.		

e of Co-author	rof. Dr. Matthew R. Tucker	
	ed the manuscript. I hereby certify that the Stateme horship is accurate.	

The B-class MADS-box gene *HvMADS16* underlies the barley multiovary locus *mov1*

C. Selva, X. Yang, N. Shirley, R. Whitford, U. Baumann, M. R. Tucker

School of Agriculture Food and Wine, University of Adelaide, Waite Campus, Urrbrae 5064, South Australia, Australia

Abstract

Floral architecture is a key determinant of a plant's reproductive strategy and is therefore one of the factors able to influence seed set. Implementing hybrid breeding in cereals such as wheat and barley requires forcing outcrossing in a plant that is otherwise autogamous. Modifying the structure of florets is a promising target to support hybrid breeding efforts, particularly in wheat, and consequently lower the cost of hybrid seed production. Despite the complexities of wheat genetics, knowledge of pathways controlling development can often be extrapolated from barley to wheat due to their close evolutionary history. Here, we analyse the barley multiovary mutant *mov1* and identify the B-class MADS-box gene *HvMADS16* as a putative candidate. The homeotic recessive mutant *mov1* transforms lodicules into bract-like organs and stamens into carpels, respectively. Developmental, histological and transcriptomic data provide clues on how *mov1* interacts with the ABC model of flower development and the processes that lead to stamen and lodicule specification in barley. Understanding flower development in related cereals with a relatively simpler genome can provide key insights for flower development in wheat.

Introduction

The most diverse and captivating floral structures in nature build upon a simple set of rules established in the ABC model of flower development. This model was predicted based on extensive studies in model species like *Arabidopsis thaliana* and *Antirrhinum majus*, and postulates that each organ within a flower is specified by the combinatorial action of specific classes of genes, referred to as the A-, B-, C-, D- and E- class genes [1-8]. These genes act in two adjacent whorls in the flower. Thus, in the outermost whorl, A- and E-class genes together specify sepals or other bract-like organs, while the action of A-, B- and E-class genes in whorl 2 give rise to petals or equivalent structures. Stamens in the third whorl are specified thanks to the action of B-, C- and E-class genes. Carpels, on the other hand, require the action of C- and E-class genes, with D-class genes being essential for ovule development. Surprisingly, the majority of genes forming the ABC model are transcription factors belonging to the MADS-box gene family.

In cereals, especially for wheat, manipulating the flower structure to increase yield represents a key target for breeding. In wheat the flower is normally formed by a pair of palea and lemma in whorl 1, two lodicules in whorl 2, three stamens in whorl 3 and a single carpel in the fourth whorl. However, numerous natural and induced mutants showing pistillody (the conversion of stamens into additional pistils) have been described [9-11]. Studies in wheat pistillody mutants often show that the pistillody phenotype correlates with changes in expression of genes belonging to the ABC model [12-14]. This is particularly true for B-, C- and D-classes which are directly implicated in stamen, carpel and ovule development.

In barley, at least three multiovary loci have been reported [15]. Mutations at these loci specifically affect flower development and cause additional carpels to grow at the total or partial expense of stamens. The *multiovary1* (*mov1*) locus was initially recorded by Tazhin in 1980 [16]. The first described allele, termed *mo5*, appeared spontaneously in the barley variety Revelatum 1886. *mo5* presented as a homeotic conversion of stamens into additional non-functional carpels, with substitution of lodicules as half-curved leaf-like structures and partial fertility [16]. Linkage with the *nud* locus positioned *mo5* on the centromeric region of chromosome 7H [16] with later mapping attempts by Soule *et al.* (1996) suggesting that the locus may encode a member of the MADS-box gene family [17]. In 1995, additional multiovary mutants appeared as a result of a fast neutron mutagenesis experiment in the cultivar Steptoe. Of the mutants identified, *mo6b* showed striking phenotypic similarity to *mo5*, despite showing complete sterility. Using an RFLP probe with sequence similarity to MADS-box genes, *mo6b* mapped to the centromeric region of chromosome 7H, at a similar location to *mo5* [18]. Because both *mo6b* and *mo5* were shown to be monofactorial recessive they were deemed to belong to the same locus, although allelism tests were not conducted due to seed unavailability [15]. Despite the availability of an early general description for the mutant phenotype and rough mapping information for the *mov1* locus, the responsible gene(s) have yet to be identified. Here we show that a total of three genes on chromosome 7HL, including the B-class gene *HvMADS16* (*HORVU7Hr1G091210*) are absent in *mov1* plants. We propose that absence of *HvMADS16* is responsible for the *mov1* phenotype. Morphological and developmental characterisation data indicate that *HvMADS16* is required for lodicule and stamen specification, as developmental defects in these organs appear very early in floral organogenesis in mutant plants. We explore the interactions of B-class genes in barley and discuss the implications of the absence of *HvMADS16* in the context of the ABC

model. Based on molecular and histological data, we postulate a model explaining how B-class genes specify stamen and lodicule development in barley.

Results:

mov1 florets show homeotic conversion of floral organs in whorls 2 and 3

Wild-type barley florets are composed of one set of palea and lemma (bract-like organs) which surrounds two lodicules, three stamens and a single carpel (**Figure 1**). By contrast, within *mov1* florets the three stamens (whorl 3) each appear to have been replaced by carpels and one pair of leaf-like organs is present instead of the lodicules (whorl 2) (**Figure 1**). However, in *mov1* florets there is no conversion of the floral organ in the fourth whorl which still develops into a central carpel like in wild type, as well as for the palea and lemma in whorl 1.

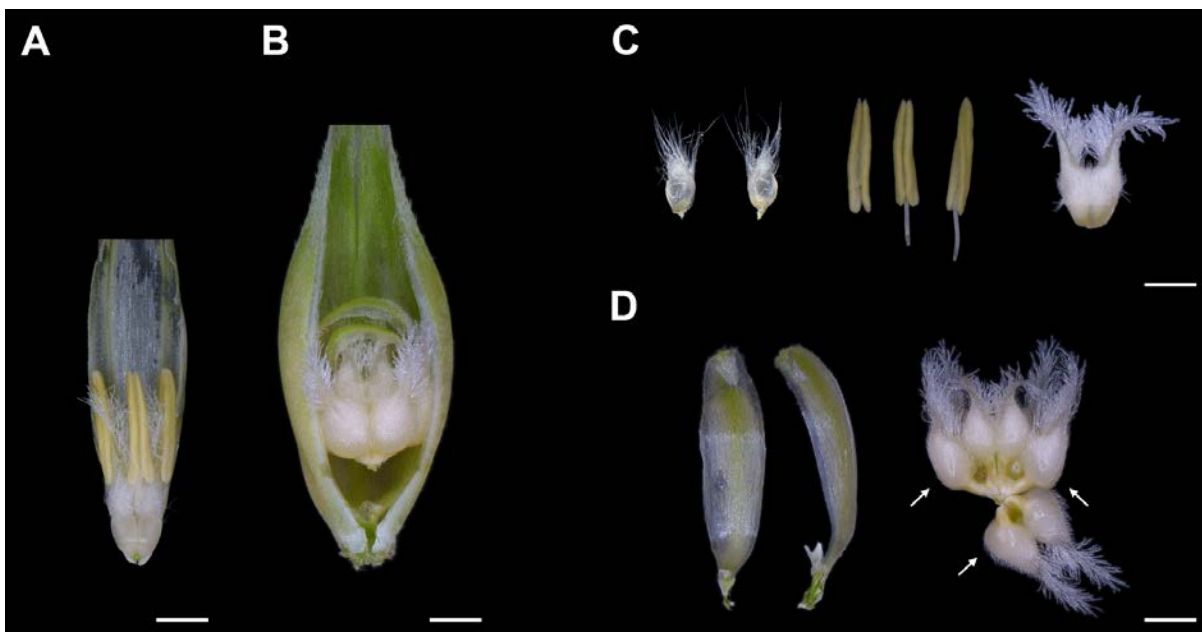


Figure 1. Exposed (A) wild-type and (B) *mov1* florets. (C) The floral organs in a wild-type floret consist of 2 lodicules, 3 stamens and 1 carpel. (D) In *mov1* florets the lodicules are converted into bract-like organs, and stamens are converted into additional carpels (arrows). Scale bars: 1000 μm .

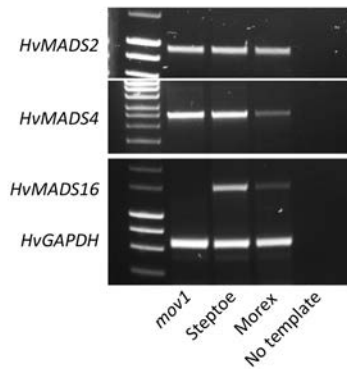
The B-class gene *HvMADS16* is absent in *mov1*

Since *mov1* shows specific homeotic conversion of floral organs in whorls 2 and 3, a PCR-based strategy, consisting of PCR amplification followed by Sanger sequencing of the amplicon, was used to survey all B-class genes (derived from Morex reference assembly Hv_IBSC_PGSB_v2) for presence/absence, structural and sequence variants in *mov1* plants relative to wild type (cv. Steptoe). Homology-based search using rice B-class genes found a total of 3 B-class genes in barley consisting of 2 GLO-like homologues and a single DEF-like homologue, confirming already published data [19]. The GLO-like homologues *HvMADS2* (*HORVU3Hr1G091000*) and *HvMADS4* (*HORVU1Hr1G063620*) did not show differences when tested for amplicon size polymorphisms by PCR as well as in Sanger sequencing between genotypes (**Figure 2A**). In contrast, the DEF-like homologue *HORVU7Hr1G091210* on chromosome 7H appeared to be absent in *mov1* mutants (**Figure 2A**). *HORVU7Hr1G091210* will be named here as *HvMADS16* based on homology to the rice B-class gene *OsMADS16/OsSPW1* (*SUPERWOMAN1*) and the high-sequence identity shared by the respective encoded proteins (88.3%) (**Appendix A, Supplementary Figure S4**). Genotyping by copy number analysis combined with phenotyping performed on a total of 583 progeny plants originating from a *mov1* heterozygote seed stock showed a 3 : 1 (wild type : multiovary) segregation, typical for a single recessive locus (**Appendix A, Supplementary Table S1**). Absence of *HvMADS16* co-segregated perfectly with the *mov1* phenotype in all plants tested. These findings indicate that *mov1* lacks a *HvMADS16* sequence, as would be expected from a large deletion, typical of fast-neuron-derived mutants.

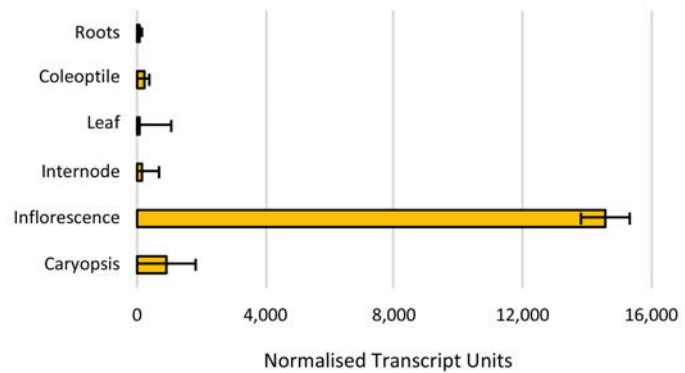
To define the size of the deletion surrounding *HvMADS16*, a similar PCR-based approach was used to test the presence/absence of neighbouring gene sequences, based on annotations from the barley reference Morex genome Hv_IBSC_PGSB_v2. The *mov1* mutant appeared to

be missing a region of approximately 0.95 Mb relative to wild type (**Figure 2C**). According to the reference sequence, this region is predicted to encode three gene sequences: *HORVU7Hr1G091190* (40S ribosomal protein), *HORVU7Hr1G091200* (undescribed protein); and *HORVU7Hr1G091210/HvMADS16* (MADS-box transcription factor 16) (**Appendix A, Supplementary Table S2**). Consistent with a role of *HvMADS16* in correct inflorescence specification, transcripts were predominantly detected in developing wild-type inflorescences when tested by quantitative Real-Time PCR (qRT-PCR) across a Steptoe tissue series (**Figure 2B**). On the other hand, transcripts for *HORVU7Hr1G091200* could not be detected in any of the tissues examined. Furthermore, a BLASTx query of the translated nucleotide sequence against NCBI non-redundant protein databases found no significant similarity to any protein in other species. The gene encoding a 40S ribosomal protein (*HORVU7Hr1G091190*) could not be assayed by PCR or qRT-PCR due to the highly repetitive nature of the sequence. As presence/absence of this gene could not be confirmed, it was included in the deletion following a more conservative approach. However, publicly available RNAseq data [20] indicate that *HORVU7Hr1G091190* is not expressed in 16 barley tissues including the inflorescence. Considering the specific homeotic conversion of floral organs in whorls 2 and 3, the absence of *HvMADS16* in *mov1* mutant plants and the role of B-class genes in other plant species, *HvMADS16* appears to be the most likely causal agent for *mov1*. Based on the reference Morex assembly Hv_IBSC_PGSA_v2 the *HvMADS16* gene is divided in seven exons and contains untranslated regions (UTRs) at both the 5' and 3' ends (**Figure 2D**).

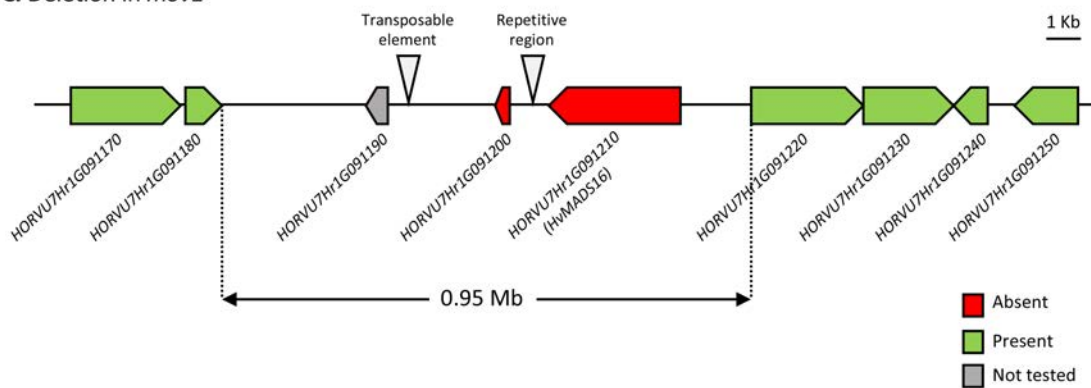
A. B-class genes in *mov1*



B. *HvMADS16* transcript abundance



C. Deletion in *mov1*



D. *HvMADS16* gene structure

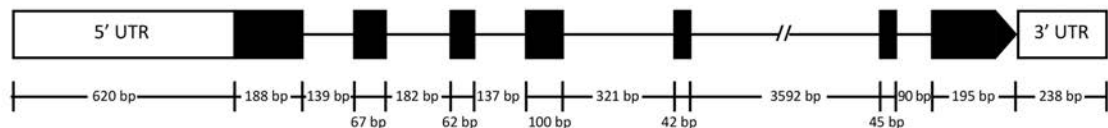


Figure 2. (A) *HvMADS16* (*HORVU7Hr1G091210*) is the only B-class gene physically absent in *mov1* when assayed by PCR. The barley glyceraldehyde 3-phosphate dehydrogenase *HvGAPDH* gene (*HORVU7Hr1G074690*) was used as reference. (B) Transcript abundance of *HvMADS16* (*HORVU7Hr1G091210*) in a Steptoe tissue series as assayed by qRT-PCR. For details about tissue sampling refer to Materials and Methods. (C) Present (green) and absent (red) genes surrounding *HvMADS16* (*HORVU7Hr1G091210*) in *mov1* as assayed by PCR. Deletion size in *mov1* is estimated to be no bigger than 0.95 Mb, based on the Morex reference assembly *Hv_IBSC_PGsb_v2*. (D) Gene structure of *HvMADS16* (*HORVU7Hr1G091210*); length in base pairs of Untranslated Regions (UTR), exons (black) and introns (solid line) is indicated.

In *mov1* the stamens are not correctly specified during floral organogenesis

Scanning electron microscopy (SEM) images of wild-type and *mov1* developing inflorescences were compared in order to determine how and when *mov1* affects floral organ development. The earliest observable difference between wild type and mutant was seen immediately preceding the appearance of stamen primordia at Waddington stage W3.0 (**Figure 3**) [21]. Following Waddington stage W3.0, wild-type meristems develop three lateral dome-shaped protrusions that are clearly observable at Waddington stage W3.5 (**Figure 3**). These protrusions are stamen primordia which subsequently differentiate into filament and anther tissues (W5.0 – 7.0). The meristematic tissue at the centre of the meristem terminally differentiates into a single carpel (W5.0 – 7.0) (**Appendix A, Supplementary Figure S1**).

When compared to wild type, a crease appears in the basal floral meristems of *mov1* inflorescences already at W3.0 (**Figure 3**). At stage W3.5 the meristems divide into protrusions that arrange into multiple concentric creases by W5.0 (**Figure 3**). As development progresses (W5.0 – 7.0) each crease will then give rise to a carpel, leading to the 4-carpel structure visible in the mature *mov1* floret (**Figure 3; Appendix A, Supplementary Figure S1**). Occasionally, it was observed that a single floral meristem in *mov1* inflorescences could give rise to two distinct florets (**Figure 3**).

To further examine if the additional carpels in *mov1* are functional and correctly developed, transverse sections of the carpel structures were observed by light microscopy. Interestingly in *mov1*, each carpel lobe contains an ovule-like structure (**Figure 4**). Nonetheless, female gametophytes are not correctly differentiated in any of the additional ovules, as well as in the ovule of the central carpel (**Figure 4**).

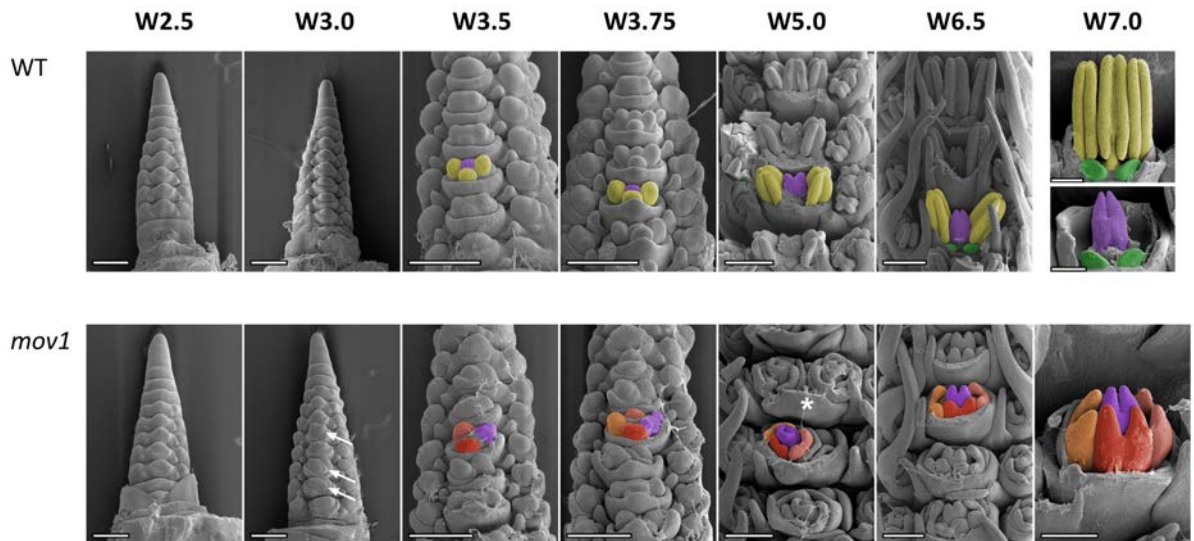


Figure 3. SEM of barley inflorescence development in wild type (WT) and *mov1*. For wild type, primordia giving rise to stamens are false-coloured in yellow, cells giving rise to the carpel in purple and lodicules in green. In *mov1*, the central carpel (purple) is retained while the stamens are converted into additional carpels. Cells giving rise to the additional carpels are false-coloured in different shades of red/orange. White arrows in *mov1* (W3.0) indicate creases in the floral meristems, while white asterisk at W5.0 indicates separation of a single floral meristem into two distinct florets. Waddington stage is indicated for each developmental timepoint, from W5.0 lemma and/or stamens have been removed to expose the carpels. Scale bars: 200 μ m.

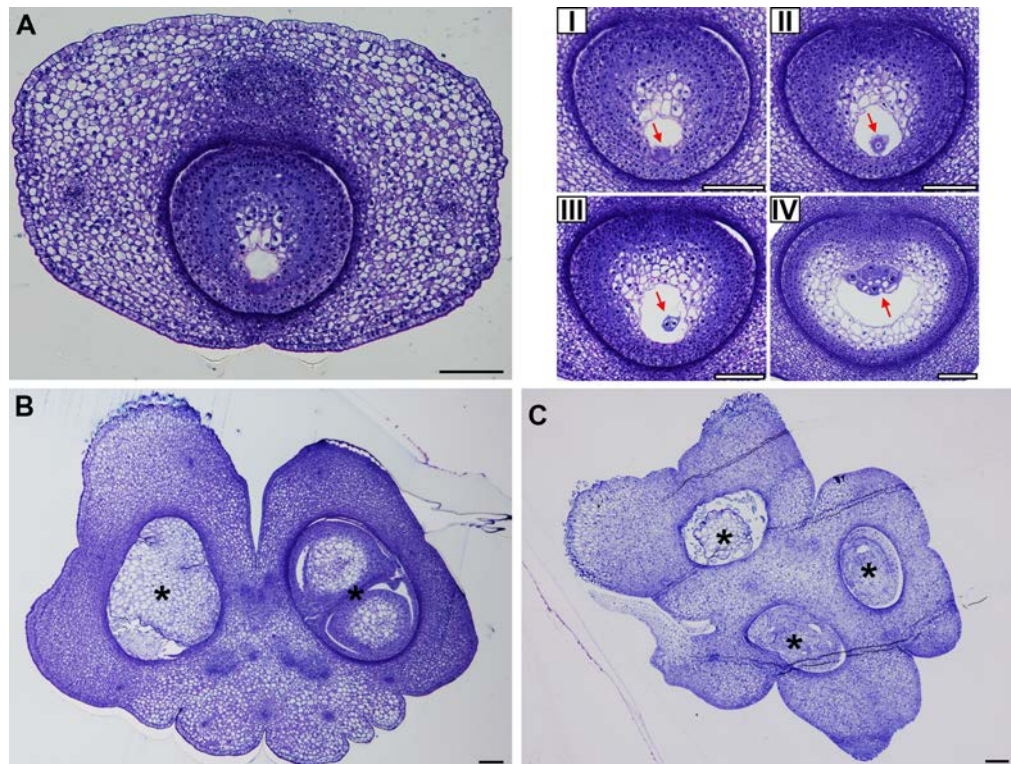


Figure 4. Transverse sections of mature (A) wild-type and (B and C) *mov1* carpels stained with toluidine blue. Black asterisks indicate the ovule-like structures formed in the multiovary mutant. Inset shows characteristics of a wild-type female gametophyte. Red arrows indicate: (I) synergid cells, (II) egg cell, (III) central cell and (IV) antipodal cells. For wild type scale bars: 50 µm; for *mov1* scale bars: 100 µm.

In *mov1* the expression of the ABC genes is disrupted

To evaluate how the intricate interplay of genes within the ABC model are influenced in *mov1*, qRT-PCR was performed on developing inflorescences (W2.0 - 6.0). These particular stages were chosen because, developmentally, they cover the specification and start of differentiation of both the male and female barley reproductive organs. A general overview of the dynamics of the ABC model was obtained by examining transcript abundance for representatives for each gene class. As expected, *HvMADS16* expression in wild type steadily increases as inflorescence development progresses, while it is completely absent in *mov1* samples (**Figure 5**). Overall, as shown in **Figure 5**, expression of B-class genes (*HvMADS2* and *HvMADS4*) and E-class genes (*HvMADS7* and *HvMADS8*) is significantly lower in *mov1* from stage W3.5 onwards (**Appendix A, Supplementary Figure S2**). Conversely, expression of C-class genes (*HvMADS3* and *HvMADS58*), D-class gene (*HvMADS13*) and the barley orthologue of the rice carpel-specific *DROOPING LEAF* (*HvDL*) gene are all significantly increased in *mov1* compared to wild type (**Figure 5; Appendix A, Supplementary Figure S2**). Perhaps not surprisingly given their role in establishing floral meristem identity, the expression pattern of A-class genes (*HvMADS14* and *HvMADS15*) remains largely unaffected, with a slight decrease in expression for *HvMADS14* (**Appendix A, Supplementary Figure S2**).

The transcript abundance of barley orthologues of genes which were shown in rice to be specifically involved in both male and female germ-line development was also examined. Expression of *HvMEL1* (*MEIOSIS ARRESTED AT LEPTOTENE1*), involved in rice in the correct

development pre-meiotic germ cells and progression of meiosis [22], is significantly reduced in *mov1* plants (**Appendix A, Supplementary Figure S2**). The *HvMSP1* (*MULTIPLE SPOROCTE*) gene shows homology to rice *OsMSP1*, which has been shown to restrict the number of cells beginning sporogenesis as well as initiating anther wall formation in rice [23]. However, no significant expression difference was detected between wild-type and *mov1* inflorescences (**Appendix A, Supplementary Figure S2**).

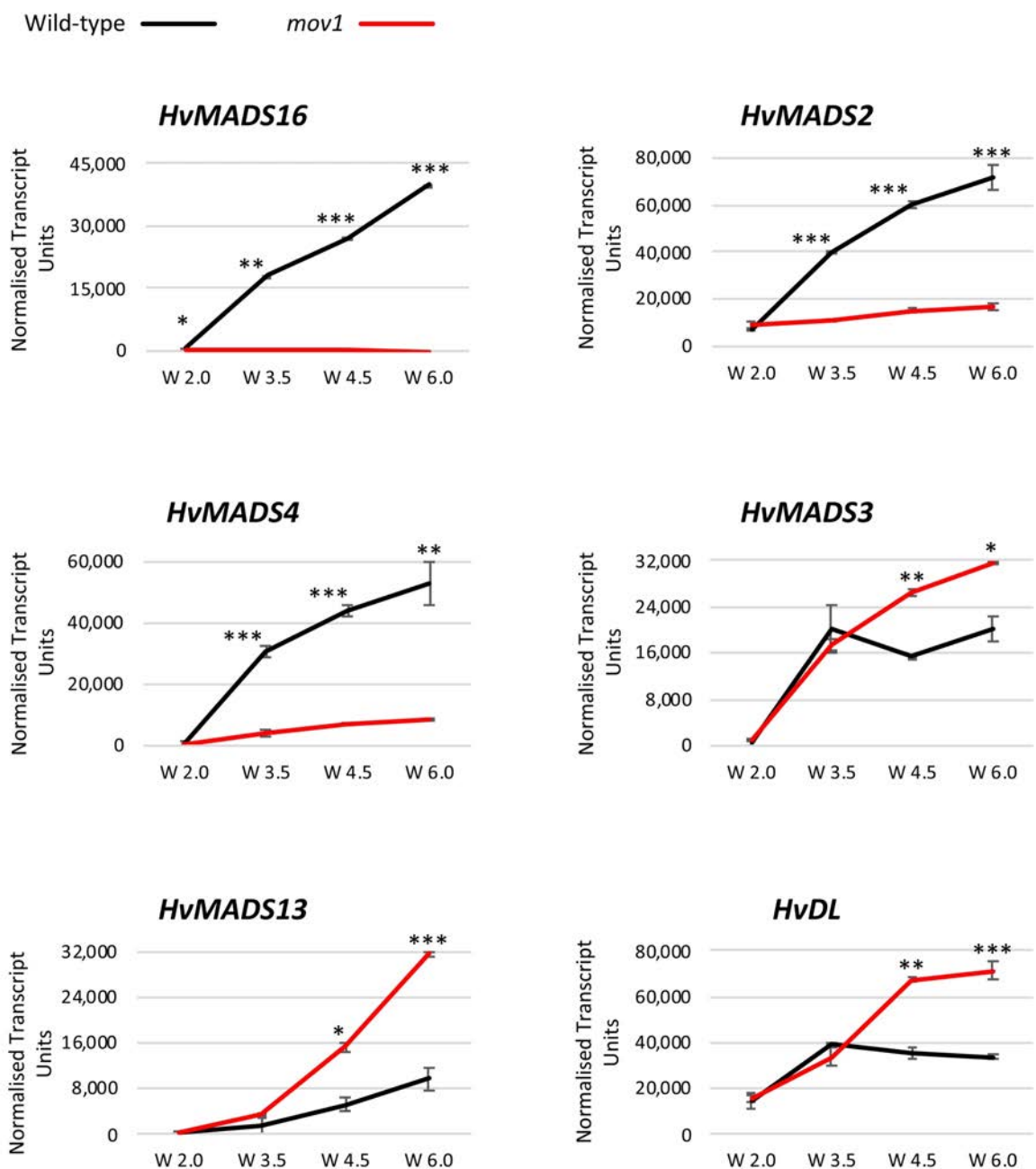


Figure 5. Transcript abundance assessed by qRT-PCR in wild-type (black) and *mov1* (red) developing inflorescences at stages W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development) for: B-class genes *HvMADS16* (*HORVU7Hr1G091210*), *HvMADS2* (*HORVU3Hr1G091000*), *HvMADS4* (*HORVU1Hr1G063620*); C-class gene *HvMADS3* (*HORVU3Hr1G026650*); D-class gene *HvMADS13* (*HORVU1Hr1G023620*) and *HvDL* (*HORVU4Hr1G067780*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**) and ≤ 0.001 (***) are shown for differences between wild type and *mov1*. For each sample *n* = 3 independent biological replicates.

To validate the qRT-PCR results and determine the spatio-temporal pattern of ABC genes during floret development, transcript location and abundance of selected genes was assessed by *in situ* hybridization in tissue series of Steptoe and the *mov1* mutant. In wild-type Steptoe, *HvMADS16* is first detected in the stamen and lodicule primordia (**Figure 6**). *HvMADS16* signal becomes weaker in whorl 3 as stamens develop, while remaining strong in the lodicules. No expression is observed in the carpel. As expected, *HvMADS16* expression is undetectable in *mov1* inflorescences in any floral organ. *HvMADS3* (C-class) and *HvMADS13* (D-class) were also selected for *in situ* hybridization analysis (**Figure 6**). The expression pattern of these two genes was found to be very similar: in wild-type florets expression is confined to the ovule, while in *mov1*, expression of both *HvMADS3* and *HvMADS13* localizes in 2 or 3 locations per floret. Notably, *HvDL* expression in wild type is restricted to the carpels and to the abaxial side of the lemma but is not expressed in lodicules or stamens (**Figure 6**). In *mov1* florets *HvDL* expression is more diffuse within the floret and remains in the lemma. For all genes tested, sense probes gave no observable signal (**Appendix A, Supplementary Figure S3**).

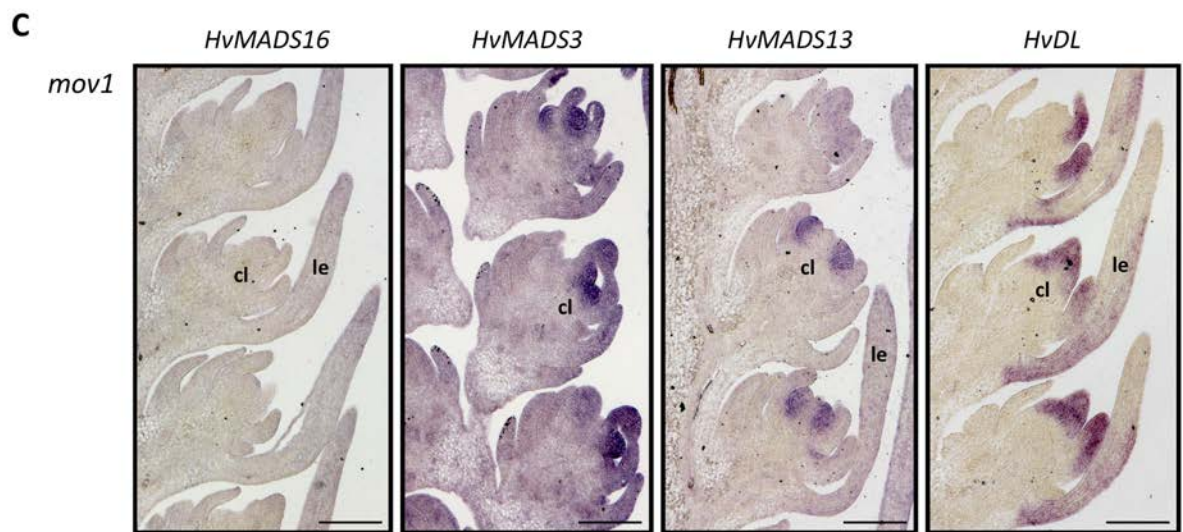
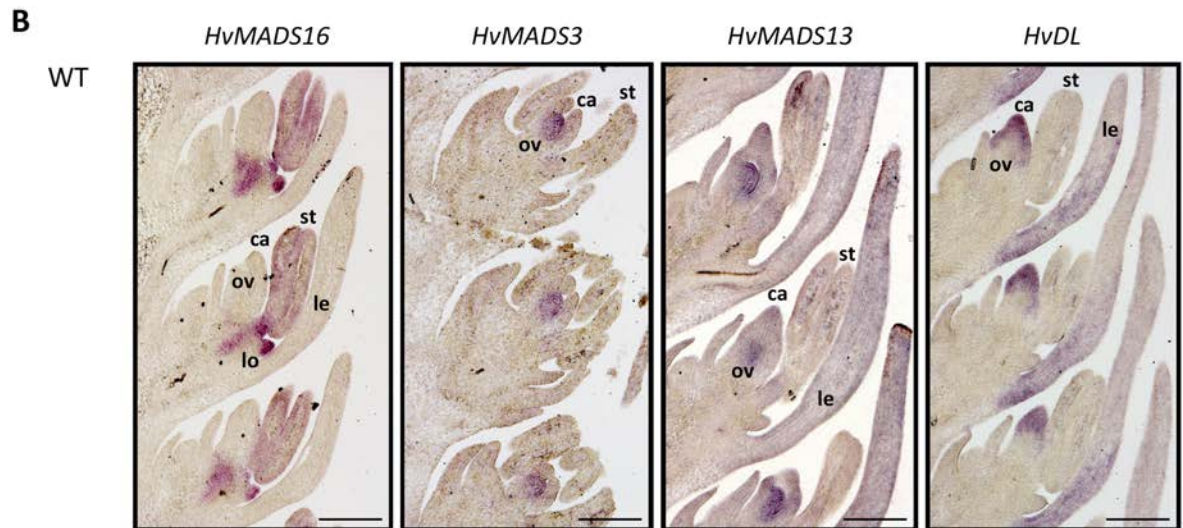
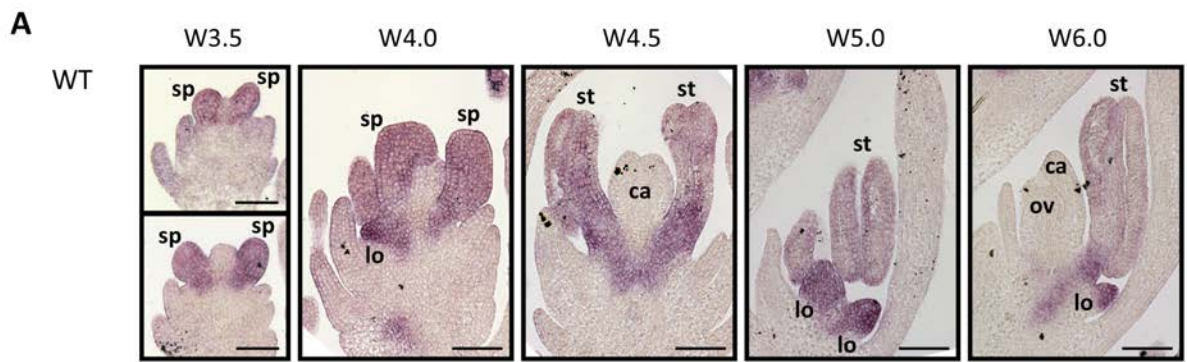


Figure 6. (A) Expression pattern as assessed by *in situ* hybridization with antisense probe for *HvMADS16* (*HORVU7Hr1G091210*) during wild-type (WT) inflorescence development. *HvMADS16* signal is initially localised in stamen primordia (sp). At later stages of development signal is confined to lodicules (lo) and more weakly in stamens (st), but not in carpels (ca) or ovules (ov). Lemma is indicated by le. Scale bars: 100 μ m. Expression pattern for *HvMADS16*, *HvMADS3* (*HORVU3Hr1G026650*), *HvMADS13* (*HORVU1Hr1G023620*) and *HvDL* (*HORVU4Hr1G067780*) in (B) wild-type and (C) *mov1* inflorescences at stage W6.0. The carpel-like complex in *mov1* is indicated with cl. Scale bars: 200 μ m.

HvMADS16 is able to form hetero-dimers with the other B-class genes

To assess the interactions of barley B-class proteins bimolecular fluorescence complementation (BiFC) was conducted in onion epidermal cells. Fluorescence signal was observed only in cells co-expressing *HvMADS16* with *HvMADS2* or *HvMADS4* (**Figure 7**). For these combinations, fluorescence was predominantly confined in the nucleus. No signal was observed when the two GLO-like genes *HvMADS2* and *HvMADS4* were co-expressed in the same cells, or with a truncated version of *HvMADS16* (Δ *HvMADS16-nYFP*), containing only the MADS-domain (**Figure 7**). Likewise, no signal could be detected when homo-dimeric combinations for each B-class gene were tested (**Figure 7**). A previously published interaction between rice the glutamyl-tRNA synthetase OsERS1 and its cofactor OsARC was used as positive control [24].

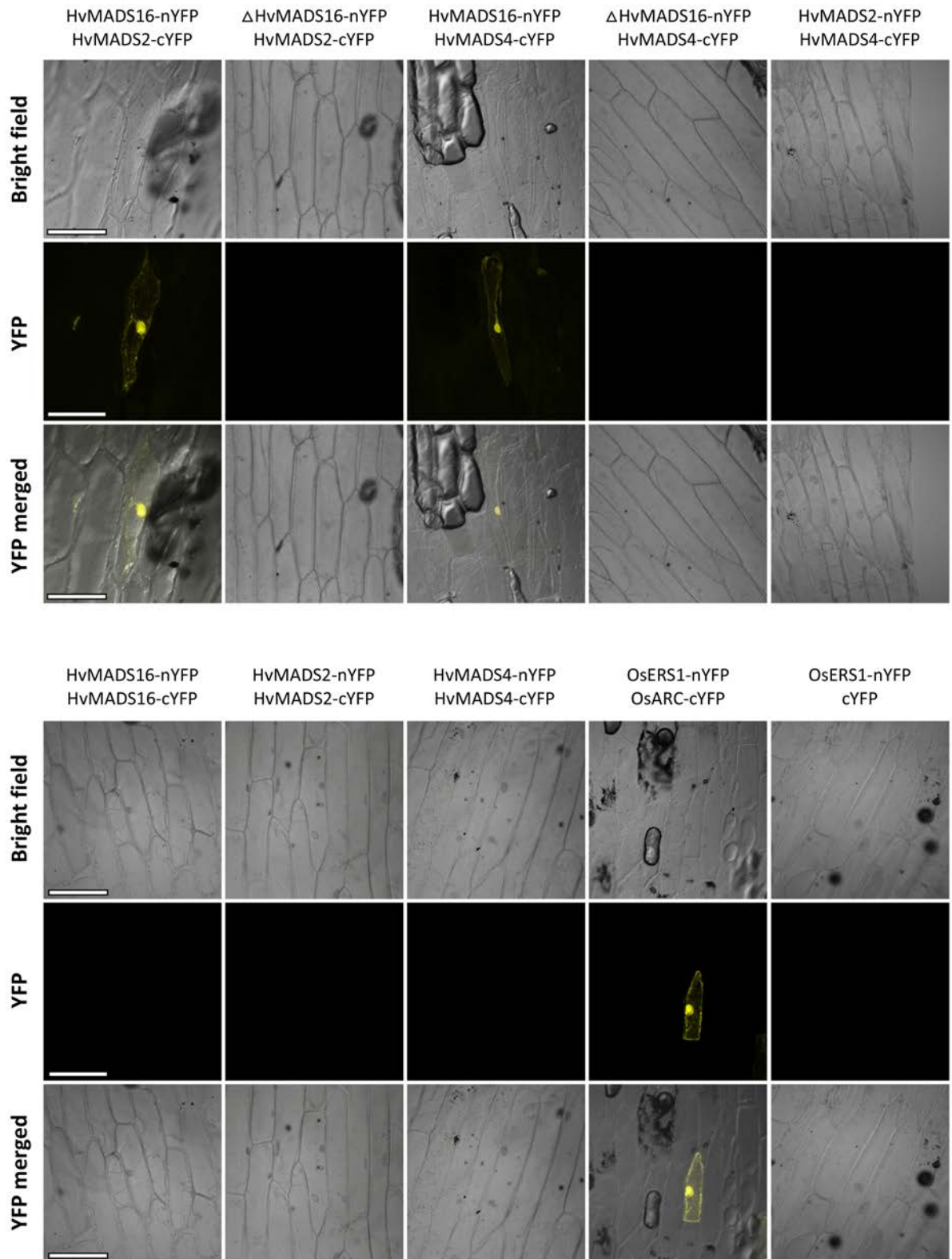


Figure 7. BiFC assays in onion epidermal cells showing interaction between barley B-class genes. nYFP indicates N-terminal of YFP (1-174), while cYFP indicates C-terminal split of YFP (175-241). Scale bars: 200 μ m. A previously published interaction between rice proteins OsERS1 and OsARC was used as positive control [24].

Discussion

mov1 is a B-class mutant

Previous mapping studies positioned the *mov1* locus, responsible for the multiovary phenotype, on the centromeric region of chromosome 7H [18]. This agrees with the present work which identifies a missing region of approximately 0.95 Mb on chromosome 7HL in *mov1* plants. However, the current study positioned the *mov1* locus with higher resolution and significantly narrowed the critical interval to only three putative genes encoding: a 40S ribosomal protein, an undescribed protein and a MADS-box transcription factor. The latter has been named here *HvMADS16* due to high sequence identity (88.3 %) to the rice B-class protein *OsMADS16/SUPERWOMAN1* (**Appendix A, Supplementary Figure S4**). Among the missing genes in *mov1*, *HvMADS16* stands out as the most likely causative candidate as *mov1* plants show specific homeotic conversion of lodicules and stamens in whorls 2 and 3, respectively (**Figure 1**). Based on the ABC model, the striking phenotype strongly suggests the lack or reduced activity of B-class genes. Additionally, the *mov1* phenotype is remarkably similar to the rice B-class mutants *superwoman1* (*spw1*) [25] and maize *silky1* [26], which also show transformation of lodicules and stamens into bract-like organs and carpels and involve DEF-like genes. This observation suggests a conserved role of B-class genes in floral organ specification in monocotyledonous cereals. Furthermore, consistent with what observed by Soule *et al.* (2000) [18], the 3:1 phenotypic segregation observed for *mov1* heterozygote seeds is typical for a single, recessive Mendelian gene which suggests that no other loci are involved in the *mov1* phenotype.

Additional molecular evidence supports our hypothesis that altered *HvMADS16* function underlies the *mov1* phenotype. Besides being physically absent in *mov1* while being present

in wild type, examination of *HvMADS16* transcript levels across a variety of tissues in wild type found expression predominantly in the inflorescence (**Figure 2**). Specifically, *HvMADS16* expression is confined to lodicules and stamens in wild-type developing inflorescences (**Figure 6**). The very specific expression pattern provides a possible explanation as to why *mov1* only shows a major phenotype in floral development, but not in any other aspect of plant growth. Another gene, *HORVU7Hr1G091200*, was identified in the fast-neutron deleted region, however, no expression could be detected for this gene by qRT-PCR. It may be that *HORVU7Hr1G091200* has a very specific expression pattern that was not included by our sampling or it may be that *HORVU7Hr1G091200* represents a pseudogene as it is currently annotated as a low-confidence gene without functional annotation in the Hv_IBSC_PGsb_v2 assembly used and no orthologues could be found in other species.

To validate the role of *HvMADS16* in flower development, a complementation strategy to rescue the *mov1* phenotype was attempted (**Appendix A, Supplementary Figure S5**). As homozygous *mov1* plants are fully sterile, the strategy consisted in genotyping segregating *mov1* plants by copy number analysis and inserting *HvMADS16* in a heterozygote *mov1* background. The transgenic plants resulting from the transformation would genotypically segregate, however the phenotypically rescued plants would have been identified using sequence polymorphisms included in the inserted construct. However, the *mov1* mutant is in the cv. Steptoe, a genotype that is not very prone to transformation, and fully developed regenerant plants could not be obtained from calli following the transformation event.

Another way to corroborate the role of *HvMADS16* is to use a CRISPR strategy to knockout *HvMADS16* in barley plants cv. Golden Promise and thus attempt to re-create the *mov1* phenotype. For this reason, two guide RNAs targeting the first exon of *HvMADS16* at

positions + 29 bp and + 72 bp from the translational start site were designed (**Appendix A, Supplementary Figure S5**). Plant transformation is currently underway, however, given the extent of barley transformation protocol, regenerant plants are still not available for analysis.

Mov1 is essential for early flower development and correct reproductive development

As seen from our SEM analysis, phenotypic differences in *mov1* inflorescences start to appear very early in flower development, corresponding approximately to stage W3.0 (lemma-floret primordium) (**Figure 3**). This is consistent with the spatial expression pattern of *HvMADS16* in wild-type inflorescences as observed by *in situ* hybridization; signal is detectable in young meristems before the differentiation of stamen and lodicule primordia (**Figure 6**). It is also in accordance with *HvMADS16* temporal expression pattern as shown by qRT-PCR, as expression in a wild-type inflorescence development series begins only after stage W2.0 (double ridge) and gradually increases until after the reproductive organs have started differentiating (W6.0) (**Figure 5**).

The end result of the developmental differences occurring in *mov1* are completely sterile florets. Other than their inherent male-sterility due to the lack of stamens, *mov1* flowers also appear to exhibit female-sterility. As shown by histochemical analyses with toluidine blue, all four carpels in *mov1* are non-functional. Although ovule-like structures are present, they are unable to complete development to produce a fully differentiated embryo sac (**Figure 4**). Instead, the ovule-like structures are filled with undifferentiated cell layers, an effect also observed in wheat pistillody lines [12]. We speculate that, as seen for *spw1* in rice, the undifferentiated cells observed in the ovule-like structures derive from overproduction of

the nucellar tissue [25]. However, further experiments are needed to address this point. Furthermore, no seeds were obtained when *mov1* was used as female parent for crosses with the other multiovary mutants *mov2* and *mov5* for allelism tests. Interestingly, this is in contrast to the *Arabidopsis* orthologue *ap3* mutant which produces viable seeds [27]. Taken together, in barley, the presence of multiovary is not indicative of functional reproductive development.

The absence of *HvMADS16* and the remodelling of the ABC model

Many studies have shown the vast and intricate interactions among the genes of the ABC model. Not surprisingly, the absence of a B-class gene in *mov1* causes alterations in the expression of its interacting partners. The changes in expression patterns seen by qRT-PCR and *in situ* hybridization are all consistent with the absence of stamens and lodicules; and the presence of additional carpels in the final *mov1* phenotype.

As expected, *HvMADS16* expression in *mov1* is not detected by qRT-PCR and the signal from *in situ* hybridization is completely lost in the mutant inflorescences. Interestingly, the expression of the other B-class genes (*HvMADS2* and *HvMADS4*) also decreases at a very early stage, from W2.0 (double ridge) onwards, indicating a concerted action in specifying lodicules and stamens (**Figure 5**). This observation is consistent with reports for numerous other species whereby DEF- and GLO-like hetero-dimers are able to activate their own expression and initiate a positive autoregulatory feedback loop [28-30]. Additional evidence supporting this hypothesis comes from the BiFC experiments whereby we show that *HvMADS16* (DEF-like) is an obligate hetero-dimer with either GLO-like proteins *HvMADS2* or *HvMADS4* (**Figure 7**). The MADS-domains alone is insufficient to establish this interaction as

Δ MADS16, a truncated version of HvMADS16 containing only the MADS-domain, did not show a signal in the BiFC assay (**Figure 7**). Furthermore, GLO-like proteins do not interact with each other (**Figure 7**).

Overall, the expression of C-class genes (*HvMADS3* and *HvMADS58*), D-class genes (*HvMADS13*) and *HvDL* increase in *mov1* inflorescences. In wild type, these gene classes are expressed in the innermost fourth whorl of the flower and function in specifying carpel and ovule development. The increase observed in *mov1* inflorescences can be explained by the combined effect of a minor repressive action of B-class genes on carpel and ovule-promoting genes, together with the expansion of the expression domain of these classes to the third floral whorl. The final effect of these expression changes is a floret with carpels in two consecutive whorls (whorl 3 and whorl 4). Interestingly though, there is no alteration in the number of organs in whorl 3, an effect that has also been observed in *spw1* mutants in rice [25].

The spatial expression pattern of *HvMADS3* detected by *in situ* hybridization also highlights a potential sub-functionalization of barley C-class genes. Both the rice (*OsMADS3*) and maize (*ZMM2*) orthologues of *HvMADS3* have been reported to be strongly expressed in stamen primordia [31] and anthers [32], respectively. In contrast, in wild-type barley inflorescences *HvMADS3* expression was specific to the ovule primordia (**Figure 6**). Furthermore, it has been shown that *OsMADS3* plays an important role in specifying stamen identity, as well as in repressing lodicule formation [33]. In our case, *HvMADS3* expression was shown by qRT-PCR to increase in *mov1* inflorescences, which wouldn't be expected if the function of *HvMADS3* were conserved with respect to its rice orthologue (**Figure 5**).

On the other hand, the expression pattern of *HvDL* is consistent with the expression pattern of the rice orthologue *OsDL* [25]. In rice florets, *OsDL* expression has been shown to be

confined to the carpel and to the central vein of the lemma [34]. A very similar expression pattern was observed for *HvDL* in wild type, with expression in the carpel and base of the lemma, thus suggesting a conserved function in carpel morphogenesis. In *mov1* though, *HvDL* expression is more diffuse throughout the developing floret (**Figure 6**).

Likewise, for *HvMADS13*, the spatial expression observed in wild-type inflorescences is consistent with the expression of the rice orthologue *OsMADS13* [35]. For both the barley ovule-specific genes *HvMADS3* and *HvMADS13* the increased expression in *mov1* detected in *mov1* inflorescences by qRT-PCR analysis may be explained by the additional ovules observed by histochemical analysis (**Figure 4; Figure 5**). Notably, this corresponds well with the spatial pattern of expression of both genes in the mutant as visualised by *in situ* hybridisation, which shows distinct expression maxima throughout the floral meristem (**Figure 6**).

During flower development, A-class genes are predominantly involved in determining the floral organs in whorls 1 (palea/lemma) and 2 (lodicules). In rice, there are four A-class genes which slightly differ in function and expression pattern. In particular, *OsMADS14* has been reported to be mainly responsible for the identity of lodicules and stamens, while *OsMADS15* is more involved in the formation of palea and lemma [36]. In this study we show that *HvMADS14* expression decreases slightly in *mov1*, while *HvMADS15* expression remains unaffected (**Appendix A, Supplementary Figure S2**). Based on these results, we propose that the A-class genes *HvMADS14* and *HvMADS15* share a conserved function in flower development to their rice counterparts.

Similarly, the rice E-class genes *OsMADS7* and *OsMADS8* are initially expressed in lodicule, stamen and carpel primordia. As development progresses, expression persists in the developing lodicules and stamens [37,38]. The significant decrease in *HvMADS7* and

HvMADS8 expression observed in *mov1* developing inflorescences is consistent with a conserved role in lodicule and stamen determination (**Figure 5**). Overall, the lack of both lodicules and stamens in *mov1* can be explained by reduced expression of all the genes involved in the formation of these floral organs.

Interestingly, expression of the barley *MEL1* orthologue, is significantly reduced in *mov1* plants (**Figure 5**). *MEL1* belongs to the *ARGONAUTE* gene family and in rice it is expressed in the male and female archesporial cells – the cells that divide to form the sporocyte - and in sporocytes themselves. *OsMEL1* functions to regulate correct cell division and correct progression of meiosis [22]. The reduction of *HvMEL1* transcript may reflect the absence of male archesporial cells and sporocytes due to lack of stamens, as well as the lack of female sporocytes in the ovule-like structures. Unexpectedly, the expression of *HvMSP1* does not seem to be affected (**Figure 5**). Rice *MSP1* encodes a receptor-like protein kinase and is crucial in limiting the number of cells entering into male and female sporogenesis by acting in the cells surrounding the sporocytes [23]. *MSP1* is also expressed in the inner wall layers in anthers where it functions in initiating anther cell wall formation [23]. If *HvMSP1* shared a conserved function with its rice counterpart, its expression would be predicted to decrease in *mov1* due to the lack of anthers and female sporocytes. It is possible that another *MSP1*-like gene fulfils this role in barley.

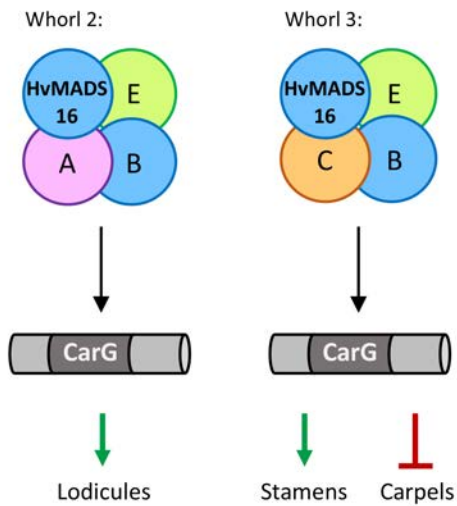
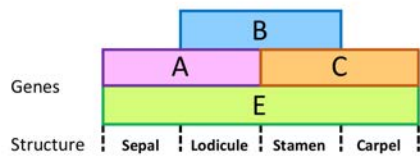
A model to explain barley flower development

Based on the results of the present study and on previous reports in barley and other plant species, we propose a model to explain the role of B-class MADS-box proteins in flower development in barley (**Figure 8**). In wild type, the B-class genes *HvMADS16*, *HvMADS2* and *HvMADS4* are expressed in the floral whorls 2 and 3. Here they can form DEF-GLO hetero-

dimers. Based on the floral quartet model, in whorl 2 the hetero-dimers associate in higher order protein complexes with A- and E-class genes. In whorl 3, instead, the floral quartet complex forms between B-, C- and E-class genes. The relevant quaternary complexes are then able to regulate the expression of target genes, eventually leading to the formation of lodicules in whorl 2 and promoting stamen development in whorl 3 while concomitantly repressing carpel formation. In *mov1*, *HvMADS16* is absent. The remaining B-class GLO-like proteins *HvMADS2* and *HvMADS4* do not interact and therefore cannot compensate for the absence of *HvMADS16*. Additionally, the lack of a DEF-GLO hetero-dimer results in an absence of the positive autoregulatory loop acting on B-class genes. Thus, the only protein complexes that can form in whorl 2 are between A- and E-class genes, and between C- and E-class genes in whorl 3. As a result of the genes regulated by these complexes of transcription factors, bract-like organs will develop in whorl 2 and carpel formation is promoted in whorl 3, leading to the multiovary phenotype observed in *mov1*, resembling a mutant in B-class function.

In conclusion, we show that *mov1* is very likely caused by the deletion of *HvMADS16*, a B-class gene necessary for specification of stamen and lodicule identity. The findings presented in this research support the hypothesis that the ABC model is perturbed in *mov1*, and that the correct higher-order protein complexes can thus not be formed appropriately. Developmentally, the absence of *HvMADS16* causes defects at the onset of meristem differentiation and results in completely sterile florets.

A. Wild type



B. *mov1*

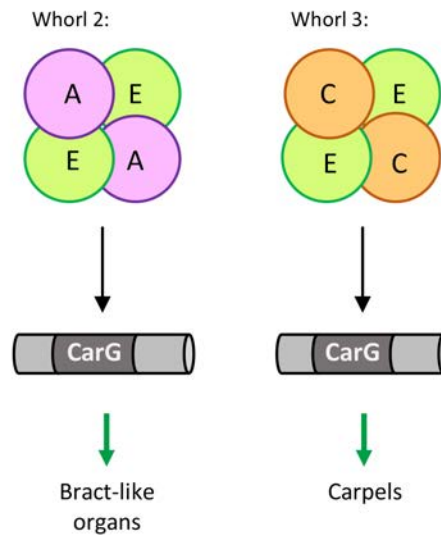
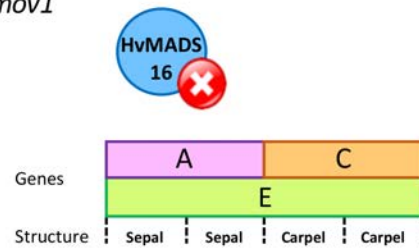


Figure 8. (A) In wild type, the B-class genes *HvMADS16* (*HORVU7Hr1G091210*), *HvMADS2* (*HORVU3Hr1G091000*) and *HvMADS4* (*HORVU1Hr1G063620*) are expressed in whorls 2 and 3. In whorl 2 the floral quartet between A-, B- and E-class genes promotes lodicule specification. In whorl 3, the floral quartet between B-, C- and E-class genes promotes stamen specification and represses carpel development. (B) In *mov1*, *HvMADS16* is absent and the remaining B-class genes do not form dimers. The only protein complex that forms in whorl 2 is between A- and E-class genes, which drives bract-like development, while the floral quartet in whorl 3 forms between C- and E-class genes, promoting carpel and ovule differentiation. *mov1* effectively resembles a mutant in B-class function.

Materials and Methods

Plant material and genotyping

Segregating seeds for the *mo6b* allele (cv. Steptoe) mutated at the *mov1* locus were kindly provided by Professor A. Kleinhofs. For phenotyping, segregating *mov1* plants were grown in the glasshouse (photoperiod maintained at ~ 12 hour over winter months with the use of artificial lighting, ~ 23/ ~ 16 °C day/night temperatures). Phenotypes were visually scored by manually opening the florets with tweezers. For a single plant, at least 6 central florets from 2-3 different spikes were scored to confirm phenotype. Images of florets were taken at Adelaide Microscopy (University of Adelaide) using a Nikon SMZ25 Stereo Fluorescence Microscope equipped with DS-Ri1 colour cooled digital camera. Image analysis and processing was carried out with the NIS-Elements AR software. For inflorescence dissection, plants were grown in a growth chamber with a 16-hour photoperiod and 23/15 °C (day/night) temperatures. All soil used was composed of 75 % (v/v) Coco Peat, 25 % (v/v) nursery cutting sand (sharp), 750 mg/L CaSO₄·2H₂O (gypsum) 750 mg/L Ca(H₂PO₄)₂·H₂O (superphosphate), 1.9 g/L FeSO₄, 125 mg/L FeEDTA, 1.9 g/L Ca(NO₃)₂, 750 mg/L Scotts Micromax micronutrients, and 2.5 g/L Osmocote Plus slow release fertilizer (16:3:9) (Scotts Australia Pty. Ltd.). pH was adjusted to 6.0 - 6.5 using 2 parts agricultural lime to 1 part hydrated lime.

Genotyping of the plants was performed by copy number analysis using TaqMan™ assay (Thermo Fisher Scientific, USA). Primers and TaqMan™ probes were designed with PrimerQuest software from IDT (Integrated DNA Technologies, USA) on the target sequence *HvMADS16* as gene of interest and barley *CONSTANS*-like *CO2* (*HORVU6Hr1G072620*) as internal positive control [38]; sequence of primers and probes can be found in **Appendix A, Supplementary Table S3A**. The reaction was set up as follows: 1x of PrecisionFast™ qRT-PCR Master Mix with Low ROX (Primerdesign Ltd., UK), 200 nM of each primer, 100 nM probe,

150 ng template DNA and water to a final volume of 10 μ L. The reaction was performed with an initial activation step at 95 °C for 2 minutes, followed by 40 cycles at: 95 °C for 5 seconds and 60 °C for 20 seconds in a QuantStudio 6 Flex Real-Time PCR machine (Thermo Fisher Scientific, USA). The detectors used were FAM-BHQ1 and HEX-BHQ1, with ROX as internal passive reference.

Nucleic acid extraction and genomic PCR

For all barley material, DNA was extracted from two weeks old plants using a phenol/chloroform method as described by Rogowsky *et al.* (1991) [40]. DNA concentration and quality were checked with NanoDrop™ (Thermo Fisher Scientific, USA).

PCR was performed using OneTaq DNA polymerase (New England BioLabs, USA), following the manufacturer's protocol in a final volume of 12.5 μ L. Primer T_m was calculated online at <http://tmcalculator.neb.com/#!/main>. Primers used are listed in **Appendix A, Supplementary Table S3B and S3C**. PCR products were stained with SYBR Safe (Thermo Fisher Scientific, USA), separated on 1 % agarose gel by electrophoresis and visualized under UV light. The DNA marker used was HyperLadder 1 Kb (Bioline, Australia). All PCR fragments were purified using ISOLATE II PCR and Gel Kit (Bioline, Australia) and sent to the Australian Genome Resource Facility (AGRF) for sequencing.

Microscopy

For scanning electron microscopy (SEM), inflorescences were manually dissected and fixed overnight in 4 % paraformaldehyde, 1.25 % glutaraldehyde in PBS, 4 % sucrose, pH 7.2. Before processing, samples were washed three times in PBS and fixed in 2 % OsO₄ in PBS for one hour. Samples were then dehydrated in a 50 - 100 % ethanol series and dried with a

critical point dryer. Dried samples were arranged on carbon tabs stuck to 12 mm aluminium stubs and coated with platinum. Samples were observed using a Hitachi FlexSem 1000 Scanning Electron Microscope at Adelaide Microscopy (University of Adelaide). For light microscopy, carpels were fixed in FAA solution (50 % Ethanol 100 %, 5 % acetic acid, 10 % formaldehyde 37 %, one drop of Tween-20) overnight and transferred to ethanol 70 % until further processing. Samples were dehydrated in an 70-100 % ethanol series and embedded in Technovit 7100 or LR white resin for wild type and *mov1*, respectively. Samples were sectioned using a Leica Rotary Microtome RM2265 at 5 μm or 1.5 μm , for Technovit 7100 and LR white resin respectively. Slides were stained with 0.1 % toluidine blue in 0.1 % sodium tetraborate for 2 minutes and rinsed three times with water. Slides were left to dry at 40 °C before mounting with DPX. After 72 hours slides were imaged at Adelaide Microscopy (University of Adelaide) with a Nikon Eclipse Ni-E optical microscope equipped with DS-Ri1 colour cooled digital camera. Image analysis and processing was carried out with the NIS-Elements AR software.

RNA extraction, cDNA synthesis and quantitative Real-Time PCR (qRT-PCR)

Wild-type and *mov1* inflorescences were manually dissected with fine-pointed tweezers at stages W2.0, W3.5, W4.0, W6.0, which correspond roughly to 17, 20, 23 and 26 Days Post Germination (DPG) according to the growing conditions used. Dissected inflorescences were snap frozen in liquid nitrogen and stored at -80 °C until further processing.

For the Steptoe tissue series, roots and coleoptiles were sampled at 2 DPG, leaves were collected at 17 DPG, inflorescences correspond to 26 DPG, internodes were sampled at 40 DPG, and caryopsis at 50 DPG.

RNA was extracted using the ISOLATE II RNA Plant Kit (Bioline, Australia), following the manufacturers' protocol. An additional DNase treatment was performed with TURBO™ DNase (Thermo Fisher Scientific, USA), followed by cDNA synthesis using SuperScript™ IV First Strand Synthesis (Thermo Fisher Scientific, USA) as per manufacturers' instructions. PCR amplification of the *HvGAPDH* (*HORVU7Hr1G074690*) gene was used to verify cDNA quality. Quantitative real-time PCR (qRT-PCR) was carried out following the method described in Burton *et al.* (2008) [41]. Namely, the specificity of each primer pair was verified by melt curve analysis and checked by HPLC purification and Sanger sequencing. For each amplicon, between four and six 20 µL PCR reaction mixtures (sequences of primers used are listed in **Appendix A, Supplementary Table S3D**) were combined for purification by HPLC using a HELIX DNA DVB 50 x 3.0 mm monolithic polymer reversed-phase column (Agilent Technologies, USA). Chromatography was performed using buffer A (100 mM triethylammonium acetate [Applied Biosystems] and 0.1 mM EDTA) and buffer B (100 mM triethylammonium acetate, 0.1 mM EDTA and 75 % acetonitrile). The size and PCR product identity were confirmed by Sanger sequencing. The product was then dried and dissolved in water to produce a 20 ng/µL stock solution which would then be used to create a dilution series covering orders of magnitude ranging from 10¹ - 10⁹ /µL copies of PCR product to use as standard.

For each timepoint or tissue, three technical replicates and at least three biological replicates were considered in the analysis. Quantitative real-time PCR (qRT-PCR) experiments were assembled by the liquid-handling robot CAS-1200 robot (Corbett Robotics, Australia) and the reaction was set up to contain 2 µL of template (either cDNA solution diluted 1:20, the diluted standard, or water as no template control), 5 µL of IQ SYBR Green PCR reagent (Bio-rad Laboratories, USA), 1.2 µL each of forward and reverse primers at 4 mM, 0.3 µL of 10x

SYBR Green in water, and water to reach a total volume of 10 μ L. Reactions were performed in a Biorad CFX384 3 minutes at 95 °C followed by 45 cycles of 1 second at 95°C, 1 second at 55 °C, 30 seconds at 72 °C, and 15 seconds at the optimal acquisition temperature.

The transcript levels of genes encoding barley glyceraldehyde 3-phosphate dehydrogenase (*HvGAPDH*), cyclophilin *HvCYCLO* (*HORVU6Hr1G012570*), α -tubulin *HvTUB* (*HORVU1Hr1G081280*), and heat shock protein 70 *HvHSP70* (*HORVU5Hr1G113180*) were used as controls for normalization. Normalization was carried out using multiple control genes as described by Burton *et al.* (2004) [42]. For each gene, transcript abundance is expressed as arbitrary units that represent the numbers of cDNA/ μ L, normalized against the geometric means of the three control genes that vary the least with respect to each other [43].

In situ hybridization

Sense and antisense RNA probes for *in situ* hybridization were PCR-amplified with the T7 promoter extension to the 5' of primers (primers used can be found in **Appendix A, Supplementary Table S3E**). The PCR product was separated on a 1 % agarose gel and purified from gel with ISOLATE II PCR and Gel Kit (Bioline, Australia). Approximately 3 μ g of template were used for *in vitro* T7 RNA polymerase (DIG RNA Labelling Kit, Roche, US) transcription to synthesize Digoxigenin (DIG)-labelled RNA probes. RNA probes were purified by precipitation with ethanol and lithium chloride and resuspended in DEPC-treated water to a concentration of approximately 500 ng/ μ L.

mov1 inflorescences were fixed in FAA solution (50 % ethanol, 4 % paraformaldehyde, 0.05 % glacial acetic acid and one drop of Tween-20 surfactant in DEPC-treated water). The tissue was dehydrated in a graded ethanol series, cleared with HistoChoice[®] Clearing Agent

(SIGMA, USA) and infiltrated with Parra pastillated Paramat Gurr™ (BDH Chemicals, UK) performed by the Meyerowitz Lab (<https://www.its.caltech.edu/~plantlab/protocols/insitu.html>). Samples were sectioned at 8 µm using a Leica Rotary Microtome RM2265 and mounted on Polysine™ glass slides (Thermo Fisher Scientific, USA).

In situ hybridization was performed using an InsituPro robot (Intavis, Germany). Slides were gently dewaxed and rehydrated, washed with PBS and treated with Proteinase K for 30 minutes at 37 °C. Slides were then washed with 4 % paraformaldehyde and dehydrated to 100 % ethanol. Slides were left to dry before incubating 16 - 20 hours with hybridization mix (50 % formamide, 300 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1x Denharts, 10 % dextranulphate, 10 mM DTT, 250 ng/mL tRNA, 100 µg/mL poly(A)) and 1000 - 1500 ng of probes at 55 °C. After incubation, slides were washed in 2x SSC/50 % formamide and treated with RNase A (ThermoFisher Scientific, USA). Immuno-labelling was performed using Anti-Digoxigenin-AP fragments (Roche, Switzerland). NBT-BCIP (Roche, Switzerland) was used for immune-detection as per manufacturer's instructions and incubated for 14 - 16 hours. Slides were imaged at Adelaide Microscopy (University of Adelaide) with a Nikon Eclipse Ni-E optical microscope equipped with DS-Ri1 colour cooled digital camera. Image analysis and processing was carried out with the NIS-Elements AR software.

Bi-molecular fluorescent complementation (BiFC)

The full-length coding sequences of *HvMADS2*, *HvMADS4*, *HvMADS16* and Δ *HvMADS16*, containing only the MADS-domain, were PCR-amplified with Q5® high-fidelity DNA polymerase (New England BioLabs, USA), from Steptoe inflorescence cDNA at 26 DPG using primers containing HindIII and XmaI restriction sites. Concomitantly, the BiFC vectors pSAT1-nEYFP-N1 (N-terminal fragment) and pSAT1-cEYFP-C1-B (C-terminal fragment) [44] were

digested using restriction enzymes HindIII and XmaI. Primers used for cloning are listed in **Appendix A, Supplementary Table S3F**. PCR fragments were digested, and then all components were stained with SYBR Safe (Thermo Fisher Scientific, USA), separated on 1 % agarose gel by electrophoresis and purified from gel using ISOLATE II PCR and Gel Kit (Bioline, Australia). Purified fragments were quantified with Nanodrop and ligated using T4 DNA Ligase (Thermo Fisher Scientific, USA) as per manufacturer's protocol.

For onion epidermal cells, the onion was cut into 2 × 2 cm squares and incubated at 28 °C on HO medium (200 mM D-sorbitol, 200 mM D-mannitol, 4.4 g/L MS media, 1.75 g/L Phytigel, pH 5.8) for 4 hours in the dark before transformation. For transformation, 15 µL of 1 µm gold particles (40 mg/mL) were coated with DNA as follows. The gold particles were transferred to 1.5 mL tubes and vortexed for 30 seconds at 1600 rpm. A mixture containing 5 µg of each plasmid DNA (1 µg/µL) was added to the gold particles and quickly vortexed at 2450 rpm. Meanwhile samples were vortexing, 5 µL of PM solution (40 % PEG-4000, 520 mM MgAc, pH 6.6) were added to the tube and additionally vortexed for 30 seconds. The DNA-gold mixture was incubated at room temperature for 20 minutes. After incubation the supernatant was gently removed, and the gold pellet was gradually resuspended in 45 µL of 100 % ethanol. The tube was centrifuged at maximum speed for 3 - 5 seconds to pellet the gold and the supernatant was promptly replaced with 20 µL of fresh 100 % ethanol. For biolistic, approximately 8 µL of DNA-gold mixture were pipetted onto the centre of a macrocarrier disc.

Biolistic was performed using a PDS-1000/He particle delivery system (Bio-Rad Laboratories, USA), with a helium pressure of 1100 psi, 28 inHg vacuum in the chamber and a 6 cm distance between the projectile and the samples. After bombardment, the onion tissue was incubated 24 hours at 28 °C before imaging. For imaging, the epidermal peel was removed from the

onion tissue using fine-pointed tweezers, placed in glycerol on glass slides and covered with a coverslip. Z-stack images of YFP were imaged using an excitation wavelength of 514 nm and emission wavelengths of 525 to 546 nm at Adelaide Microscopy (University of Adelaide) using a Nikon A1R Laser Scanning confocal microscope equipped with DS-Ri1 colour cooled digital camera. Image analysis and processing was carried out with the NIS-Elements C software.

Gene knockout by CRISPR

Vectors were provided by Prof. Yao-Guang Liu (South China Agricultural University). Guide RNA design and cloning for *HvMADS16* CRISPR knockout was performed as described by Ma *et al.* (2015) [45]. The primer sequences used for cloning are listed in **Appendix A, Supplementary Table S3G.**

Acknowledgments

This research was supported by the APA Scholarship from the Australian Government Research Training Program Scholarship and the ACPFG Supplementary Scholarship. We thank Dr. Gwenda Mayo from Adelaide Microscopy for the support and expertise. We would also like to thank Dr. Xiujuan Yang for protocols, material and technical advice essential to this project and Dr. David Marshall for providing gene sequences. We appreciate the invaluable help from Chao Ma and critical revision of the manuscript by Margaret Pallotta.

References

1. **Carpenter R, Coen ES.** Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes & Development*. 1990;4:1483–93.
2. **Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H.** Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*. *Science*. 1990;250:931–6.
3. **Coen ES, Meyerowitz EM.** The war of the whorls: genetic interactions controlling flower development. *Nature*. 1991;353:31–7.
4. **Colombo L, Franken J, Koetje E, van Went J, Dons HJ, Angenent GC, et al.** The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell*. 1995;7:1859–68.
5. **Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF.** B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*. 2000;405:200–3.
6. **Pinyopich A, Ditta GS, Savidge B, Liljgren SJ, Baumann E, Wisman E, et al.** Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*. 2003;424:85–8.
7. **Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, et al.** MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell*. 2003;15:2603–11.
8. **Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF.** The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol*. 2004;14:1935–40.
9. **Murai K, Euphytica KT.** Photoperiod-sensitive cytoplasmic male sterility in wheat with *Aegilops crassa* cytoplasm. *Euphytica*. 1993;67:41–8.
10. **Murai K, Takumi S, Koga H, Ogihara Y.** Pistillody, homeotic transformation of stamens into pistil-like structures, caused by nuclear-cytoplasm interaction in wheat. *Plant Journal*. 2002;29:169–81.
11. **Peng Z-S.** A new mutation in wheat producing three pistils in a floret. *Journal of Agronomy and Crop Science*. 2003;189:270–2.
12. **Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S, Murai K.** Class D and B_{sister} MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). *Plant Mol. Biol*. 2009;71:1–14.
13. **Wang QH, Yang ZJ, Wei SH, Jiang ZY, Yang YF, Hu ZS, et al.** Molecular cloning, characterization and expression analysis of *WAG-1* in the pistillody line of common wheat. *Genet. Mol. Res*. 2015;14:12455–65.
14. **Yang Z, Peng Z, Wei S, Liao M, Yu Y, Jang Z.** Pistillody mutant reveals key insights into stamen and pistil development in wheat (*Triticum aestivum* L.). *BMC Genomics*. BioMed Central; 2015;16:211–10.
15. **Barley Genetics Newsletter.** 2013;pp.48–223.
16. **Tazhin.** The linkage of the genes *mo5** and *n* in barley. *Barley Genetics Newsletter*. 1980;10(II):69–72.
17. **Soule JD, Skodova, Kudrna DA, A K, Kleinhofs A.** Molecular and genetic characterization of barley flower development mutants. *Barley Genetics Newsletter*. 1996;76–80.
18. **Soule JD, Kudrna DA, Kleinhofs A.** Isolation, mapping, and characterization of two barley multiovary mutants. *J Hered*. 2000;91:483–7.
19. **Callens C, Tucker MR, Zhang D, Wilson ZA.** Dissecting the role of MADS-box genes in monocot floral development and diversity. *Journal of Experimental Botany*. 2018;69:2435–59.
20. **Colmsee C, Beier S, Himmelbach A, Schmutzer T, Stein N, Scholz U, et al.** BARLEX - The barley draft genome explorer. *Molecular Plant*. 2015;8:964–6.
21. **Waddington SR, Cartwright PM.** A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*. 1983;51:119–30.
22. **Nonomura K-I, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, et al.** A germ cell-specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell*. 2007;19:2583–94.
23. **Nonomura K-I, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, et al.** The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*. 2003;15:1728–39.
24. **Yang X, Li G, Tian Y, Song Y, Liang W, Zhang D.** A rice glutamyl-tRNA synthetase modulates early anther cell division and patterning. *Plant Physiology*. 2018;177:728–44.
25. **Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, et al.** *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development*. 2003;130:705–18.
26. **Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ.** Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell*. 2000;5:569–79.

27. **Jack T, Brockman LL, Meyerowitz EM.** The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*. 1992;68:683–97.
28. **Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, et al.** Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal*. 1992;11:251–63.
29. **Goto K, Meyerowitz EM.** Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Development*. 1994;8:1548–60.
30. **McGonigle B, Bouhidel K, Irish VF.** Nuclear localization of the *Arabidopsis* APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes & Development*. Cold Spring Harbor Lab; 1996;10:1812–21.
31. **Dreni L, Pilatone A, Yun D, Erreni S, Pajoro A, Caporali E, et al.** Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. *Plant Cell*. 2011;23:2850–63.
32. **Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ.** Diversification of C-function activity in maize flower development. *Science*. 1996;274:1537–40.
33. **Yamaguchi T.** Functional diversification of the two C-class mads box genes *OSMADS3* and *OSMADS58* in *Oryza sativa*. *Plant Cell*. 2006;18:15–28.
34. **Ohmori Y, Toriba T, Nakamura H, Ichikawa H, Hirano H-Y.** Temporal and spatial regulation of *DROOPING LEAF* gene expression that promotes midrib formation in rice. *Plant Journal*. 2011;65:77–86.
35. **Lopez-Dee ZP, Wittich P, Enrico Pè M, Rigola D, Del Buono I, Gorla MS, et al.** *OsMADS13*, a novel rice MADS-box gene expressed during ovule development. *Dev. Genet*. 1999;25:237–44.
36. **Wu F, Shi X, Lin X, Liu Y, Chong K, Theissen G, et al.** The ABCs of flower development: mutational analysis of *AP1/FUL*-like genes in rice provides evidence for a homeotic (A)-function in grasses. *Plant Journal*. 2017;89:310–24.
37. **Pelucchi N, Fornara F, Favalli C, Masiero S, Lago C, Pè EM, et al.** Comparative analysis of rice MADS-box genes expressed during flower development. *Sex Plant Reprod*. 2002;15:113–22.
38. **Kang H-G, Jang S, Chung J-E, Cho Y-G, An G.** Characterization of two rice MADS box genes that control flowering time. *Molecules & Cells*. 1997;7.
39. **Bartlett JG, Alves SC, Smedley M, Snape JW, Harwood WA.** High-throughput *Agrobacterium*-mediated barley transformation. *Plant Methods* 2008;4:1–12.
40. **Rogowsky PM, Guidet FL, Langridge P, Shepherd KW, Koebner RM.** Isolation and characterization of wheat-rye recombinants involving chromosome arm 1DS of wheat. *Theor Appl Genet*. 1991;82:537–44.
41. **Burton RA, Jobling SA, Harvey AJ, Shirley NJ, Mather DE, Bacic A, et al.** The genetics and transcriptional profiles of the cellulose synthase-like *HvCslF* gene family in barley. *Plant Physiology*. 2008;146:1821–33.
42. **Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB.** The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiology*. 2004;134:224–36.
43. **Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*. 2002;3.
44. **Citovsky V, Lee L-Y, Vyas S, Glick E, Chen M-H, Vainstein A, et al.** Subcellular localization of interacting proteins by Bimolecular Fluorescence Complementation *in planta*. *Journal of Molecular Biology*. 2006;362:1120–31.
45. **Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, et al.** A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant*. 2015;8:1274–84.

Chapter 4



Statement of Authorship

Title of Paper	The role of <i>HvSTAMENLESS1 (HvSL1)</i> in regulating barley floral development
Publication Status	Unpublished and unsubmitted work written in manuscript style.

By signing the Statement of Authorship, each author certifies that:

- i. each author's contribution to the manuscript is accurate; and
- ii. permission is granted to include the manuscript in the thesis

Principal Author

Name of Principal Author	Caterina Selva		
Contribution to the Paper	Designed and performed the experiments. Analysed and interpreted the results. Wrote the manuscript.		
Percentage of Contribution	80 %		
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	18/09/2019

Co-author Contributions

Name of Co-author	Dr. Neil Shirley		
Contribution to the Paper	helped in SNP discovery for mapping, performed the qRT-PCR experiment and analysed the qRT-PCR data. I hereby certify that the Statement of Authorship is accurate.		

U

Contribution to the Paper	supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of		
Signature		Date	3/09/2019

Name of Co-author	Dr. Ute Baumann		
	edited the manuscript. I hereby certify that the Statement of authorship is accurate.		

Contribution to the Paper	supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of authorship is accurate.		
Signature		Date	3/09/2019

The role of *HvSTAMENLESS1* in regulating barley flower development

C. Selva, N. Shirley, R. Whitford, U. Baumann, M. R. Tucker

School of Agriculture Food and Wine, University of Adelaide, Waite Campus, Urrbrae 5064, South Australia, Australia

Abstract

We report mapping of the barley multiovary locus *mov2* to the short arm of chromosome 3H and identify *HvSL1*, a C2H2 zinc finger transcription factor, as the most likely candidate. Florets of mutant *mov2.g* plants lack stamens, but exhibit functional supernumerary carpels leading to multiple seeds per floret when artificially pollinated. Based on developmental, genetic and molecular results we propose a model for stamen and carpel specification in barley, with *HvSL1* regulating the B-class gene *HvMADS16*, particularly in floral whorl 3.

Introduction

Understanding flower development is the first step towards providing the knowledge about how floral structures can be modified for agricultural use. This can be achieved through prediction of key genes and CRISPR-knockout by reverse genetics, but also by using forward genetics to exploit naturally variable germplasm and diverse mutant resources. Recent advances in genomic technologies have highlighted the utility of mutant populations as a

tool for gene discovery in diploid cereal species such as barley [1,2], and complex polyploid species such as wheat [3]. However, historical mutant screens uncovered a considerable array of phenotypes (<https://www.nordgen.org/bgs/>), some of which led to a dramatically altered number of reproductive organs. Until now, many barley multiovary mutants have been partially described in the literature [4-8]. However; their genetic location, gene content and inter-relationship are still largely unknown. Among them, the *multiovary2* (*mov2*) locus has been roughly mapped to the short arm of chromosome 3H and suggested to be a MADS-box containing gene [9]. In this study we characterize the recessive *mov2.g* allele, derived from a fast neutron irradiated population of the cultivar Steptoe generated in the early 1990s [8]. *mov2.g* plants have previously been reported to lack stamens and to contain complex carpel-like structures [8,9]. In this study, we map the *mov2* locus to a region on 3H containing a C2H2 zinc finger transcription factor, named here as *HvSL1* based on homology to rice *OsSL1* (*STAMENLESS1*). As the *mov2.g* phenotype closely resembles that of plants with reduced B-class activity, we report in detail how *mov2.g* affects flower development and investigate interactors of *HvSL1* in the context of the ABC model, in particular promoters of B-class gene sequences.

Results

The multiovary phenotype of *mov2.g* results in additional fertile ovules and multiple viable seeds per floret

The floral organs in a wild-type barley floret are arranged in a defined pattern. The outermost whorl (whorl 1) contains the palea and lemma, two lodicules are present in whorl 2, while three stamens in whorl 3 surround a single carpel at the centre of the floret (whorl 4). In contrast, florets of *mov2.g* present a multiovary phenotype whereby there is a homeotic

conversion of all stamens into supernumerary carpels, thus rendering the plant unable to self-pollinate (**Figure 1**).

The number of carpels within *mov2.g* florets varies, but typically ranges from 5 - 7 carpels or carpel-like structures. Visually, it was not possible to distinguish the central carpel from the additional organs as all carpels were irregularly shaped, joined at the base, and appeared smaller relative to wild type (**Figure 1**). In contrast, development and morphology of the lemma and palea were unaffected in *mov2.g*. Likewise, the development of lodicules remained largely unchanged. Sporadically, *mov2.g* florets contained one wild-type lodicule and one lodicule showing partial conversion into a bract-like organ. Importantly, when *mov2.g* plants were used as a female recipient in artificial pollination, some florets were able to produce multiple seeds per floret (**Figure 1**). After manual pollination, a maximum of three seeds were observed to develop within a single floret. All three seeds were viable, and after germination gave rise to mature plants. In contrast to wild-type plants, however, *mov2.g* homozygous seedlings tended to have reduced early vigour. In most cases, it was also observed that seeds from heterozygous plants, still segregating for the *mov2.g* phenotype, presented an overall germination rate of ~ 50 % or lower (**Appendix B, Supplementary Table S1**). Of the seeds that germinated, the proportion of *mov5.o* was usually observed to be lower (6 – 10 %) than expected for a single locus recessive trait (25 %) (**Appendix B, Supplementary Table S2A**).

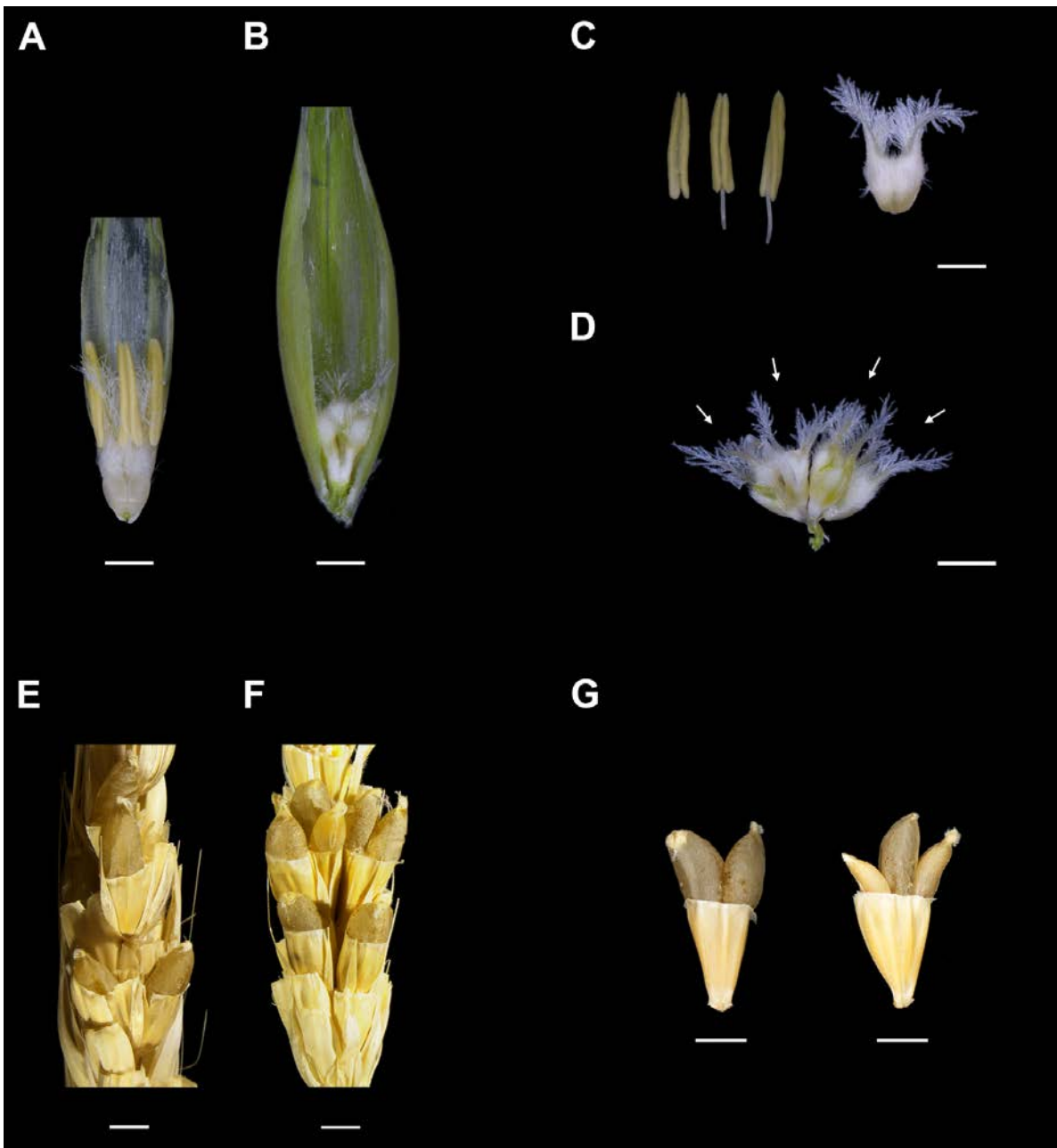


Figure 1. Exposed (A) wild-type and (B) *mov2.g* florets. (C) The reproductive organs in a wild-type floret consist of 3 stamens and 1 carpel. (D) In *mov2.g* florets the stamens are converted into additional supernumerary carpel-like structures (arrows). Scale bars: 1000 μm . (E) A wild-type spike produces one seed per floret. (F-G) Artificially pollinated *mov2.g* spikes can produce multiple seeds per floret. Scale bars: 2000 μm .

Developmental defects appear after meristem differentiation in *mov2.g* inflorescences

To understand the basis of the multiovary phenotype, scanning electron microscopy (SEM) was used to observe wild-type (cv. Steptoe) and *mov2.g* inflorescence development. Immature inflorescences of both genotypes were morphologically comparable at the very early stages of spike development, corresponding to Waddington stages W3.0 (lemma/floret primordium) [10] (**Figure 2**). At stage W3.5 – 3.75 (stamen primordia), both wild-type and *mov2.g* floral meristems displayed lateral protrusions, consistent with the appearance of stamen primordia from the floral meristem (**Figure 2**). Following this stage, the first morphological differences were identified. In wild-type barley, the lateral protrusions give rise to mature stamens, while the central area of meristematic cells proliferate and differentiate into a single ovule-containing carpel (W5.0 – 8.5) (**Figure 2**). As each wild-type floral meristem develops, it maintains a vertical symmetry along the central inflorescence rachis. However, when compared to wild type, as development proceeds (W5.0 – 8.5) the lateral protrusions in *mov2.g* floral meristems do not differentiate into stamens, but instead give rise to organs which follow the characteristic morphogenesis of carpels. Specifically, these carpel-like organs appear to initially develop an ovule primordium which is later surrounded by the growing carpel tissue. These additional carpel-like structures also differentiate stigmas bearing stigmatic papillae (W8.5) and appear to develop synchronously with the central carpel, resulting in a unit formed by multiple carpel-like structures, each basally joined together and to the central carpel (**Figure 2**). Moreover, as floral organs develop there is frequently a loss in the vertical symmetry of florets along the central inflorescence rachis and distortion of the inflorescence tip, which is reflected in a shorter and broader spike morphology compared to wild type (**Appendix B, Supplementary Figure S1**).

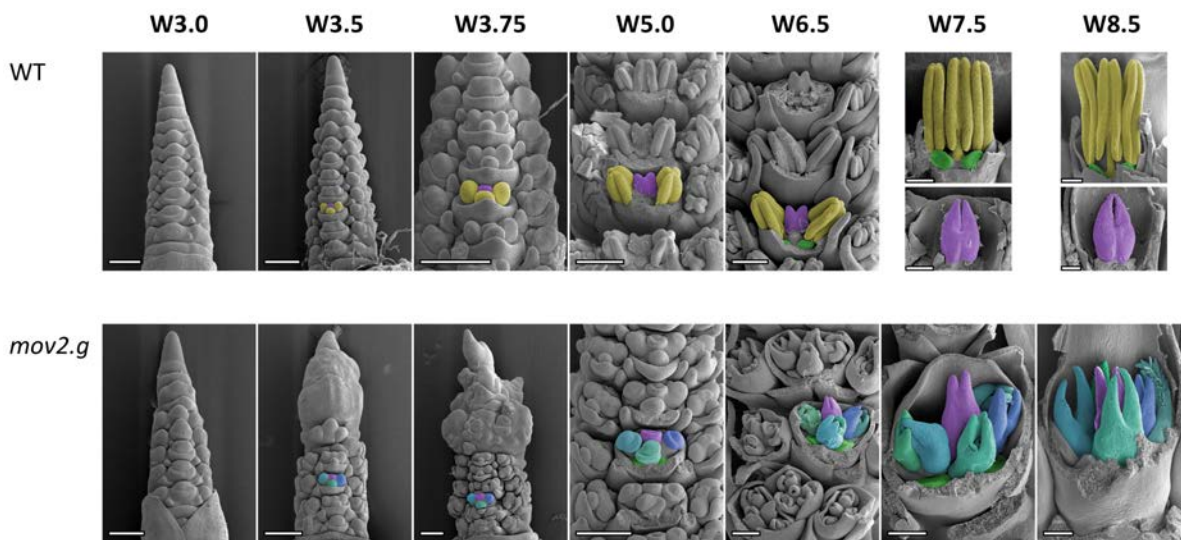


Figure 2. SEM of barley inflorescence development in wild type (WT) and *mov2.g*. For wild type, primordia giving rise to stamens are false-coloured in yellow, cells giving rise to the carpel in purple and lodicules in green. In *mov2.g*, the central carpel (purple) and lodicules (green) appear to be retained, cells giving rise to the additional carpel-like structures are false-coloured in blue (different shades). Waddington stage is indicated for each developmental timepoint, from W5.0 lemma and/or stamens have been removed to expose the carpels. Scale bars: 200 μ m.

A C2H2 zinc finger transcription factor underlies the *mov2* locus

A previous study mapped the *mov2* locus to the telomeric region on the short arm of chromosome 3H (3HS), between molecular markers ABC171A and JS001B [9]. However, the genomic interval in which *mov2* resides (\sim 28 Mb) was deemed too large to reliably identify the underlying causative gene sequence(s). In order to map *mov2* at a higher genetic resolution, a *mov2.g* (cv. Steptoe) x Morex bi-parental population was developed. For the purpose of confirming if the parental cross was successful, six seeds that set on the *mov2.g* male-sterile female parent were assayed both phenotypically and genotypically. For all six F_1 plants heterozygosity was confirmed by KASPTM marker analysis across two known SNPs, at positions chr3H_1006543 and chr3H_28805649 according to the Morex reference assembly (Hv_IBSC_PGSB_v2) and by Sanger sequencing across a known SNP within gene *HORVU7Hr1G091210*. Phenotypically, F_1 individuals possessed intermediate growth

characteristics (**Appendix B, Supplementary Figure S2**) and flowering time (personal observation) relative to both parental lines. The mutant multiovary phenotype was not observed in these individuals suggesting that the *mov2.g* mutation is completely recessive. In total, 352 F₂ plants derived from four confirmed F₁ individuals (88 F₂ plants from each F₁ individual) were grown until maturity and subsequently used for genetic mapping. For each F₂ plant, phenotyping was performed at flowering by visual inspection of 5 - 13 florets across 1 - 3 individual spikes. The *mov2.g* multiovary phenotype segregated as a single Mendelian recessive gene (3 wild-type : 1 mutant) in F₂ plants derived from only two of the F₁ lines (V144_XE8-2 and V144_XE8-3) (**Appendix B, Supplementary Table S2B**). Phenotypic segregation of F₂ individuals derived from the other two F₁ lines (V144_XE8-5 and V144_XE8-3-8) was observed to be skewed from the expected 3:1, with only 14 % (12/88) and 6 % (5/88) of F₂ plants expressing the multiovary phenotype (**Appendix B, Supplementary Table S2B**). For fine-mapping of the *mov2* locus, SNPs spanning the 3HS region were identified and developed as KASP™ markers. SNPs were identified either from a published Steptoe x Morex dataset [11] or from in-house Steptoe leaf transcriptomic data that was mapped to the reference Hv_IBSC_PGSB_v2 Morex assembly. Confirmation that *mov2* mapped to the correct 3H locus was achieved using seven KASP™ markers designed to span the previously published interval (~ 28 Mb). These were tested on a subset of 48 plants from the *mov2.g* x Morex F₂ population, of which 20 individuals presented the *mov2.g* phenotype. Following this, an additional 10 KASP™ markers were designed to saturate the previously published interval, facilitating the reduction of the critical *mov2* containing region to approximately 1.9 Mb, based on flanking markers at positions chr3H_9095799 and chr3H_11039299 (**Figure 3**). Further resolution of the locus was achieved using 179 F₃ individuals derived from carefully selected recombinant F₂ plants that were genotypically heterozygous across the target locus.

These F₃ individuals segregated for the multiovary phenotype and were genotyped with six additional KASP™ markers spanning the 1.9 Mb region. Comparing phenotypic to genotypic segregation across these F₃ plants allowed reduction of the critical interval to roughly 449 Kb, based on the Morex reference sequence (barley genome Scaffold_1432, *personal communication* Dr. Martin Mascher, IPK Gatersleben, Germany), between markers chr3H_9748112 and chr3H_10289104 (Figure 3). The aforementioned scaffold contains 20 annotated gene sequences (Appendix B, Supplementary Table S3).

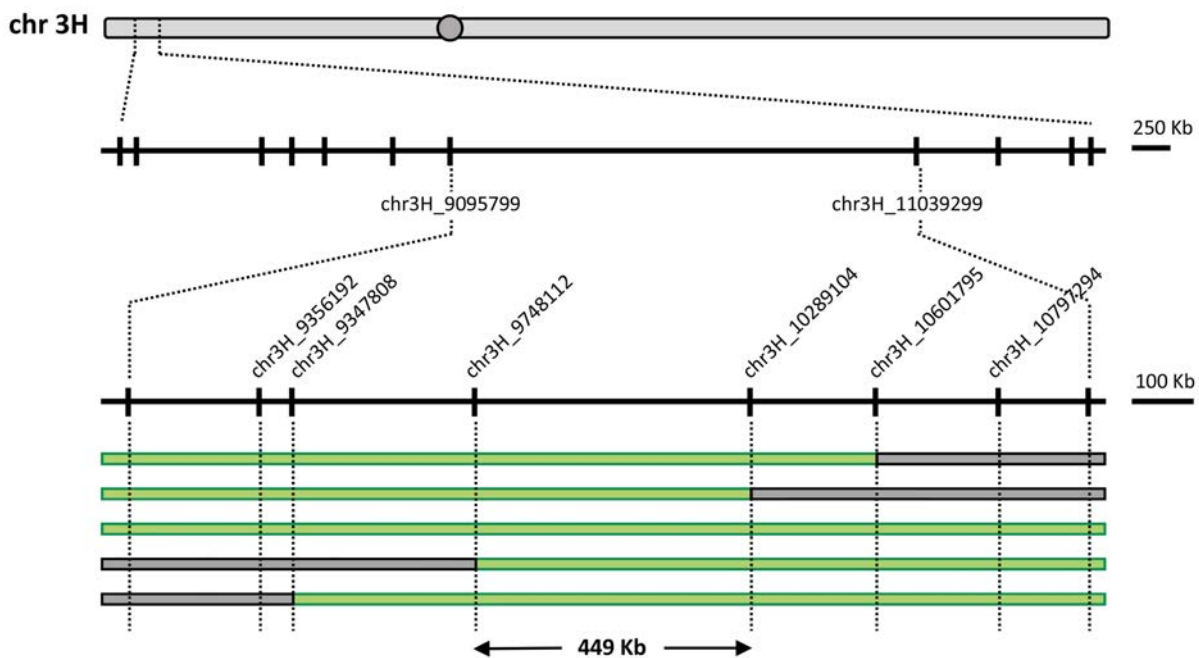


Figure 3. Mapping of the *mov2* locus in a *mov2.g* x Morex bi-parental population. *mov2* was initially mapped to a ~ 1.9 Mb interval between markers chr3H_9095799 and chr3H_11039299 based on 352 F₂ segregants. Fine mapping using 179 F₃ recombinants reduced the critical interval to ~ 449 Kb between markers chr3H_9748112 and chr3H_10289104. Marker order is based on the genetic map. Examples of mapping in F₃ recombinants exhibiting *mov2.g* phenotype are shown; green: *mov2.g* allele; grey: Morex allele.

For the purpose of identifying the causative gene sequence, each of the 20 annotated gene sequences within the interval were assessed for expression in floral tissues, especially in

tissue types affected in *mov2.g* florets. Barley expression datasets were obtained from both in-house and unpublished resources, including Steptoe and *mov2.g* leaf (Zadoks stage Z22) [12] transcriptomes (unpublished), as well as transcriptomes from 2-week old seedlings (Z12) [13], different stages of developing inflorescences (W2.0, W3.5 and W8.0 - 8.5) [13] and from developing pistils (W8.0 – 10.0) (*L. Wilkinson, M. Tucker, unpublished*).

Overall, ten of the 20 annotated gene sequences showed expression in either pistil or inflorescence tissues. Of the nine gene sequences, only *HORVU3Hr1G003740* was identified to be uniquely expressed in reproductive tissues and not vegetative tissues (leaf and seedling stage) nor in *mov2.g* leaf (**Figure 4**). Sequence analysis indicated that *HORVU3Hr1G003740* encodes a putative C2H2 zinc finger transcription factor (**Figure 5A**) sharing 65.6 % sequence identity with the rice protein LOC_Os01g03840, also known as STAMENLESS1 (SL1). In rice, *OsSL1* is known to play a crucial role in floral development, with loss of *OsSL1* function leading to a multiovary phenotype [14]. Thus, *HORVU3Hr1G003740* is hereafter referred to as *HvSL1*.

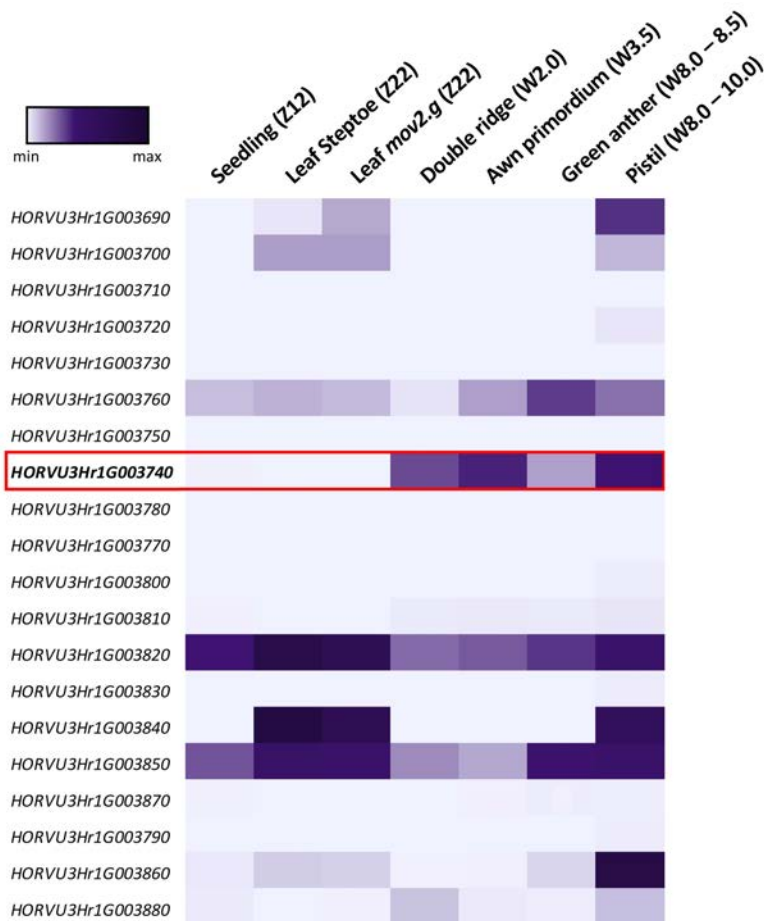
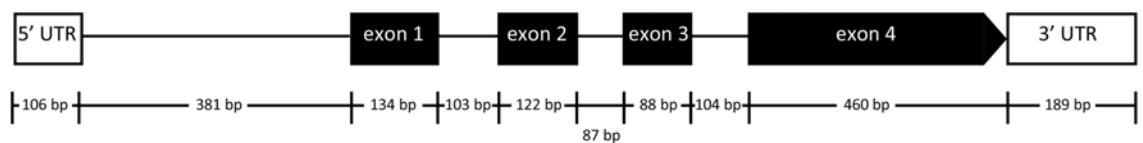


Figure 4. Heatmap of gene expression (RNAseq values) in different plant tissues for genes present in the mapped critical interval for *mov2*. Gene order is based on Morex Scaffold_1432 (Dr. Martin Mascher, IPK Gatersleben, Germany). *HvSL1* (*HORVU3Hr1G003740*) is indicated in bold. For each tissue, development based on the Zadoks or Waddington growth scale is provided.

To determine if a mutation in *HvSL1* might contribute to the *mov2.g* phenotype, PCR primers were designed to amplify a 2199 bp region spanning the entire *HvSL1* coding sequence. Comparison of PCR results from Morex, Steptoe and *mov2.g* genomic DNA suggested that *HvSL1* is absent in *mov2.g* plants (**Figure 5B**). By contrast, predicted high-confidence neighbouring gene sequences were still present, indicating that *HvSL1* could be the only deleted gene at the *mov2* locus (**Appendix B, Supplementary Table S3**). However, presence of low-confidence neighbouring genes could not be tested due to repetitiveness of the sequence. For additional confirmation, *HvSL1*-specific PCR was repeated on 36 critical

recombinant F₃ individuals derived from the *mov2.g* x Morex mapping population, of which 16 (44.4 %) exhibited the multiovary phenotype. For all samples, the absence of *HvSL1* corresponded with the mutant phenotype. Although this alone does not provide conclusive proof, it appears likely that the multiovary phenotype in *mov2.g* is due to deletion of the *HvSL1* gene.

A. *HvSL1* gene structure



B. *HvSL1* in *mov2.g*

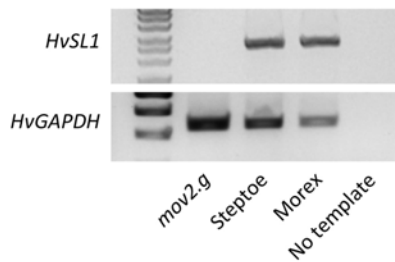


Figure 5. (A) Gene structure of *HvSL1* (*HORVU3Hr1G003740*); length in base pairs of Untranslated Regions (UTR), exons (black) and introns (solid line) is indicated (B) *HvSL1* (*HORVU3Hr1G003740*) appears to be absent in *mov2.g* plants when assayed by PCR. The barley glyceraldehyde 3-phosphate dehydrogenase *HvGAPDH* gene (*HORVU7Hr1G074690*) was used as reference.

HvSL1 is expressed early in flower development and the expression of ABC genes is altered in *mov2.g*

To establish how the *mov2.g* mutation might influence the expression of floral regulators from the ABC-class genes, quantitative real-time PCR (qRT-PCR) was performed on developing inflorescences from wild type and *mov2.g* at stages W2.0 to W6.0. This time-

course captures the specification and differentiation stages for both the male and female reproductive organs.

First, *HvSL1* expression was investigated. In wild type, *HvSL1* transcript abundance steadily increased between stages W2.0 (double ridge) through to W3.5 (stamen primordia) and W4.5 (carpel primordium) after which it decreased towards stage W6.0 (carpel development). As expected, *HvSL1* expression was completely absent in developing inflorescences from *mov2.g* plants (**Figure 6**).

When comparing transcript abundance for the three barley B-class genes, the greatest difference was observed for *HvMADS16*, which was significantly reduced in *mov2.g* relative to wild type as development progressed (**Figure 6**). Although the overall transcript abundance was reduced, it did temporally increase with development. A similar trend was observed also for the B-class gene *HvMADS2*, while for the remaining B-class gene *HvMADS4*, transcript abundance was initially lower in *mov2.g* until stamen primordia specification (stage W3.5). After stage W3.5, transcript abundance for *HvMADS4* steadily increased to match wild-type levels by stage W6.0 (**Figure 6**). These results suggest that the *mov2.g* mutation affects transcript levels of all three B-class genes.

This contrasts with the trend observed for carpel-specific (*HvMADS58*) and ovule-specific (*HvMADS3*) MADS-box genes for which transcript abundance in the *mov2.g* mutant showed a delayed accumulation relative to wild type, followed by an increase which eventually exceeded wild type (**Figure 6; Appendix, Supplementary Figure S3**). Developing *mov2.g* inflorescences also exhibited significantly higher levels of the meristem-specific gene *HvOSH1* (*HOMEODOMAIN 1*) transcript at most timepoints, particularly as development progressed (**Figure 6**). Conversely, transcript abundance of ovule specific D-class gene *HvMADS13*, A-class genes *HvMADS14* and *HvMADS15*, in addition to the carpel-specific

YABBY transcription factor *HvDL* (*DROOPING LEAF*) and the germline regulator *HvMSP1* (*MULTIPLE SPOROCTE*), remained unaffected in *mov2.g* samples (**Figure 6; Appendix, Supplementary Figure S3**). Likewise, transcript abundance of the E-class genes *HvMADS7* and *HvMADS8* reached similar levels in wild type and *mov2.g*, despite a delay in accumulation (**Appendix, Supplementary Figure S3**). Strikingly, transcripts for another germline gene *HvMEL1* (*MEIOSIS ARRESTED AT LEPTOTENE1*) showed a distinct, but not significant, reduction in *mov2.g* inflorescences (**Appendix, Supplementary Figure S3**).

Wild-type — *mov2.g* —

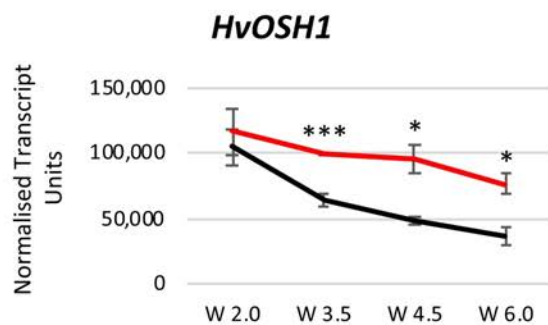
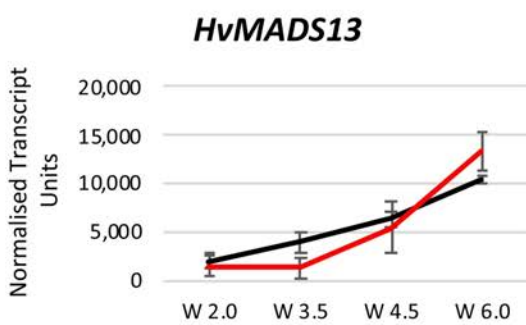
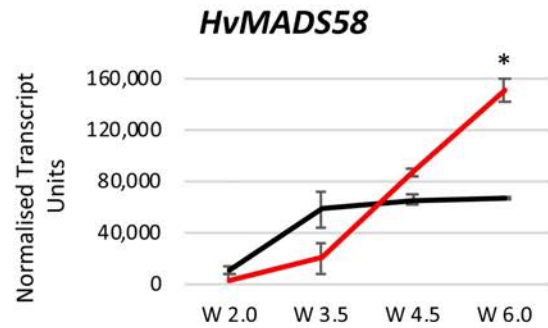
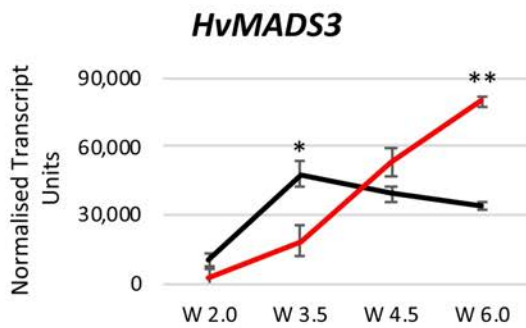
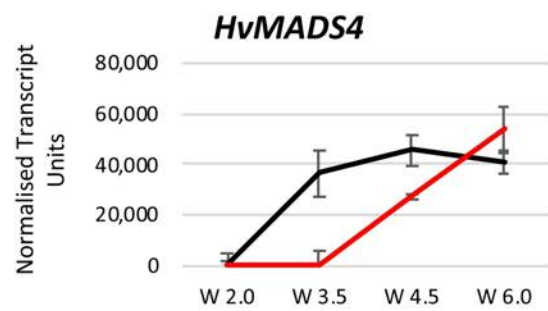
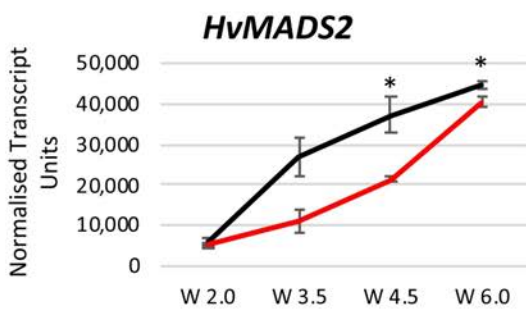
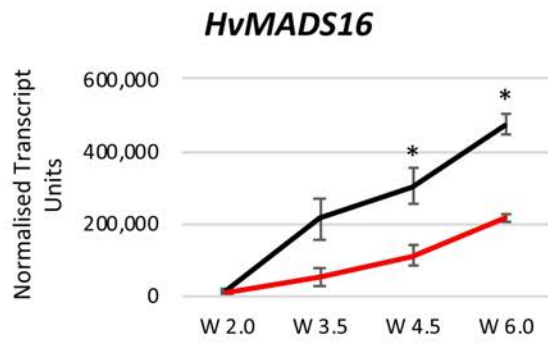
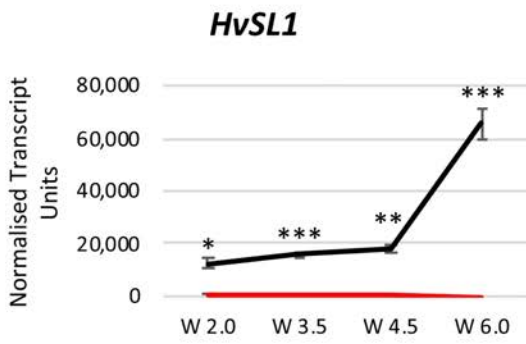


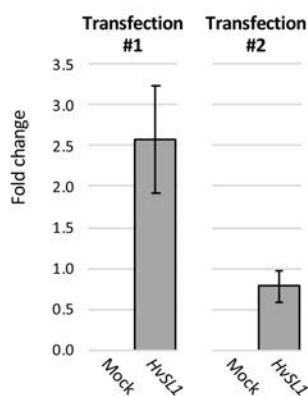
Figure 6. Transcript abundance assessed by qRT-PCR in wild-type (black) and *mov2.g* (red) developing inflorescences at stages W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development) for: *HvSL1* (*HORVU3Hr1G003740*), B-class genes *HvMADS16* (*HORVU7Hr1G091210*), *HvMADS2* (*HORVU3Hr1G091000*), *HvMADS4* (*HORVU1Hr1G063620*); C-class genes *HvMADS3* (*HORVU3Hr1G026650*) and *HvMADS58* (*HORVU1Hr1G029220*); D-class gene *HvMADS13* (*HORVU1Hr1G023620*) and *HvOSH1* (*HORVU4Hr1G009730*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**) and ≤ 0.001 (***) are shown for differences between wild type and *mov2.g*. For each sample *n* = 3 independent biological replicates.

HvSL1 alone is not sufficient to directly activate the promoters of MADS-box B-class genes

Studies conducted in rice suggest that OsSL1 acts as an upstream positive regulator of *OsMADS16* transcription [14]. An *in silico* analysis of the 3 Kb putative promoter sequence upstream of the translation start site for *HvMADS2*, *HvMADS4* and *HvMADS16* predicted multiple potential C2H2 transcription factor binding sites (TFBS) for each gene. For *HvMADS2*, the identified TFBSs showed a 75 % similarity score to the *Arabidopsis* motif used as input, while the similarity score reached ≥ 80 % for the TFBSs predicted for both *HvMADS4* and *HvMADS16* (**Appendix B, Supplementary Figure S4**). To determine whether a direct interaction occurs in barley between *HvSL1* and MADS-box B-class genes, a dual luciferase reporter assay was performed in isolated barley protoplasts. Firstly, protoplast transfection efficiency was assayed using a plasmid designed to constitutively express YFP. For each sample, transfection efficiency was calculated by counting and averaging the number of protoplasts expressing YFP in three representative images. Six independent transfections suggested that transformation efficiency was variable, ranging from 59.8 to 96.8 %, but successful nonetheless (**Appendix B, Supplementary Figure S5**). Likewise, constitutive expression of *HvSL1* driven by pCaMV35S promoter worked when transfected alone but appeared to be quite variable leading to a 0.8 to 2.6-fold increase in *HvSL1* expression

compared to expression of housekeeping genes (n = 2 independent transfections). As expected, no *HvSL1* expression was detected in protoplasts transfected with MMG solution (mock) (**Figure 7**). For each B-class gene, the 3 Kb putative promoter sequence was fused to the dual luciferase reporter system. Results indicated that for all B-class genes, the relative luciferase activity did not significantly change in the presence of *HvSL1* over-expression, compared to protoplasts transfected with the luciferase reporter alone (n = 3 independent transfections) (**Figure 7**).

A. *HvSL1* expression in protoplasts



B. DLR assay in protoplasts

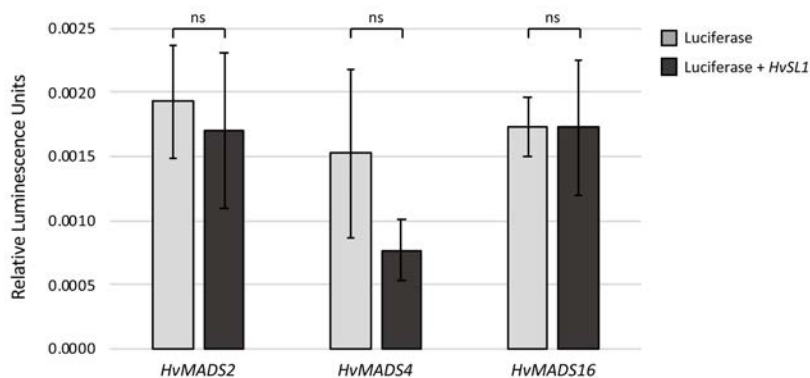


Figure 7. (A) *HvSL1* (*HORVU3Hr1G003740*) expression by qRT-PCR in protoplasts transfected with mock or with a construct driving constitutive *HvSL1* expression. Fold change is reported relative to housekeeping genes barley glyceraldehyde 3-phosphate dehydrogenase and cyclophilin. **(B)** DLR assay in protoplasts transfected with the sole dual luciferase reporter (light grey) or co-transfected with both the dual luciferase reporter and *HvSL1* (dark grey). The luciferase reporter is regulated by the putative promoter (3 Kb) of each B-class gene, while *HvSL1* is constitutively expressed by the Cauliflower Mosaic Virus 35S promoter. ns indicates no significant difference. For further vector details refer **Supplementary Figure S5**.

Discussion

The barley multiovary *mov2.g* phenotype is likely caused by deletion of the zinc finger transcription factor *HvSL1*

In previous studies, *mov2* was mapped to the short-arm telomeric region of chromosome 3H (3HS) [9]. In the present study, the location of *mov2* was confirmed in a *mov2.g* (cv. Steptoe) x Morex F₃ population, further refining its position to a 449 Kb interval (**Figure 3**). Analysis of this region in the Morex reference sequence was unable to confirm previous suggestions that a MADS-box gene resides within the *mov2* locus. Rather, sequence analysis identified the HORVU3Hr1G003740 gene, which encodes a putative C2H2 zinc finger transcription factor and is the likely orthologue of *OsSL1*. The HORVU3Hr1G003740 gene, referred to as *HvSL1*, appears to be the most likely candidate responsible for the *mov2.g* phenotype. Indeed, among the genes annotated in the mapped interval, *HvSL1* appears to be the only gene that is physically absent in *mov2.g* plants (**Figure 5**). This type of deletion is consistent with fast-neutron mutagenesis. The absence of *HvSL1* completely co-segregates with the mutant trait as determined by analysis of 36 F₃ lines derived from critical recombinants across the locus interval. *HvSL1* expression is also completely abolished in developing inflorescences from *mov2.g* plants compared with wild type (**Figure 6**). Furthermore, the *mov2.g* multiovary phenotype closely resembles that of *stamenless1* (*sl1*) mutants in rice [14], whose causative gene sequence is a direct orthologue of *HvSL1*. It is worth noting, however, that slight phenotypic differences are present between *mov2.g* and rice *sl1*. In particular, we did not observe a high rate of abnormal lodicules, bract-like organs between the additional carpels, nor incomplete conversions of stamens into carpels in *mov2.g* plants. Similarly, phenotypic differences are present between *HvSL1* and the respective orthologues in *Arabidopsis* *JAGGED* (*JAG*) and *NUBBIN* (*NUB*). Although both *JAG* and *NUB* seem to be involved in correct

stamen and carpel development, lack of function in these genes does not alter organ identity, but rather organ morphogenesis [15,16]. Furthermore, the role of *JAG* and *NUB* as shape determinants is not restricted to the reproductive organs but also affects the outer floral whorls as well as vegetative tissues, indicating a more general role in proper lateral organ shape, especially for *JAG* [15,16].

Despite the strong evidence presented in this work, conclusive proof is needed in order to confirm that *HvSL1* is responsible for the *mov2.g* multiovary phenotype. For this reason, although it is beyond the timeframe of the current study, experiments have been initiated to generate *Hvs/1*-knockout plants by CRISPR/Cas9 technology. Two different guide RNAs, positioned at + 39 bp and + 269 bp downstream of the transcriptional start site, were designed to target the first and second *HvSL1* exon, respectively (**Appendix B, Supplementary Figure S6**). Transformation of barley (cv. Golden Promise) is currently being performed.

The most striking feature of *mov2.g* spikes is the ability to produce multiple seeds per floret upon cross-pollination, with up to three seeds developing simultaneously within the same floret (**Figure 1**). These seeds are able to germinate and produce plants, suggesting that the ovules and female gametophytes of the additional carpels are correctly differentiated. Production of multiple seeds per floret is in accord with initial reports in the 1950s of multiovary *mo* mutants in the barley cv. Trebi [5,6]. Nonetheless, allelism tests between *mov2.g* and the Trebi-derived *mo* multiovary mutant are yet to be performed due to seed unavailability [17]. Although the description of the early multiovary mutants does not completely reflect the phenotype of *mov2.g* florets, this might be a consequence of differences in genetic backgrounds (Steptoe and Trebi), or types of mutation possibly due to

different kinds or doses of mutagen. In particular, *mo* mutants (cv. Trebi) exhibited incomplete penetrance of the multiovary phenotype, with some florets retaining one or two normal stamens, or showing incomplete conversion of stamens to carpels [5]. Furthermore, florets of *mo* plants were commonly observed to contain a single lodicule [6]. These characteristics were not observed in *mov2.g* (cv. Steptoe) plants in our growing conditions. Although the *mov2.g* trait has been reported to be genetically monofactorial recessive [17], we observed that in some cases the frequency of detectable mutants from segregating heterozygous parents was much lower (6 – 10 %) than expected (25 %). This was seen both in F₂ individuals derived from the *mov2.g* x Morex population as well as during genotyping by copy number of plant material used for the other experiments (**Appendix B, Supplementary Table S2**). Indeed, a previous study on the same mutant also reported a distorted segregation ratio in F₂ plants [8,9]. It was suggested that the observed low proportion of mutant plants might be a consequence of likely additional mutations and/or chromosomal rearrangements that would have a negative impact on survival or transmission of gamete carrying the *mov2.g* allele [9]. An alternative hypothesis could be that the skewed segregation seen in some segregating families and the low germination (≤ 50 %) of *mov2.g* heterozygous seeds could be indicative of an additional role of *HvSL1* in seed viability and/or germination apart from that in floral development. However, additional studies are needed to explore this hypothesis further.

Shifting the balance of the ABC genes possibly contributes to the multiovary *mov2* phenotype

Mov2 appears to be necessary for the correct development of stamens. As seen from SEM, the initial phases of inflorescence differentiation are similar in *mov2.g* and wild type until

stage W3.75 (stamen primordia). At this stage, floral meristems of both *mov2.g* and wild type develop lateral protrusions (**Figure 2**). However, it is hard to distinguish if the lateral protrusions in *mov2.g* are indeed stamen primordia which fail to develop due to insufficient levels of B-class genes or if the protrusions are in reality a symptom of the incipient initiation of additional carpels. Further experimentation, possibly looking at expression domain of ABC genes by *in situ hybridization* at these early stages is needed to discern between these two alternative hypotheses.

Overall, the overview of the ABC genes obtained from transcript abundance studies in *mov2.g* is in accordance with the resulting phenotype. In particular, the reduced expression in *mov2.g* inflorescences of all three B-class genes *HvMADS2*, *HvMADS4* and *HvMADS16* is likely to explain the lack of stamens. Nevertheless, it is worth noting that expression of these genes is not completely abolished and could be sufficient to drive normal lodicule development. Indeed, transcript abundance of both *HvMADS2* and *HvMADS4* reach levels that are comparable to wild type at stage W6.0 (carpel development) (**Figure 6**). Based on the qRT-PCR results and phenotype of *mov2.g* florets, we speculate that the action of *HvSL1* is prominent in the third floral whorl (stamens). Following this hypothesis, the occasional partial conversion of lodicules into bract-like organs is probably a secondary effect resulting from the combined low levels of *HvMADS2* and *HvMADS16* rather than a direct action of *HvSL1* in whorl 2 (lodicules). This would be in contrast to rice *SL1* which has been shown by *in situ hybridization* to be expressed in both whorls 2 and 3 [14]. However, it is still plausible that *HvSL1* could have a role in lodicule development, even if less striking. To this effect, it would be ideal to perform *in situ hybridization* in wild-type and *mov2.g* inflorescences to determine the expression domain of *HvSL1*. The initial lower levels of *HvMADS2* and

HvMADS4 seen by qRT-PCR could also be explained by the low levels of *HvMADS16* driving a slow start of the positive autoregulatory feedback loop typical of B-class genes [18,19].

Noticeably, *HvOSH1*, which is predicted to act in the floral meristem based on homology to rice, is present at higher levels in *mov2.g* (**Figure 6**). This result suggests that a larger meristem niche is present in the multiovary mutant, or that the meristem is maintained for longer. In either case, an increase in *HvOSH1* expression is in agreement with the higher number of floral organs (5 - 7 carpels and lodicules) present in *mov2.g* compared to wild-type florets. Furthermore, the increased transcript abundances of carpel and ovule-specific transcription factors belonging to the C-class (*HvMADS3* and *HvMADS58*) MADS-box genes (**Figure 6**) also support the presence of additional carpels and ovules in *mov2.g* florets. Interestingly, despite the phenotype, the expression levels of the D-class (*HvMADS13*) MADS-box gene and of the carpel-specific *HvDL* gene remain unaffected in *mov2.g* with respect to wild type (**Figure 6; Appendix B, Supplementary Figure S3**). This may indicate that *HvMADS13* and *HvDL* play a minor role than other MADS-box genes in the formation of additional ovules and carpels in *mov2.g*. Alternatively, *HvDL* in *mov2.g* may have slower expression dynamics that might have not been captured entirely. Indeed, most of the investigated genes seem to have a delayed response in *mov2.g*. This observation, together with the smaller plant size for homozygote mutants might be indicative of delayed floral development in the multiovary mutant.

Transcript abundance levels of A-class (*HvMADS14* and *HvMADS15*) and E-class (*HvMADS7* and *HvMADS8*) genes in *mov2.g* are comparable to wild-type levels for most timepoints (**Appendix B, Supplementary Figure S3**). Since A-class genes are predominantly involved in specification of whorls 1 and 2, the qRT-PCR results are consistent with the observation that *mov2.g* florets do not show any major phenotypic disruption in these outermost whorls. In

the case of the E-class genes, the only developmental stage in which transcript abundance decreases significantly in *mov2.g* compared to wild type is at W3.5. This stage corresponds to the initiation of stamen primordia initiation and is in accord with a role of *HvMADS7* and *HvMADS8* in stamen and lodicule differentiation as seen for rice [20,21]. The rescue of their expression at later timepoints may indeed explain the normal lodicule development seen in *mov2.g*.

Based on studies in rice, *HvMEL1* is predicted to regulate correct male and female germ cell development [22]. Accordingly, expression of *HvMEL1* is decreased in *mov2.g* plants with respect to wild type (**Appendix B, Supplementary Figure S3**), possibly due to lack of the male germline as a result of the absence of anthers. Lower *HvMEL1* expression is still retained in *mov2.g* developing inflorescence, indicative that the female germline may be present in the additional carpel-like structures.

Finally, as seen for *mov1* (Chapter 3), transcript abundance of *HvMSP1* gene, presumed from rice to regulate sporogenesis [23], does not show significant differences between mutant and wild-type plants (**Appendix B, Supplementary Figure S3**). Once more, this argues in favour of *HvMSP1* having an alternate function to its rice counterpart.

Even though the lack of *HvSL1* in *mov2.g* broadly affects the genes involved in the ABC model, results from the dual luciferase reporter assay in isolated barley protoplasts suggest that *HvSL1* alone is not sufficient to regulate the activity of B-class genes (**Figure 7**). This observation may indicate either that *HvSL1* does not regulate the activity of B-class genes or that regulation of B-class genes by *HvSL1* is indirect. Alternatively, *HvSL1* may act as a direct regulator by recruitment of other co-factors or via additional *cis*-elements missing in our experiment. It is thus a starting point for new hypotheses which should be addressed by

further experimentation. It is worth mentioning that in our particular case, *in silico* analysis of barley putative promoter sequences (3 Kb) using an input motif from *Arabidopsis* did not provide reliable predictions of TFBSs. However, this could not be known *a priori* and choice was determined by the restricted input motifs that were available.

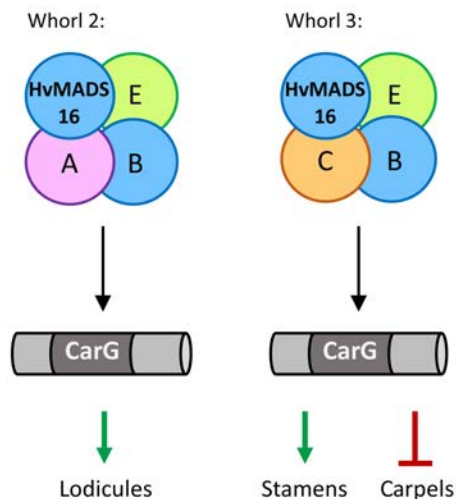
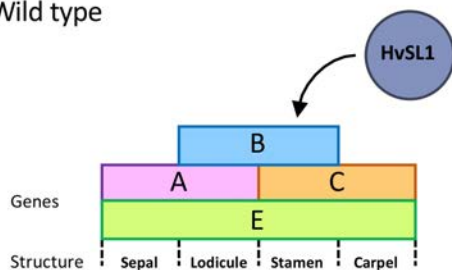
Adding components to barley flower development

When considering the arrangement of floral organs into whorls, *mov2.g* plants only show distinct homeotic conversion of stamens (whorl 3), while floral organs in the other whorls remain unaffected. For this reason, we propose that *Mov2* predominantly functions in whorl 3 and is involved in correct regulation of B-class genes, particularly *HvMADS16*. In wild type, formation of the correct hetero-dimer between a *DEFICIENS*-like (*HvMADS16*) and a *GLOBOSA*-like (*HvMADS2* and *HvMADS4*) B-class genes in whorls 2 and 3 can support the assembly of appropriate quaternary complexes of MADS-box transcription factors to initiate the cascade of downstream signalling [24]. This results in lodicules being formed in whorl 2 and stamens developing in whorl 3 (**Figure 8**).

The lack of *HvSL1* in *mov2.g* plants results in a shift in the balance of the ABC model. At a molecular level, we speculate that B-class activity in *mov2.g* flowers is predominantly maintained in whorl 2 (lodicules) where enough *HvMADS16* is present to give rise to normal lodicules as in wild type. At the same time, the reduced levels of *HvMADS16* in whorl 3 are not sufficient to successfully form functional B-class hetero-dimers. As a consequence, the predominant quaternary complexes forming in whorl 3 are between C- and E- class genes which will promote carpel formation (**Figure 8**). Also, concomitant possible expansion of the expression domain of ovule-promoting genes to whorl 3 and/or a greater meristem niche in *mov2.g* florets results in multiple carpels having a functional female gametophyte.

In conclusion, we have demonstrated that a C2H2 zinc finger transcription factor, termed here *HvSL1*, is absent in homozygous *mov2.g* plants and is located at the *mov2* locus. This finding is in contrast to previous studies proposing that a MADS-domain containing gene sequence is responsible for the *mov2* multiovary phenotype [9]. In *mov2.g* plants, defects appear early in inflorescence development but remain restricted to the third floral whorl, suggesting that *Mov2* is necessary for normal stamen development in barley. Indeed, even though *Mov2* alone is not able to directly influence B-class gene activation, lack of *Mov2/HvSL1* perturbs expression of genes that give rise to the reproductive organs. Phenotypically, the final outcome of the underlying genetic changes is the presence of a male-sterile floret containing supernumerary carpels capable of supporting production of multiple seed set per floret.

A. Wild type



B. *mov2.g*

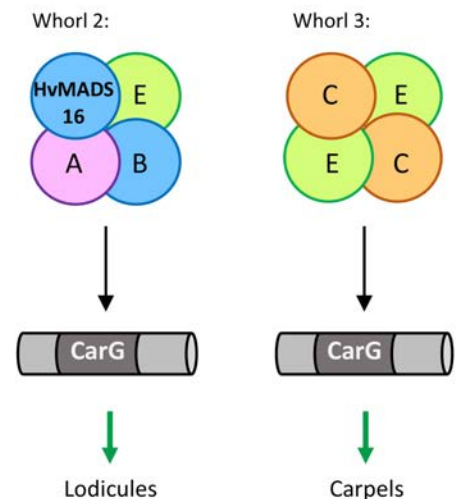
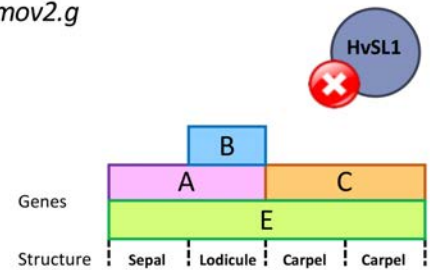


Figure 8. (A) In wild type, the B-class genes *HvMADS16* (HORVU7Hr1G091210), *HvMADS2* (HORVU3Hr1G091000) and *HvMADS4* (HORVU1Hr1G063620) are expressed in whorls 2 and 3. In whorl 2 the floral quartet between A-, B- and E-class genes promotes lodicule specification. In whorl 3, the floral quartet between B-, C- and E-class genes promotes stamen specification and represses carpel development. (B) In *mov2.g*, *HvMADS16* expression is mostly confined to whorl 2 due to absence of the positive regulator *HvSL1* (HORVU3Hr1G003740). The remaining B-class genes and residual expression of *HvMADS16* in whorl 3 is not sufficient to form hetero-dimers. As a result, the most frequent protein complex that forms in whorl 3 is between C- and E-class genes, promoting carpel and ovule differentiation.

Materials and Methods

Plant material and genotyping

Segregating seeds for the *mov2.g* allele (cv. Steptoe) mutated at the *mov2* locus were kindly provided by professor A. Kleinhofs. Growing conditions, soil preparation and phenotyping of plant material were performed as described in Chapter 3. Similar growing conditions and phenotyping methods were also used for the bi-parental mapping population. DNA for KASP assay was extracted from two weeks-old leaf tissue. KASP™ assays were designed using Primer Picker (LGC Genomics) and prepared using the LGC Genomics SNPLINE™. Assays were performed using KASP™ Master mix, following the manufacturer's instructions. Sequence of KASP™ markers can be found in **Appendix B, Supplementary Table S4**.

Genotyping of segregating *mov2.g* plants was performed by copy number analysis using TaqMan™ assay (Thermo Fisher Scientific, USA). Primers and TaqMan™ probes were designed with PrimerQuest software from IDT (Integrated DNA Technologies, USA) on the target sequence *HvSL1* as gene of interest and barley *CONSTANS*-like *CO2* (*HORVU6Hr1G072620*) as internal positive control [25]. Sequence of primers and probes can be found in **Appendix B, Supplementary Table S5A**. The reaction set up and reaction conditions used are the same as those described in Chapter 3. The reaction was performed in a QuantStudio 6 Flex Real-Time PCR machine (Thermo Fisher Scientific, USA), using FAM-BHQ1 and HEX-BHQ1 as detectors, and ROX as internal passive reference.

Nucleic acid extraction and PCR

For all barley material, genomic DNA was extracted from freeze-dried 2-week old plant material as described in Kovalchuk (2014) [26]. PCR reactions were prepared, performed,

separated on agarose gel and Sanger sequenced as mentioned in Chapter 3. Primers used for PCR are listed in **Appendix B, Supplementary Table S5B**.

RNA extraction, cDNA synthesis and quantitative real-time PCR (qRT-PCR)

Wild-type and *mov2.g* inflorescences were manually dissected with fine-pointed tweezers at timepoints 17, 20, 23 and 26 Days Post Germination (DPG), which correspond approximately to developmental stages W2.0, W3.5, W4.0, W6.0 in the growing conditions used. Dissected inflorescences were processed as described in Chapter 3 for RNA extraction, cDNA synthesis, qRT-PCR and output analysis. Sequence of primers used for qRT-PCR are listed in **Appendix B, Supplementary Table S5C**.

Scanning electron microscopy (SEM)

For SEM, wild-type and *mov2.g* inflorescences were processed and imaged as mentioned in Chapter 3.

In silico analysis of putative promoters

Zinc finger transcription factor binding sites were predicted *in silico* with JASPAR 7th release (<http://jaspar.genereg.net>) [27]. For each B-class gene, a 3 Kb sequence upstream of the translational start site was scanned for the only available zinc finger motif, which belonged to *Arabidopsis thaliana* (ID MA1277.1).

Protoplast isolation

Isolation of barley leaf protoplasts was performed as described by Yoo *et al.* (2007) [28] with minor modifications. Briefly, the adaxial epidermal layer of leaves from 11-day old barley

seedlings (cv. Golden Promise) were individually peeled and the leaf was cut into approximately 2 cm x 0.5 cm strips using surgical blades. The leaf segments of approximately 10 plants were immediately transferred in a Petri dish containing 15 mL of 0.6 M mannitol for 30 minutes at room temperature to induce plasmolysis. After incubation, the leaf segments were transferred to another Petri dish containing 10 mL of freshly prepared enzyme solution [0.55 M mannitol, 40 mM MES-KOH at pH 5.7, 20 mM KCl, 2.0 % cellulase R10 (Yakult Pharmaceutical, Japan), 0.75 % macerozyme R10 (Yakult Pharmaceutical, Japan), 10 mM CaCl₂, 0.1 % BSA] and incubated for 3 hours in the dark at 28 °C with gentle shaking (40-60 rpm). After enzymatic digestion, forceps were used to gently remove the remaining epidermis and leaf debris from the enzyme solution. An equal volume (10 mL) of W5 solution (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl and 2 mM MES-KOH at pH 5.7) was slowly added to the protoplasts and the solution was filtered with a 100 µM nylon mesh to a 50 mL round-bottom tube. Volume was adjusted by adding 5 mL of W5 solution. The filtered protoplasts were collected by centrifugation at 600 g for 3 minutes. Supernatant was replaced with 15 mL of fresh W5 and the protoplasts resuspended by gentle shaking. Protoplasts were allowed to pellet by gravity for 30 minutes in ice. After incubation, the supernatant was promptly removed and substituted with MMG solution (0.6 M mannitol, 15 mM MgCl₂ and 4 mM MES-KOH at pH 5.7) at a concentration of 10⁶ cells/mL, determined by counting cells in 12 µL of a 1:10 dilution of protoplast solution with a haematocytometer.

PEG-mediated transfection of barley protoplasts

PEG-mediated transfections were mostly carried out as described by Bai *et al.* (2014) [29]. Firstly, 200 µL of PEG-Ca²⁺ solution [40 % (w/v) PEG 4000, 0.4 M mannitol, 0.1 M CaCl₂] were pre-loaded in pipette tips. Secondly, 100 µL of protoplast solution (approximately 10⁵ cells)

were added to 20 µg of each plasmid DNA in a 2.0 mL tube. The pre-loaded PEG-Ca²⁺ solution was immediately added to the protoplast-DNA mixture, mixed gently and incubated for 15 minutes at room temperature in the dark. The transfection process was stopped by adding 840 µL of W5 and centrifuged at 600 g for two minutes. Cells were resuspended in 500 µL W5 and transferred to multi-well plates previously coated with 5 % (v/v) FBS. Protoplasts were cultured at 28 °C for 40 - 48 hours in the dark.

To test transfection efficiency, protoplasts were transfected with pUbi-YFP-rbcS in six independent transfections. 40 - 48 hours after transfection the protoplasts were visualized under UV and bright light at Adelaide Microscopy (University of Adelaide) using a Nikon Eclipse Ni-E optical microscope equipped with DS-Ri1 colour cooled digital camera. Image analysis and processing was carried out with the NIS-Elements AR software.

Cloning of luciferase reporters and dual luciferase assay

Constructs for transfection in barely protoplasts were cloned by restriction digestion. For the transcription factor *HvSL1*, the entire coding sequence was PCR-amplified with Q5[®] high-fidelity DNA polymerase (New England BioLabs, USA), from Steptoe inflorescence cDNA at 26 DPG with primers containing HindIII and EcoRI sites. The empty vector pSAT1-nEYFP-N1 [30] was used to amplify pCaMV35S and t35S with primers containing SacII and HindIII restriction sites for the promoter and EcoRI and Ascl sites for the terminator. Primers used for cloning can be found in **Appendix B, Supplementary Table S5D**. An empty pENTR/D-TOPO (Thermo Fisher Scientific, USA) digested with SacII and Ascl was used as the vector's backbone. PCR fragments were digested, and then all components were stained with SYBR Safe (Thermo Fisher Scientific, USA), separated on 1 % agarose gel by electrophoresis and purified from gel using ISOLATE II PCR and Gel Kit (Bioline, Australia). Purified fragments were

quantified with Nanodrop and ligated using T4 DNA Ligase (Thermo Fisher Scientific, USA) as per manufacturer's protocol to form pENTR-pCaMV35S:HvSL1:t35S vector.

For the luciferase reporters, a 3 Kb promoter sequence for genes *HvMADS2*, *HvMADS4*, *HvMADS16* was amplified using Q5[®] high-fidelity DNA polymerase (New England BioLabs, USA) from Morex genomic DNA using primers containing Apal and SacII restriction sites. Vector pGreenII-0800-LUC from Hellens *et al.* (2005) [31] was digested with Apal and SacII enzymes to use as backbone. Like mentioned before, PCR fragments were digested, and then all components were stained with SYBR Safe (Thermo Fisher Scientific, USA), separated on 1 % agarose gel by electrophoresis and purified from gel using ISOLATE II PCR and Gel Kit (Bioline, Australia). Purified fragments were quantified with Nanodrop and ligated using T4 DNA Ligase (Thermo Fisher Scientific, USA) as per manufacturer's protocol to form pGreen-0800-HvMADS2-LUC, pGreen-0800-HvMADS4-LUC and pGreen-0800-HvMADS16-LUC vectors.

pENTR-pCaMV35S:HvSL1:t35S and all luciferase reporters were transformed in 25 μ L of OneShot Mach1 T1 *E. coli* competent cells (Thermo Fisher Scientific, USA) and the plasmids were purified, checked by digestion and Sanger sequenced at the Australian Genome Resource Facility (AGRF).

To assess *HvSL1* expression, protoplasts were transfected in two independent experiments with pENTR-pCaMV35S:HvSL1:t35S alone. 40-48 hours after incubation RNA was extracted using a phenol-chloroform method. Briefly, in a 1.5 mL tube samples were lysed and homogenized in 500 μ L TRIzol (Thermo Fisher Scientific, USA) before adding 150 μ L chloroform. Samples were vortexed and centrifuged at maximum speed at 4°C for 15 minutes. The aqueous phase was carefully transferred to a new 1.5 mL tube with 250 μ L

isopropanol and incubated at 4 °C for 10 minutes before centrifuging at maximum speed at 4 °C for 20 minutes. The precipitated RNA was washed with 500 µL 100 % ethanol, allowed to air-dry and resuspended in 20 µL DEPC-treated water. cDNA synthesis and qRT-PCR were performed as previously described in Chapter 3. Changes in *HvSL1* gene expression are reported as fold change relative to the most stable reference genes glyceraldehyde 3-phosphate dehydrogenase *HvGAPDH* (*HORVU7Hr1G074690*) and cyclophilin *HvCYCLO* (*HORVU6Hr1G012570*), as calculated using the qBASE+ software [32].

For Dual Luciferase® Reporter Assay System – DLR (Promega Corporation, USA) – the reagents were prepared following the manufacturer’s instructions. Transfected barley protoplasts after 40 - 48 hours incubation at 28 °C in the dark were transferred to 2.0 mL tubes and lysed by adding 80 µL of PLB and incubating for 15 minutes at room temperature with occasional shaking. 35 µL of supernatant were then dispensed in OptiPlate-96 Black (PerkinElmer, USA) multi-well plates. DLR assay was performed following the manufacturer’s suggestions, in a GloMax® 96 Microplate Luminometer (Promega corporation, USA) using an injection volume of 100 µL for both reagents LARII and Stop&Glo, a delay between wells of 0.4 seconds and a reading time of 10 seconds for each well. Measurements are the average of three independent transfections.

Gene knockout by CRISPR

Vectors were provided by Prof. Yao-Guang Liu (South China Agricultural University). Guide RNA design and cloning for *HvSL1* CRISPR knockout followed as described in Chapter 3. The primer sequences used for cloning are listed in **Appendix B, Supplementary Table S5E**.

Acknowledgments

This research was supported by the APA Scholarship from the Australian Government Research Training Program Scholarship and the ACPFG Supplementary. We are grateful to Dr. Xiujuan Yang, Dr. Guillermo Garcia-Gimenez and Anzu Okada for providing technical assistance with CRISPR knockouts and protoplast isolation. We would like to thank Dr. Laura Wilkinson and Dr. Gang Li for sharing transcriptomic datasets. We acknowledge Dr. Gwenda Mayo from Adelaide Microscopy, Dr. Martin Mascher for providing scaffold sequence and Margaret Pallotta for critical revision of the manuscript.

References

1. Szarejko I, Szurman-Zubrzycka M, Nawrot M, Marzec M, Gruszka D, Kurowska M, *et al.* Creation of a TILLING population in barley after chemical mutagenesis with sodium azide and MNU. Jankowicz-Cieslak J, Tai TH, Kumlehn J, Till BJ, editors. *Biotechnologies for Plant Mutation Breeding: Protocols*. Springer. 2017. pp. 91–111.
2. Schreiber M, Barakate A, Uzrek N, Macaulay M, Sourdille A, Morris J, *et al.* A highly mutagenised barley (cv. Golden Promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods*. 2019;15:99.
3. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, *et al.* Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci.* 2017;114:E913–21.
4. Gregory FJ, Purvis ON. Abnormal flower development in barley involving sex reversal. *Nature*. 1947;160:221–2.
5. Moh CC, Nilan RA. Multi-ovary in barley: a mutant induced by atomic bomb irradiation. *J Hered.* 1953;44:183–4.
6. Kamra OP, Nilan RA. Multi-ovary in barley floral anatomy and embryo-sac development. *J Hered.* 1959;50:159–66.
7. Tazhin. The linkage of the genes *mo5** and *n* in barley. *Barley Genetics Newsletter*. 1980;10(II):69–72.
8. Soule JD, Skodova, Kudrna DA, A K, Kleinhofs A. Molecular and genetic characterization of barley flower development mutants. *Barley Genetics Newsletter*. 1996;76–80.
9. Soule JD, Kudrna DA, Kleinhofs A. Isolation, mapping, and characterization of two barley multiovary mutants. *J Hered.* 2000;91:483–7.
10. Waddington SR, Cartwright PM. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*. 1983;51:119–30.
11. Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, *et al.* Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics*. 2009;10:582.
12. Zadoks JC, Chang TT, Konzak CF. A decimal code for the growth stages of cereals. *Weed Research*. John Wiley & Sons, Ltd (10.1111); 1974;14:415–21.
13. Liu H, Li G, Yang X, Kuijter HNJ, Liang W, Zhang D. Transcriptome profiling reveals phase-specific gene expression in the developing barley inflorescence. *The Crop Journal*. 2019;1–16.
14. Xiao H, Tang J, Li Y, Wang W, Li X, Jin L, *et al.* *STAMENLESS 1*, encoding a single C2H2 zinc finger protein, regulates floral organ identity in rice. *Plant Journal*. 2009;59:789–801.
15. Ohno CK, Reddy GV, Heisler MGB, Meyerowitz EM. The *Arabidopsis JAGGED* gene encodes a zinc finger protein that promotes leaf tissue development. *Development*. 2004;1111-1122.
16. Dinneny JR, Weigel D, Yanofsky MF. *NUBBIN* and *JAGGED* define stamen and carpel shape in *Arabidopsis*. *Development*. 2006;1645-1655.
17. *Barley Genetics Newsletter*. 2013;pp.48–223.
18. Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, *et al.* Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal*. 1992;11:251–63.
19. McGonigle B, Bouhidel K, Irish VF. Nuclear localization of the *Arabidopsis APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes & Development*. 1996;10:1812–21.
20. Kang H-G, Jang S, Chung J-E, Cho Y-G, An G. Characterization of two rice MADS box genes that control flowering time. *Molecules & Cells*. 1997;7.
21. Pelucchi N, Fornara F, Favalli C, Masiero S, Lago C, Pè EM, *et al.* Comparative analysis of rice MADS-box genes expressed during flower development. *Sex Plant Reprod*. 2002;15:113–22.
22. Nonomura K-I, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, *et al.* A germ cell-specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell*. 2007;19:2583–94.
23. Nonomura K-I, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, *et al.* The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*. 2003;15:1728–39.
24. Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, *et al.* Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci.* 2012;109:1560–5.
25. Bartlett JG, Alves SC, Smedley M, Snape JW, Harwood WA. High-throughput *Agrobacterium*-mediated barley transformation. *Plant Methods*. 2008;4:1–12.

26. **Kovalchuk N.** High-throughput analysis pipeline for achieving simple low-copy wheat and barley transgenics. Fleury D, Whitford R, editors. *Crop Breeding: Methods and Protocols*. Springer New York; 2014. pp. 239–52.
27. **Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al.** JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research*. 2017;46:D260–6.
28. **Yoo S-D, Cho Y-H, Sheen J.** *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols*. 2007;2:1565–72.
29. **Bai Y, Han N, Wu J, Yang Y, Wang J, Zhu M, et al.** A transient gene expression system using barley protoplasts to evaluate microRNAs for post-transcriptional regulation of their target genes. *Plant Cell Tiss Organ Cult*. 2014;119:211–9.
30. **Citovsky V, Lee L-Y, Vyas S, Glick E, Chen M-H, Vainstein A, et al.** Subcellular localization of interacting proteins by bimolecular fluorescence complementation *in planta*. *Journal of Molecular Biology*. 2006;362:1120–31.
31. **Hellens RP, Allan AC, Friel EN, Bolitho K, Grafton K, Templeton MD, et al.** Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods*. 2005;1:13.
32. **Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J.** qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*. 2007;8:R19–14.

Chapter 5



Statement of Authorship

Title of Paper	Mapping of the <i>mov5</i> locus in barley
Publication Status	Unpublished and unsubmitted work written in manuscript style.

By signing the Statement of Authorship, each author certifies that:

- i. each author's contribution to the manuscript is accurate; and
- ii. permission is granted to include the manuscript in the thesis

Principal Author

Name of Principal Author	Caterina Selva		
Contribution to the Paper	Designed and performed the experiments. Analysed and interpreted the results. Wrote the manuscript.		
Percentage of Contribution	90 %		
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	20/09/2019

Co-author Contributions

Name of Co-author	Dr. Neil Shirley		
Contribution to the paper	helped in SNP discovery for mapping, performed the qRT-PCR experiment and analysed the qRT-PCR data. I hereby certify that the Statement of Authorship is accurate.		

(Handwritten mark)

Contribution to the paper	supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of		
Signature		Date	20/09/2019

Name of Co-author	Dr. Ute Baumann		
Contribution to the paper	the manuscript. I hereby certify that the Statement of authorship is accurate.		

Contribution to the paper	supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of authorship is accurate.		
Signature		Date	20/09/2019

Mapping of the *mov5* locus in barley

C. Selva, N. Shirley, R. Whitford, U. Baumann, M. R. Tucker

School of Agriculture Food and Wine, University of Adelaide, Waite Campus, Urrbrae 5064, South Australia, Australia

Abstract

This is the first report mapping the multiovary *mov5* locus in barley, which was located to the long arm of chromosome 2H. Multiple genes were identified within the mapped critical interval including *HvLFY*, the barley orthologue of *LEAFY* in *Arabidopsis thaliana*. The role of *HvLFY* in barley flower development was investigated. Based on sequence polymorphisms found in *HvLFY* in *mov5.o* plants, as well as transcriptomic analysis and the known role of LFY transcription factors in other species, we propose *HvLFY* as the most likely candidate for the *mov5.o* phenotype.

Introduction

Among the *multiovary* (*mov*) mutants *mov1*, *mov2* and *mov5*, the most unusual is *mov5*. As for other *mov* mutants, the phenotype presents itself as abnormal flower development. However, in the case of *mov5*, inflorescences show characteristics that closely resemble the sex reversal mutants initially observed in East Anglia and Britain as reported by Gregory and Purvis in 1947 [1]. Since 1947, no further studies were undertaken to establish the genetic

basis for this kind of abnormal flower development, and no studies have reported the genomic location of *mov5*.

In the present study we characterize the *mov5.o* allele of the *mov5* locus, created by fast neutron irradiation in the barley cultivar Morex [2]. We provide a detailed phenotypic description of *mov5.o*, which shows varying degrees of phenotypic expressivity between florets of the same spike, between spikes belonging to the same plant and across individual mutant plants. By assaying transcript abundance, we investigate how the balance of the ABC model is shifted in *mov5.o* inflorescences. Most importantly, we map the *mov5.o* allele to a 3.8 Mb region on the telomeric end of chromosome arm 2HL, which contains 115 annotated genes. One of these genes encodes a homologue of the *LEAFY* transcription factor from *Arabidopsis thaliana*, and the likely orthologue of *OsAPO2/RFL* from rice. Mutation of *OsAPO2/RFL* also leads to pleiotropic defects in rice inflorescence development, including formation of multiple ectopic carpels. Sequencing of *HvLFY* in *mov5.o* identified a single nucleotide polymorphism compared to *HvLFY* in cv. Morex, which leads to an amino acid change in a highly conserved residue located near the interface for HvLFY dimerization. In the present study, we discuss this finding in the context of *mov5.o* and present a testable model to explain the interaction between HvLFY and ABC-class genes.

Results

Floral organization is disrupted in *mov5.o* florets

The *mov5.o* phenotype is most striking when comparing florets from mutant and wild-type plants. In wild-type barley florets, each floral organ occupies a specific position (whorl) and displays a defined identity. The most external organs (whorl 1) are a pair of glumes called the palea and lemma which function to protect the internal reproductive organs. Palea and

lemma are followed by a pair of lodicules (whorl 2) that contribute to floret opening. The lodicules are positioned at the base of three stamens (whorl 3), which surround a central carpel (whorl 4) containing a single ovule (**Figure 1**).

In contrast to wild type, *mov5.o* inflorescences develop abnormal flowers that show different levels of phenotypic severity (**Figure 1**). Within the same spike, some *mov5.o* florets show severely disrupted floral organ identity while a phenotype could not be observed in other florets which appeared undistinguishable from wild type. Overall, the lemma and palea seem to remain unaffected in *mov5.o* as these organs are present and normal in all florets examined. However, the total number of lodicules in each floret varied from zero to two (**Table 1**). In some cases, partial conversion of lodicules into bract-like organs was observed and in extreme cases lodicules showed both bract-like and stamen-like characteristics and/or stigmatic hairs (**Figure 1D**). Similarly, the number of stamens varies, ranging from zero to three with occasional stamen-carpel mosaic organs (**Table 1; Figure 1F**). When present, stamens in *mov5.o* florets appear smaller and paler compared to wild-type stamens (**Figure 1F**), however pollen is still produced albeit in lower quantities. In addition to a partial reduction in stamen number, ectopic carpels were also formed in whorl 3. Visually, the majority of these ectopic carpels appear to be degenerate and bear no ovules (**Figure 1H**). Although additional carpels are occasionally produced in whorl 4, the prevalent structure in the centre of *mov5.o* florets is a single carpel-like organ with three, instead of two, feathery stigmas resulting in a triangular symmetry. If fertilized, the triangular symmetry is maintained in the developing seed (**Figure 1I-J**), which still supports germination of a single viable embryo into a mature plant. Phenotypic variation was also identified between *mov5.o* florets growing on the same spike, between spikes belonging to the same *mov5.o* plant and between *mov5.o* plants. Interestingly, it was also observed that in most cases the *mov5.o*

phenotype was associated with a distinctly twisted and curled spike peduncle (**Appendix C, Supplementary Figure S1**).

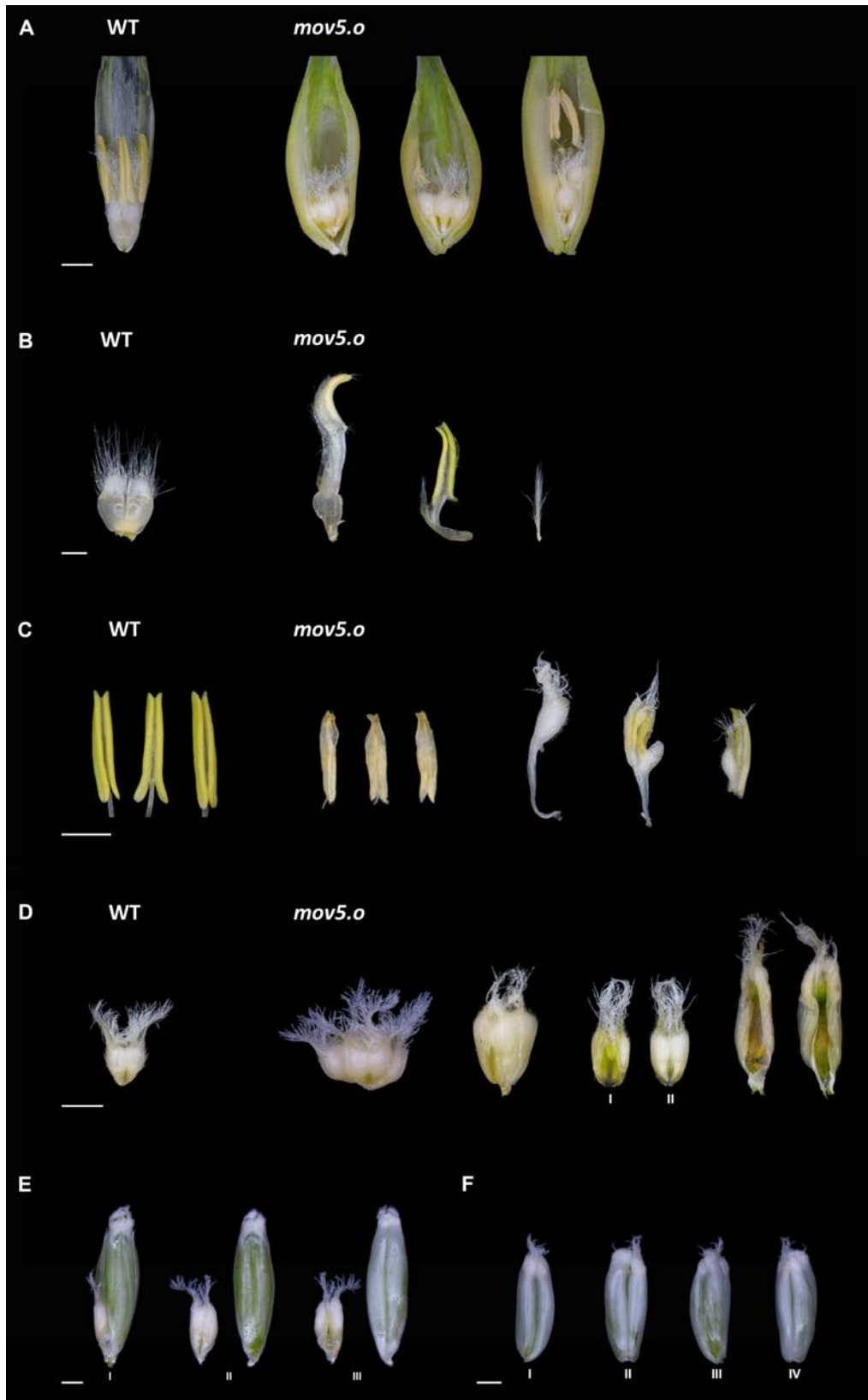


Figure 1. (A) Exposed wild-type (WT) and *mov5.o* florets. *mov5.o* florets can have a varying number of stamens and carpel-like structures. (B) Lodicules in wild type (WT) and *mov5.o*. Lodicules in *mov5.o* can show partial conversion into anthers, bracts and/or bear stigmatic hairs. (C) Stamens in wild type (WT) and possible variations in *mov5.o*. When present, stamens in *mov5.o* appear smaller and paler compared to wild type. Chimaeric organs having characteristics of both stamens and carpels are also present. (D) Carpels in wild-type (WT) and carpel-like complexes in *mov5.o* which can be composed of both normal carpels and “empty” carpels. (E-F) Seeds developing in *mov5.o* florets showing abnormal symmetry. Roman numerals indicate different sides of the same floral organ. Scale bars: 1000 µm.

Table 1. Frequency of floral organs in *mov5.o* florets from different spikes. WT indicates wild-type barley floret.

Floret	Lodicule	Stamen	Single carpel	Fused carpels	Comments
WT	2	3	1	0	
1	2	1	2	0	One lodicule/anther/bract partial conversion
2	1	2	2	1	One carpel developing into a single seed
3	0	3	0	1	One carpel/bract partial conversion
4	0	3	0	1	
5	0	2	1	1	
6	0	2	1	0	Single carpel appears "empty", one developing seed
7	0	1	1	2	
8	0	1	1	1	
9	0	1	1	1	One anther/carpel partial conversion
10	0	1	0	1	Fused carpels developing into a single seed
11	0	1	0	1	Fused carpels developing into a single seed
12	0	1	2	1	Single carpels appear "empty", Fused carpels developing into a single seed
13	0	1	3	0	Single carpels appear "empty", one developing seed
14	0	0	3	0	
15	0	0	1	1	
16	0	0	1	1	Two anther/carpel partial conversions

Histological clearing of *mov5.o* and wild-type mature carpels demonstrated that the female gametophyte can still be occasionally produced in either the additional or central carpel of *mov5.o* florets. Cases were also recorded in which multiple ovaries within a single *mov5.o* carpel-structure concomitantly developed fully differentiated embryo sacs (**Figure 2**).

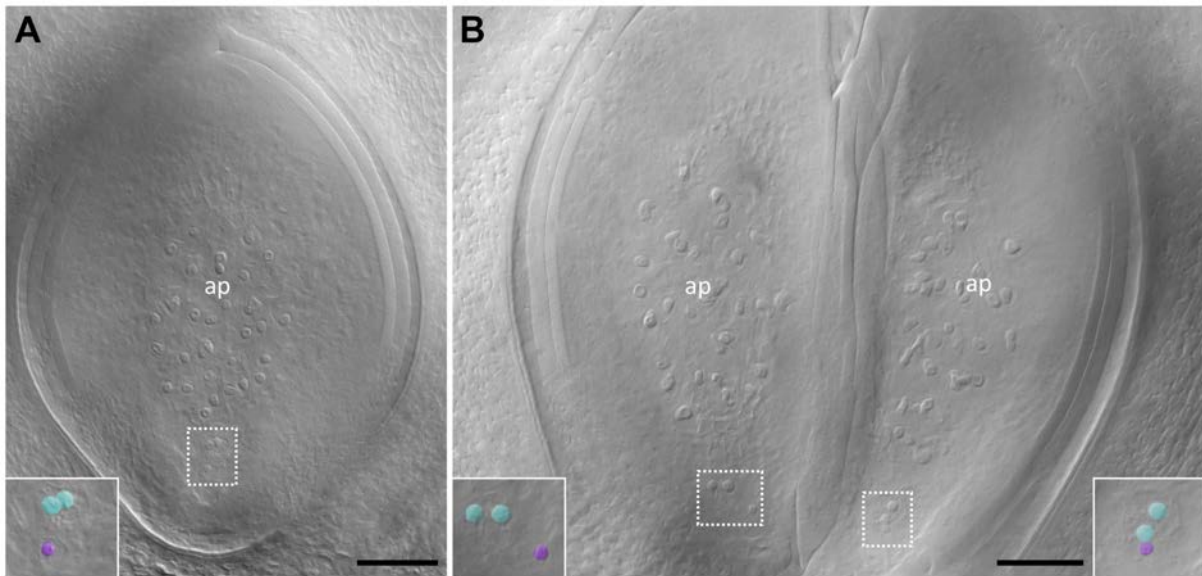
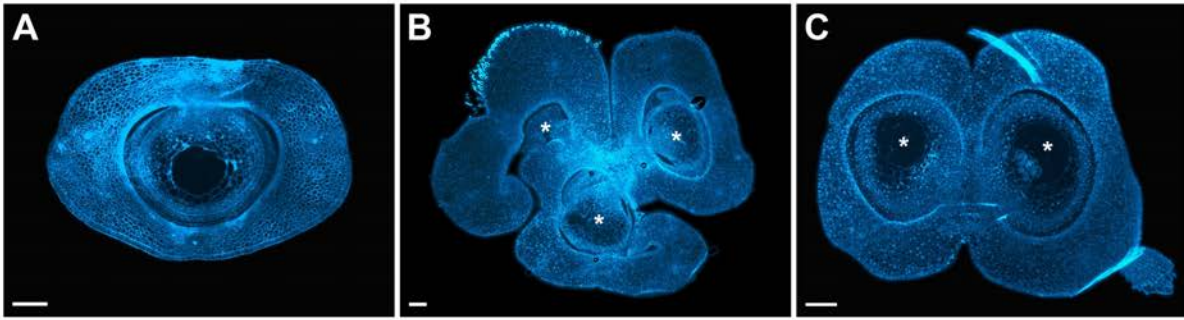


Figure 2. Cleared (A) wild-type and (B) *mov5.o* mature carpels. Inset focuses on the egg cell (purple) and polar nuclei (blue). Antipodal cells (ap) are indicated. Scale bars: 100 μ m.

However, as noted above, the majority of the ectopic carpels in *mov5.o* florets appear to degenerate prior to maturity or, if maturity is reached, contain an ovule lacking a female gametophyte. To further address whether ovule tissue identity and gametophyte development progress correctly in ectopic *mov5.o* carpels, immunolabelling was performed using antibodies that recognise specific cell wall epitopes including (1,3;1,4)- β -D-glucan and de-methylesterified pectin, which were previously shown to accumulate in defined ovule tissues (L. Wilkinson, M Tucker, unpublished). Thin transverse serial sections of multiovary structures provided an overview of the central and ectopic carpels and their contents (**Figure 3A-C**). In ovules from wild-type plants, (1,3;1,4)- β -D-glucan was detected in the inner and outer integuments, as well as in the nucellus cells adjacent to the embryo sac (**Figure 3D**). De-methylesterified pectin was mainly distributed in the nucellus cells surrounding the embryo sac, in a pattern that partially overlaps with (1,3;1,4)- β -D-glucan labelling (**Figure 3D**). As expected, considerable variation in ovule development was seen among *mov5.o* carpels. Mature *mov5.o* ovules containing a fully formed embryo sac showed a similar

pattern of (1,3;1,4)- β -D-glucan and de-methylesterified pectin accumulation compared to wild type (**Figure 3E**). However, distribution of both polysaccharides in the nucellus appeared less confined to cells surrounding the embryo sac, compared to wild type. Other carpels contained an ovule that appeared to have produced both integuments and nucellus but no gametophyte. These ovules still accumulated (1,3;1,4)- β -D-glucan and de-methylesterified pectin in the nucellus cells and (1,3;1,4)- β -D-glucan in the integuments (**Figure 3F**), suggesting that gametophyte development is not required *per se* for the accumulation or modification of these specific epitopes in the nucellus. In extreme cases, additional *mov5.o* carpels contained ovules comprised of over-proliferating integument tissues, which showed labelling of (1,3;1,4)- β -D-glucan but not of de-methylesterified pectin (**Figure 3G**). Although these results are indicative, histological clearing and immunolabelling should be repeated on a larger sample size when considering the high phenotypic variability of *mov5.o* carpels.



(1,3;1,4)-beta-D-glucan

De-methylesterified pectin

Merged

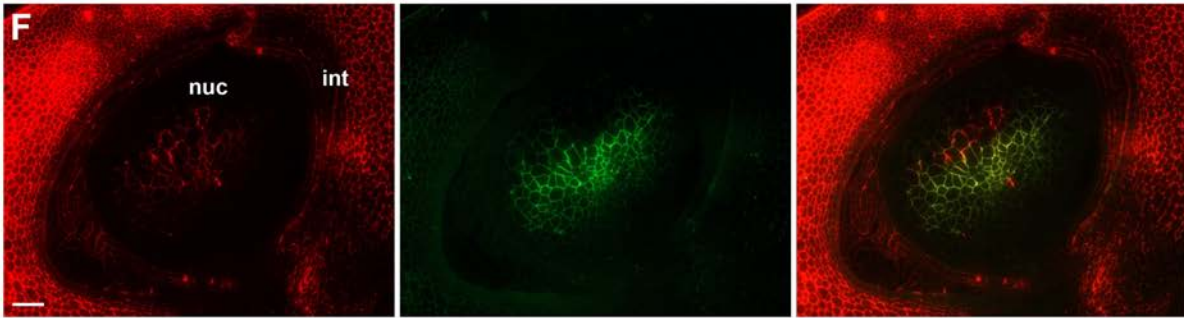
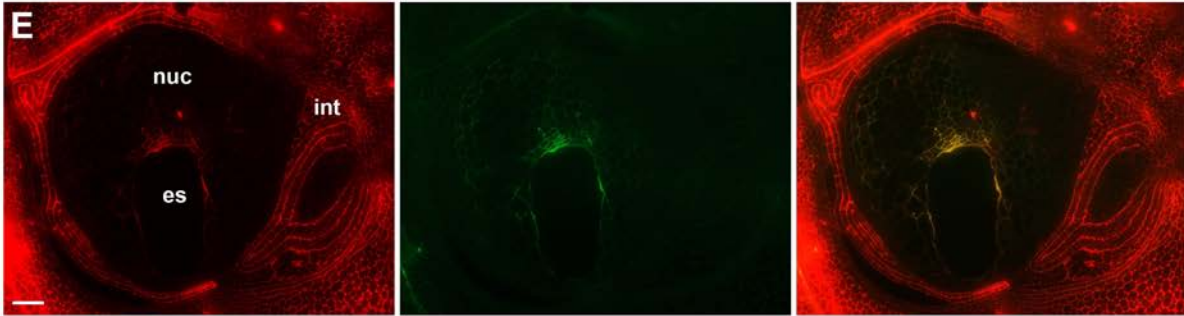
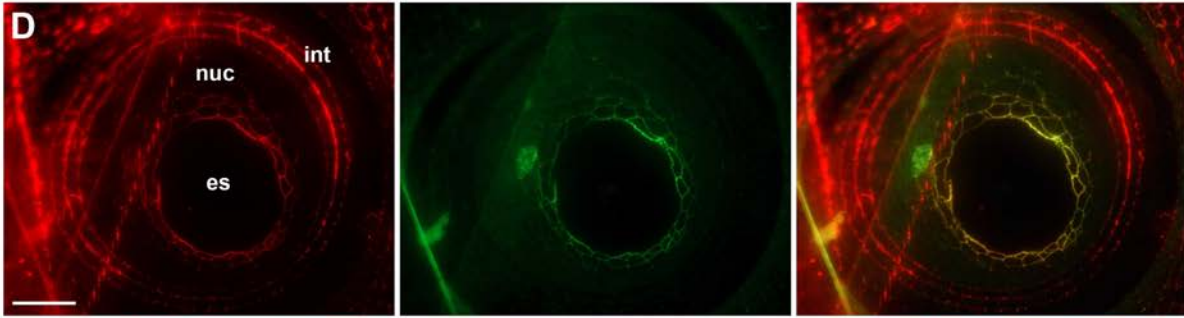


Figure 3. Transverse section of (A) wild-type and (B-C) *mov5.o* mature carpels showing autofluorescence. Ovule-like structures in *mov5.o* are indicated with asterisks. Scale bars: 100 μm . Immunolabelling of (D) wild-type and (E-G) *mov5.o* ovule structures showing distribution pattern of (1,3);(1,4)- β -D-glucan (red) and de-methylesterified pectin (green) in the ovule integuments (int), nucellus (nuc) and embryo sac (es). Scale bars: 50 μm .

The *mov5* locus is located on the long arm of chromosome 2H

To date, the genomic position of *mov5* remains unknown. To map this locus, a bi-parental F_2 mapping population derived from a cross between *mov5.o* (cv. Morex) and Steptoe was developed. To verify successful parental crosses, heterozygosity of eight F_1 plants was confirmed by KASPTM marker analysis across six previously known SNPs located on chromosome 3H, at positions: chr3H_1006543, chr3H_7767159, chr3H_8787424, chr3H_11702941, chr3H_14011512 and chr3H_28805649 according to the Morex reference assembly Hv_IBSC_PGSEB_v2. Phenotypically, flowers from all eight F_1 individuals appeared similar to wild type. In total, 88 F_2 plants from each of four confirmed F_1 lines were grown and used for mapping, totalling 352 F_2 individuals. Surprisingly, 13 (14.8 %) and 16 (18.2 %) F_2 individuals derived from two of the F_1 lines (V5317_S3-5 and V5317_S3-6) exhibited stunted growth (**Appendix C, Supplementary Figure S2**). For these plants, growth either arrested shortly after cotyledon expansion, or seedlings developed only two or three leaves that were particularly short and spindly. In all such cases, development progressed only for a one to two weeks before the seedlings died. Given these plants could not be grown to flowering for inflorescence phenotyping they were discarded from further analyses, leaving 323 F_2 individuals for mapping. Each remaining F_2 plant was phenotyped at heading or flowering stage by manually opening florets and visually inspecting 6 to 22 individual florets derived from 1 to 3 spikes. Phenotyping showed that the *mov5.o* multiovary trait segregated as a single Mendelian recessive locus (3:1) (**Appendix C, Supplementary Table S1**).

Due to the lack of knowledge regarding the location of *mov5*, KASP™ markers were designed to be evenly distributed across all barley chromosomes. To obtain a rough estimate of *mov5* chromosome locality, all markers were used to genotype a subset of 21 F₂ individuals, 11 of which exhibited the *mov5.o* phenotype. This genome-wide screening revealed association of the *mov5.o* phenotype with KASP™ marker chr2H_659264767 on chromosome 2H. To map the *mov5* locus at higher resolution, 22 additional KASP™ markers were designed surrounding chr2H_659264767 and were subsequently used to genotype all 323 F₂ individuals (**Figure 4**). This allowed the identification of a critical interval spanning approximately 3.8 Mb at the telomeric end of the long arm of chromosome 2H, between markers chr2H_695884392 and chr2H_699725902. It is worth noting that within this interval, marker chr2H_697042015 completely co-segregates with the *mov5.o* phenotype. However, partly due to the repetitiveness of the sequence in this region, no additional SNPs could be identified to further reduce the size of the critical interval.

A. *mov5* mapping

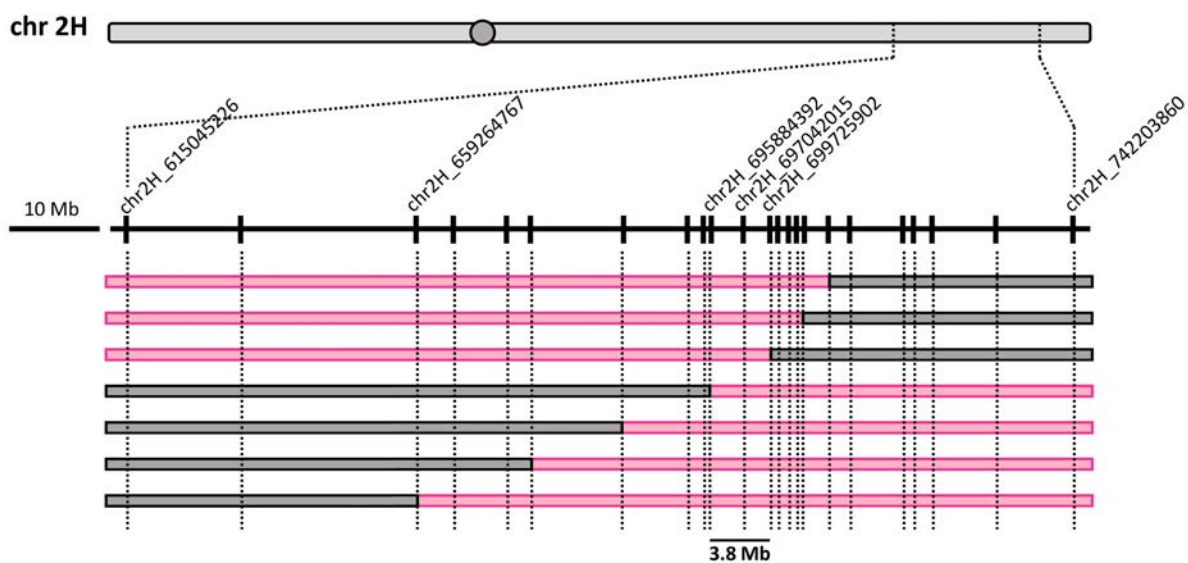


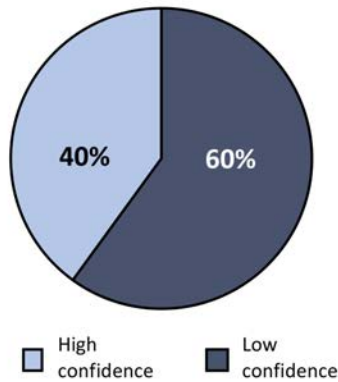
Figure 4. Mapping of the *mov5* locus in a *mov5.o* x Steptoe bi-parental population. *mov5* was mapped to a ~ 3.8 Mb interval between markers chr2H_695884392 and chr2H_699725902 based on 323 F₂ segregants. Marker chr2H_697042015 within *HvLFY* (*HORVU2Hr1G102590*) completely co-segregates with the *mov5.o* phenotype. Marker order is based on the genetic map. Examples of mapping in F₂ segregants exhibiting *mov5.o* phenotype are shown; pink: *mov5.o* allele; grey: Steptoe allele.

The *mov5* locus comprises multiple genes including the transcription factor *HvLFY*

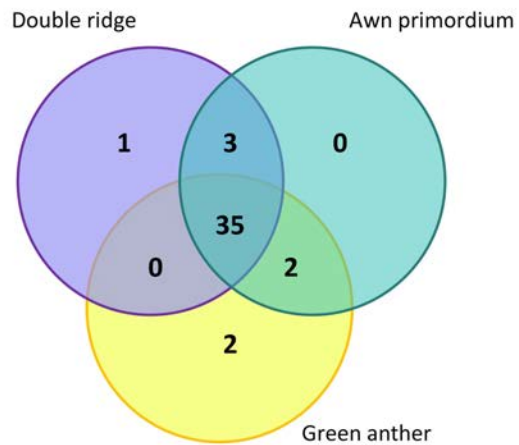
Based on comparisons between genetic and physical maps from cv. Morex, the region of interest on chromosome 2H is predicted to contain 115 annotated genes (**Appendix C, Supplementary Table S2**) of which 69 (60 %) are annotated as high confidence (HC) gene sequences (**Figure 5A**). Towards identifying the *mov5.o* causative gene sequence(s), the expression level of the 69 HC genes was investigated using published transcriptome data [3] from three developmental stages of wild-type inflorescences, corresponding to double ridge (Waddington stage W2.0), awn primordium (W3.5) and green anther (W8.0 – 8.5) stage, as well as from 2-week old seedlings (Zadok stage Z12) [4,5]. Of the 69 HC genes, only 39 (56.5 %) are expressed at double ridge with a FPKM \geq 1. Likewise, 40 (58 %) and 39 (56.5 %) genes are expressed with a FPKM \geq 1 at the awn primordium and green anther stages, respectively. Interestingly, of these expressed HC genes, 38 are in common between double ridge and awn primordium samples (**Figure 5B**). These 38 genes were selected for further consideration. Genes specific for the green anther stage were not analysed further as it captures a timepoint that was considered too late in flower development to underlie the *mov5.o* phenotype. Of the 38 HC gene sequences expressed at both double ridge and awn primordium stages, *HORVU2Hr1G102590* was identified as a possible candidate to explain the *mov5.o* phenotype based on expression pattern and annotation. Overall, *HORVU2Hr1G102590* expression is highest at the double ridge stage (ranking as the second highest expressed gene 117.69

FPKM) and noticeably decreases at the awn primordium stage (24.84 FPKM) and green anther stage (0.21 FPKM), as well as having very low expression in vegetative tissue (0.11 FPKM) (**Figure 5C**).

A. Gene annotation in *mov5* interval



B. Gene distribution among inflorescence stages



C. *HORVU2Hr1G102590* expression pattern

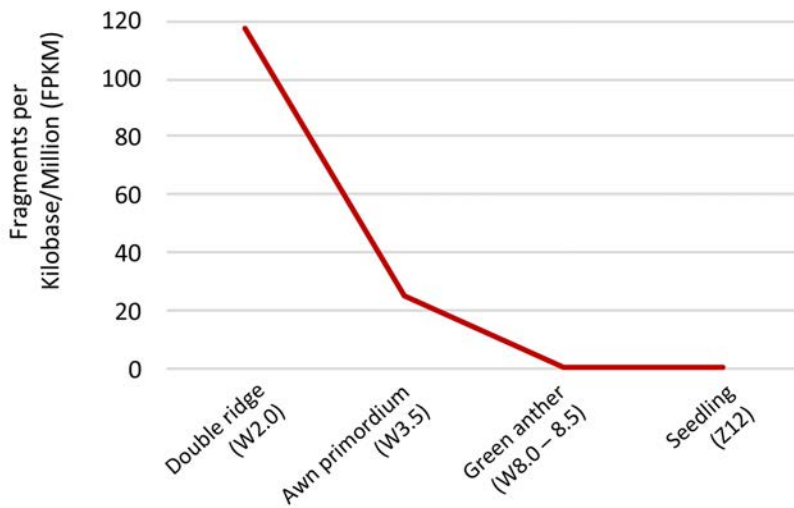


Figure 5. (A) The mapped interval for *mov5* contains 115 genes, of which only 69 (40 %) are annotated with high confidence, based on Morex reference assembly *Hv_IBSC_PGSB_v2*. (B) Most high confidence genes with FPKM ≥ 1 are expressed in all stages: double ridge (W2.0), awn primordium (W3.5) and green anther (W8.0 - 8.5). (C) *HORVU2Hr1G102590* (*HvLFY*) expression pattern from Liu *et al.* (2019) [3].

HORVU2Hr1G102590 encodes a protein with very high sequence identity (84.3%) to rice ABERRANT PANICLE ORGANIZATION 2/RFL (APO2), belonging to the FLORICAULA/LEAFY (FLO/LFY) plant-specific transcription factor family [6]. Henceforth, gene *HORVU2Hr1G102590* will be referred to as *HvLFY*. Based on the reference Morex assembly Hv_IBSC_PGSA_v2 the *HvLFY* gene is divided in three exons and a 5'UTR (**Figure 6**). The exon-intron structure was found to be conserved with previously characterised *FLO/LFY* genes from *Antirrhinum majus*, *Arabidopsis thaliana*, maize (*Zea mays*) and rice. Comparison of the FLO/LFY protein sequence among these species shows 64.1 % residue identity, with particularly high sequence conservation towards the C-terminal domain, including the final part of the second exon and the third exon (**Appendix C, Supplementary Figure S3**).

Since the *mov5.o* mutant was created by fast neutron irradiation [2], which typically causes genomic deletions and disruptions of varying sizes, the presence of *HvLFY* was tested by PCR. Multiple replicates indicated that gene *HvLFY* was still present in *mov5.o* individuals. The presence of another 14 genes located within the *mov5* region and surrounding *HvLFY*, which also show high levels of expression at double ridge/awn primordium stages, was also tested (**Appendix C, Supplementary Table S2**). Similar to *HvLFY*, all tested genes were found to be present in *mov5.o* plants.

To investigate whether *mov5.o* plants contain any sequence polymorphisms that could explain the mutant phenotype, *HvLFY* was amplified from *mov5.o*, Morex and Steptoe plants and Sanger sequenced. Sequencing covered the entire gene sequence and was extended to 500 bp upstream of the translation start site, inclusive of the 5'UTR, and 150 bp downstream of the stop codon. No sequence variations were identified within the upstream, 5'UTR and

downstream regions. One SNP [C/T] within the third exon was detected between the *mov5.o HvLFY* gene and both Steptoe and Morex sequences (**Figure 6**). At the protein level, this SNP causes a Pro → Leu non-synonymous substitution at position 364 (P364L) in HvLFY. The same SNP was used to design KASP™ marker chr2H_697042015 which was observed to completely co-segregate with the *mov5.o* phenotype (**Figure 4**).

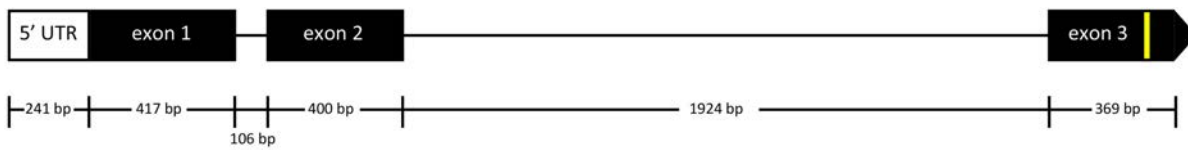


Figure 6. Gene structure for *HvLFY* (*HORVU2Hr1G102590*); length in base pairs of 5' Untranslated Region (UTR), exons (black) and introns (solid line) is indicated. Yellow line shows the position of the [C/T] SNP found in *mov5.o* plants.

Interaction of *Mov5* with the ABC genes

To investigate whether *mov5.o* influences the expression pattern of floral homeotic genes, transcript abundance of the major genes involved in flower development was tested by qRT-PCR. Wild-type and *mov5.o* developing inflorescences were collected at stages W2.0, W3.5, W4.5 and W6.0 to include organogenesis of both stamens, carpels and ovules. Interestingly, expression of *HvLFY* was not significantly altered in *mov5.o* samples compared to wild type (**Figure 7**). In both genotypes, expression was highest at W2.0 (double ridge) and decreased as development progressed. On the other hand, levels of B-class genes tended to increase in abundance with development. In *mov5.o*, all B-class genes, *HvMADS2*, *HvMADS4* and *HvMADS16* showed a reduction in transcript abundance which was most significant for stage W4.5 (carpel primordium) and most marked for *HvMADS16* (**Figure 7**). A similar trend was seen for both E-class genes tested (*HvMADS7* and *HvMADS8*). Also in the case of these two E-class genes, transcript abundance increased with development, with lower abundance in

mov5.o compared to wild type, especially at W4.5 (**Figure 7**). Carpel-specific genes (*HvMADS58* and *HvDL*) were also assayed. *HvMADS58* showed an overall decrease in transcript abundance in *mov5.o* inflorescences, while *HvDL* was significantly reduced only at the timepoint of stamen primordia initiation (W3.5). Transcript levels did not greatly differ from wild type for all other timepoints (**Figure 7; Appendix C, Supplementary Figure S4**). The ovule-specific *HvMADS13* and *HvMADS3* genes appeared to follow the same expression dynamics in *mov5.o* as in wild type. However, while *HvMADS3* levels remained unaffected, *HvMADS13* appeared to be slightly but significantly increased in *mov5.o* at stage W2.0 and significantly reduced at stage W4.5 (**Appendix C, Supplementary Figure S4**). Transcript abundance for A-class genes (*HvMADS14* and *HvMADS15*), the meristem-specific *HvOSH1* (*HOMEODOMAIN 1*) and germ-line associated *HvMSP1* (*MULTIPLE SPOROCTE*) genes remained unaffected in *mov5.o* samples (**Appendix C, Supplementary Figure S4**). The other germ-line gene tested, *HvMEL1* (*MEIOSIS ARRESTED AT LEPTOTENE1*) showed reduced abundance in *mov5.o* from stage W3.5 onwards (**Appendix C, Supplementary Figure S4**), however this reduction was not statistically significant.

Wild-type — *mov5.o* —

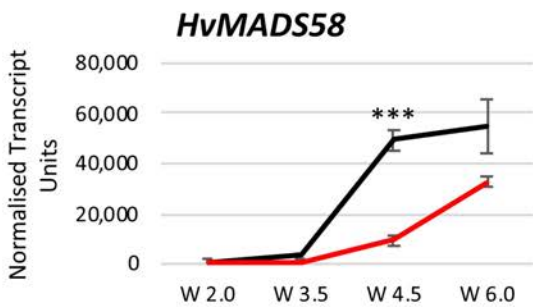
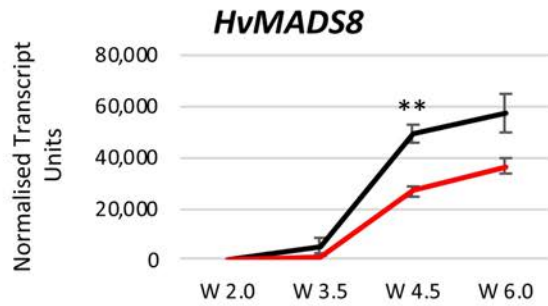
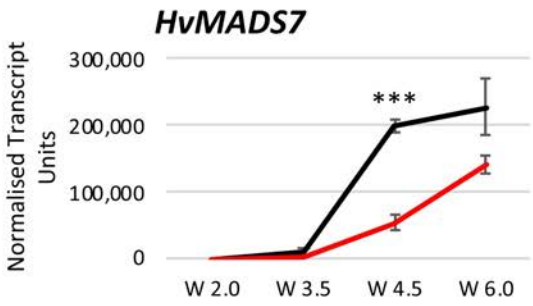
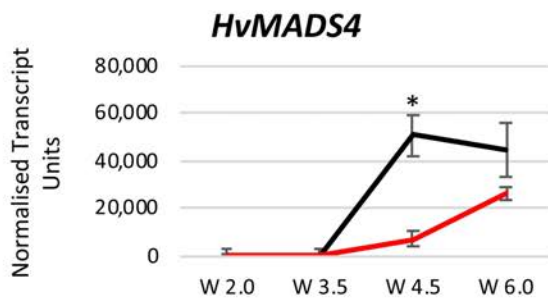
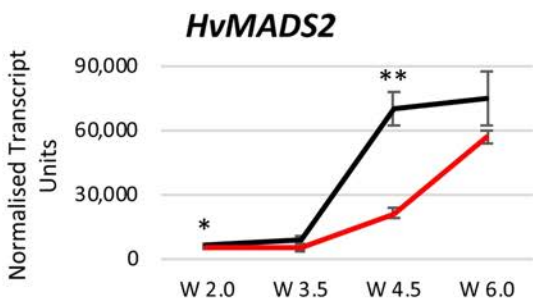
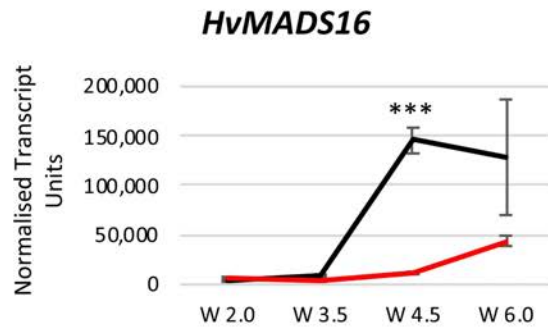
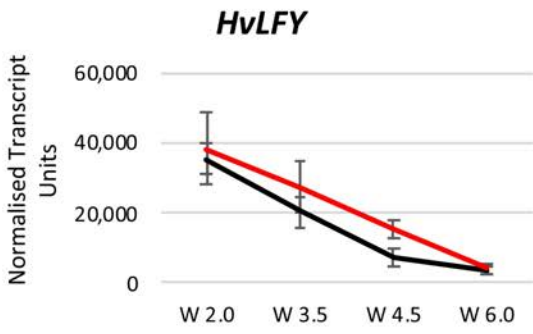


Figure 7. Transcript abundance assessed by qRT-PCR in wild-type (black) and *mov5.o* (red) developing inflorescences at stages W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development) for: *HvLFY* (*HORVU2Hr1G102590*), B-class genes *HvMADS16* (*HORVU7Hr1G091210*), *HvMADS2* (*HORVU3Hr1G091000*), *HvMADS4* (*HORVU1Hr1G063620*); E-class genes *HvMADS7* (*HORVU5Hr1G076400*), *HvMADS8* (*HORVU7Hr1G054220*) and C-class gene *HvMADS58* (*HORVU1Hr1G029220*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**) and ≤ 0.001 (***) are shown for differences between wild type and *mov5.o*. For each sample *n* = 3 independent biological replicates.

Discussion

mov5.o is involved in correct floral development

Despite showing normal vegetative growth, *mov5.o* plants develop dysfunctional florets. In *mov5.o* florets, the inconsistent loss of specific floral organs (stamens and lodicules), the production of mosaic floral organs due to incomplete conversion of lodicules to bracts, lodicules to stamens and stamens to carpels, as well as the development of supernumerary carpel-like structures (**Figure 1**) indicate that floral development is severely compromised in this mutant. Based on phenotypic evidence, we speculate that *mov5.o* affects floral organs in whorl 2 (lodicules), whorl 3 (stamens) and whorl 4 (carpel). However, it is worth noting that the number of floral organs in *mov5.o* do not usually equate to the expected number based on a simple homeotic conversion, but rather *mov5.o* florets show an increase in organ number. The phenotype observed in *mov5.o* individuals is strikingly similar to early descriptions of barley multiovary mutants by Gregory and Purvis (1947) [1]. In both cases, abnormal florets do not grow along the entire spike and there is phenotypic variability in the partial conversion of stamens to carpels, ranging from weak defects whereby the stamens exhibit a few apical stigmatic hairs to marked transitions in which stamens resemble hollow carpels lacking an ovule. For *mov5.o* and earlier described multiovary mutants, however, no allelism tests or information regarding possible causative genes is thus far available. Despite such a severe floral phenotype, *mov5.o* florets occasionally manage to develop viable seeds upon self- or cross-pollination. This observation indicates that although the majority of *mov5.o* additional carpels appear hollow, the single or fused carpel structure that forms in whorl 4 still harbours a functional embryo sac, although the symmetry of the developing seed remains abnormal (**Figure 1**). Results from histological clearing and immunolabelling of mature carpels confirm that *mov5.o* individuals are able to occasionally produce a variable

number of functional embryo sacs and that cell identity within the *mov5.o* ovules is maintained (**Figure 2**). Furthermore, successful reciprocal crosses with wild type cv. Morex indicate that limited, but viable pollen is still produced from *mov5.o* stamens (personal observation).

The *mov5* locus is located on chromosome 2H to a region containing a LEAFY-like transcription factor

mov5.o florets also closely resemble the floret disruptions reported for the rice mutant *aberrant panicle organization 2/rfl* (*apo2/rfl*). In rice, *apo2/rfl* flowers show partial loss of floral determinacy, as well as conversion of lodicules into glumes, lodicule-stamen mosaic organs, reduced number of stamens and ectopic carpel formation [6]. Similarly in maize, mutants in the corresponding APO2/RFL genes, termed *Zea FLO/LFY 1* (*ZFL1*) and *Zea FLO/LFY 2* (*ZFL2*), exhibit disruption in floral organ identity and patterning. In the female florets of the maize double mutant *zf1 zf2* stamen primordia fail to develop, accompanied by a concomitant proliferation of carpel-like organs and organs of unclear identity [7]. On the other hand, male flowers of the double mutant either lack stamens or produce twisted stamens with a reduced number and size of locules [7]. Rice *APO2/RFL* and maize *ZFL1* and *ZFL2* encode orthologues of the *FLORICAULA* (*FLO*) and *LEAFY* (*LFY*) genes in *Antirrhinum* and *Arabidopsis*, respectively. The *FLO/LFY* genes belong to a plant-specific transcription factor family with prominent roles in reproductive transition, flower development and plant architecture in various species.

In this study, generation and analysis of a *mov5.o* (cv. Morex) x Steptoe F₂ mapping population positioned *mov5* to the telomeric end of chromosome 2HL spanning a critical region of approximately 3.8 Mb (**Figure 3**). Among the 115 annotated genes within the

identified interval, preferential expression at double ridge and awn primordium stage in wild-type inflorescences was used to filter for possible causative candidates underlying *mov5* (**Figure 4**). One such gene was *HORVU2Hr1G102590*, termed here *HvLFY* due to its high sequence identity with previously characterised FLO/LFY proteins. Based on the high protein sequence identity, the conserved exon-intron structure of *HvLFY* with other characterized *FLO/LFY* genes, the extensive documented evidence of a role of *FLO/LFY* genes in flower development, and the phenotypic similarity of *mov5.o* plants to described *flo/lfy* mutants, we speculate that *HvLFY* is a plausible candidate to explain the *mov5.o* phenotype. Consistent with this hypothesis, sequencing of *HvLFY* in *mov5.o* plants revealed a [C/T] SNP in the third exon which results in a P364L change in the amino acid sequence (**Figure 5**). KASP marker chr2H_697042015 designed using the identified SNP fully co-segregates with *mov5.o* phenotype (**Figure 3**). Furthermore, segregation of the *mov5.o* phenotype within the F₂ population is indicative of a single-locus recessive trait (**Appendix C, Supplementary Table S1**).

FLO/LFY proteins usually contain a weakly conserved N-terminal domain that mediates LFY oligo-dimerization and possibly influences DNA binding [8] and a highly conserved C-terminal region that forms the DNA-binding domain which recognizes pseudo-palindromic sequence elements in promoters of target genes [9]. Consistent with this, sequence comparison of FLO/LFY proteins from barley, rice, maize, *Antirrhinum majus* and *Arabidopsis thaliana* showed a higher degree of similarity in the C-terminal domain compared to the N-terminal domain (**Appendix C, Supplementary Figure S3**).

A study looking at the crystal structure of *Arabidopsis* LFY bound to target promoter elements determined that LFY binds DNA as a cooperative dimer. Binding occurs between a unique helix-turn-helix fold in the LFY C-terminal DNA-binding domain which forms base-specific

interactions with both the major and minor grooves of the DNA [9]. The helix-turn-helix fold is able to promote LFY dimerization by itself and was shown to be defined by two short β -strands (β 1 and β 2) followed by seven α -helices (α 6 - α 12) connected by short loops [9]. Pro364 is found in the C-terminal domain of HvAPO2, located proximally to α 12, which is involved in the interface between LFY dimers in *Arabidopsis* (**Appendix C, Supplementary Figure S3**) [9]. Despite the evolutionary rates of *LFY/FLO* genes in grass species being significantly higher than those in Angiosperms [10], Pro364 is conserved in all species considered in this study and more, including Angiosperms *Nymphaea odorata* and *Brownea grandiceps*, the gymnosperm *Welwitschia mirabilis*, the fern *Matteuccia struthiopteris* and the bryophyte *Physcomitrella patens* [9]. Conservation of Pro364 across the plant kingdom is indicative of the functional importance of this specific residue. Furthermore, there are numerous reported examples whereby other single amino acid substitutions in LFY are sufficient to affect DNA-binding affinity or interactions between amino acids and lead to mutant phenotypes *in planta* [9]. Consequently, we speculate that the identified SNP resulting in P364L could influence HvLFY functionality in *mov5.o* plants. Decreased or altered HvLFY function during barley inflorescence development could explain the inconsistent phenotype observed for *mov5.o* florets.

Although it has been reported, especially in rice, that the precise spatial-temporal expression of *FLO/LFY* genes can be complex [11,12], we do not believe that the mutant phenotype in *mov5.o* plants is due to mis-regulation of *HvLFY* expression. Indeed, transcript abundance of *HvLFY* generally does not significantly differ between mutant and wild-type developing inflorescences (**Figure 6**). This result is consistent with the hypothesis that alteration in HvLFY function in *mov5.o* occurs at the protein level, hence after transcription. In both *mov5.o* and

wild-type inflorescences, *HvLFY* is most abundant at early developmental stages, corresponding to approximately W2.0 (double ridge) and decreases as the differentiation and development of both male and female reproductive organs progresses (**Figure 7**). These results corroborate the transcriptomic data utilised, where *HvLFY* expression was highest at the double ridge stage and decreases with inflorescence development (**Figure 5**). Expression of *HvLFY*, which precedes and coincides with floral organogenesis, suggests that this gene acts quite early in flower development and is consistent with the floral defects of *mov5.o* florets.

Based on KASPTM genotyping, the stunted and lethal phenotype seen in low proportion (14.8 % and 18.2 %) only in some F₂ plants of the mapping population was not linked to the *mov5* locus or caused by altered *HvLFY* function. We suspect it might be a result of a secondary background mutation caused during the original mutagenesis experiment. Apart from a twisted peduncle for *mov5.o* spikes (**Appendix C, Supplementary Figure S1**), no other pronounced pleiotropic effects were noticed in the vegetative phase or other growth aspects of *mov5.o* plants. This is in contrast to rice *apo2/rfl* mutants, which apart from aberrant floral organ identities and loss of floral meristem determinacy, also show additional pleiotropic phenotypes. In addition to its role in flower development, APO2/RFL in rice also appears to promote tillering, flowering, phyllotaxy and panicle branching [6,13]. Likewise in maize, ZFL1 and ZFL2 play a role in inflorescence architecture by promoting tassel branching and are also involved in regulating reproductive transition [7]. Given that spike architecture in barley differs from rice and maize, and that no other major defects were observed in *mov5.o* plants, we speculate that *HvLFY* in barley may have a distinct role relative to *FLO/LFY* genes of rice and maize. However, it is worth considering that the mild phenotype in *mov5.o* plants could

be attributed to a weak, but functional allele of *HvLFY* as opposed to a complete absence of *HvLFY* function, which could give rise to more severe phenotypes. To this end, future studies investigating the effects of absence of *HvLFY* will be pursued.

HvLFY shows partial functional conservation with other FLO/LFY genes

The function of FLO/LFY orthologues appears to have been modulated in several species. For example: in *Arabidopsis* LFY controls reproductive transition and floral organ identity [14-16], FLO in *Antirrhinum* confers floral meristem identity and is involved in correct phyllotaxy [17], ABERRANT LEAF AND FLOWER (ALF) in *Petunia hybrida* controls floral identity and branching pattern [18], PpLFY1 and PpLFY2 in the moss species *Physcomitrella patens* control the first zygotic cell division [19] and NICOTIANA FLO/LFY1 and 2 (NFL1 and NFL2) in tobacco (*Nicotiana tabacum*) are required for correct development and branching of both shoot apical and floral meristems [20]. The orthologue UNIFOLIATA (UNI) in pea (*Pisum sativum*) has a role in vegetative structures in promoting compound leaf development [21,22], an effect also seen for FALSIFLORA (FA) in tomato (*Solanum lycopersicum*) even if less pronounced [23]. Nonetheless, although divergent species have specified FLO/LFY function to confer additional developmental roles, all FLO/LFY proteins characterised so far retain a conserved function in floral development acted through the regulation of floral organ identity genes belonging to the ABC model. For example, in *Arabidopsis*, LFY induces expression of *AP1* (A-class), *AP3* (B-class) and *AG* (C-class) [24-28] genes. Another example can be found in rice, whereby APO2/RFL suppresses expression of A-class genes *OsMADS14*, *OsMADS15* and *OsMADS18*, as well as expression of the E-class gene *OsMADS34* [6].

In this study, we show that the expression of B-class genes (*HvMADS2*, *HvMADS4* and *HvMADS16*) and E-class genes (*HvMADS7* and *HvMADS8*) is affected in *mov5.o* inflorescences (**Figure 6**). These genes show decreased expression in *mov5.o* compared to wild-type inflorescences, an observation that is in accord with a role of these genes in stamen and lodicule organogenesis and differentiation, as documented in rice [29-33], and consistent with the barley mutant phenotype. Likewise, the reduced expression seen for *HvMEL1* (**Appendix C, Supplementary Figure S4**), the barley orthologue of rice *MEL1*, is in accord with the *mov5.o* phenotype as *MEL1* in rice is expressed in both the male and female archesporial cells and sporocytes [34]. The apparent lower expression level of *HvMEL1* may be indicative of fewer archesporial cells produced due to the reduced number of anthers developing in *mov5.o* florets.

Interestingly, expression of all tested genes involved in carpel (*HvMADS58* and *HvDL*) and ovule development (*HvMADS3* and *HvMADS13*) as well as meristem maintenance (*HvOSH1*) seem to be reduced or unaltered in *mov5.o* plants (**Figure 6, Appendix C, Supplementary Figure S4**). This result was unexpected as *mov5.o* florets produce supernumerary carpels and increased numbers of floral organs, which might be expected to manifest through increased C- and D- class gene function. It is also worth noting that for all genes assayed, the most significant changes in transcript abundance occurred at stage W4.5, corresponding approximately to carpel primordium initiation. These results highlight the need for additional studies to elucidate the precise effects of *mov5.o* and *HvLFY* on meristem development, flower development and, particularly, on the organs of the fourth (innermost) whorl. Unlike rice, transcript levels of A-class genes *HvMADS14* and *HvMADS15* do not undergo significant changes in *mov5.o* plants (**Appendix C, Supplementary Figure S4**). If *HvLFY* is the causative gene, this might be due to the nature of the allele. Considering *HvLFY* is still present and

expressed in *mov5.o* inflorescences, it is plausible that *mov5.o* represents a weak *HvLFY* allele, in which function is partially impaired due to the amino acid substitution. Additionally, unchanged levels of A-class genes, which are mostly involved in whorls 1 and 2 and in establishing the floral context [35], agree with the observation of no phenotypic defects in the palea and lemma of *mov5.o* florets. Finally, transcript abundance of *HvMSP1*, predicted from rice to be involved in regulating male and female sporogenesis [36] appears to remain unaffected in *mov5.o* plants despite having reduced anthers (**Appendix C, Supplementary Figure S4**). This result is similar to those previously obtained for *mov1* and *mov2.g*, in which *HvMSP1* levels were similarly unaffected and thus argues against a conserved function of this particular gene with rice. Further studies will be required to determine if this is indeed the correct *OsMSP1* orthologue.

In *Arabidopsis*, LFY requires the co-factor UNUSUAL FLORAL ORGANS (UFO), an F-box protein, to directly activate proper *AP3* expression for petal and stamen formation [37,38]. This regulatory mechanism between LFY and UFO appears to be widely conserved as orthologues have been identified in numerous plant species. In particular, a similar interaction for regulation of the B-class genes *DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*) has been proposed in *Antirrhinum*, between FLO and FIMBRIATA (*FIM*), the UFO ortholog [39,40]. In rice protoplasts, APO2/RFL has been shown to regulate and physically interact with ABERRANT PANICLE ORGANIZATION1 (*APO1*), the rice homolog of UFO [6]. Mutant analysis established that APO2 and APO1 jointly control floral meristem determinacy [6]. LFY and UFO orthologues have also been observed in petunia and pea [41,42]. Thus, we suggest that also in barley *HvLFY* might require the action of an F-box protein as transcriptional co-factor for direct regulation of B-class genes *HvMADS2*, *HvMADS4* and *HvMADS16*. In this context, the

barley gene *HORVU7Hr1G108970* is a likely F-box candidate based on sequence homology to rice *APO1*. However, further studies confirming interaction between HvLFY and *HORVU7Hr1G108970* and/or if LFY binding sites are present in promoters of B-class genes are needed to test our hypothesis.

The role of HvLFY in flower development

Based on the results presented in this study, we propose a model to explain the putative function of *HvLFY* in barley flower development. In wild-type barley, *HvLFY* is expressed very early in inflorescence development, even before the primordia of floral organs arise, and positively regulates the activity of E-class genes. E-class genes are typically expressed in all floral whorls and mediate the interactions between MADS-box proteins, thus at these early stages *HvLFY* plays a role in establishing the suitable molecular context for MADS-box transcription factors. As the floral meristem differentiates, *HvLFY* expression gradually decreases. However, *HvLFY* is still expressed at the initiation of stamen primordia, whereby it directly activates B-class genes, particularly *HvMADS16*, via interaction with the F-box protein *HORVU7Hr1G108970*. In floral whorls 2 and 3, the B-class genes interact with A- and E-class genes, or C- and E-class genes, to form the floral quartets necessary to specify lodicule and stamen differentiation, respectively.

In *mov5.o*, the identified P364L amino acid substitution alters *HvLFY* function. As *HvLFY* function is modified, but not abolished, the regulation of downstream targets is stochastic, thus compromising the establishment of the molecular context for MADS-box functionality. Based on the case-specific molecular context that is formed, the identity and abundance of specific MADS-box floral quartets can be favoured or hindered in any floral whorl, thus explaining the varying degree of phenotypic severity observed in *mov5.o* florets.

In conclusion, we have mapped the *mov5* locus to the long arm of chromosome 2H and have identified *HvLFY* as a plausible candidate. The *HvLFY* allele in *mov5.o* contains a single SNP causing a Pro → Leu non-synonymous substitution at a protein residue that is highly conserved between LFY orthologues across species from Angiosperms to bryophytes, and therefore possibly alters HvLFY function. Fully functional *Mov5* is required for correct expression of floral homeotic genes, especially B-class genes, which is consistent with the function of LFY from other species. Despite these correlative findings, further experimental evidence is required to confirm that *HvLFY* is *Mov5*. This might be achieved through transformation of a HvLFY rescue construct into the *mov5.o* background or by generation of *Hvlfy* alleles by CRSIRP/Cas9. At this point, we speculate that *HvLFY* is most likely *Mov5*, and that *HvLFY* function might have diverged slightly from other monocotyledons like rice and maize. We propose a model to explain the putative function of HvLFY in floral organogenesis, although further studies are needed to test the precise role of HvLFY in flower and vegetative development.

Materials and Methods

Plant material

Segregating seeds for the *mov5.o* allele (cv. Morex) mutated at the *mov5* locus were kindly provided by Professor A. Kleinhofs. Growing and phenotyping of plant material was performed as described in Chapter 3. Development of the *mov5.o* x Steptoe bi-parental mapping population, SNP discovery and KASP™ marker design was performed as mentioned in Chapter 4. Sequence of all KASP™ markers can be found in **Appendix C, Supplementary Table S3**.

Nucleic acid extraction and PCR

For all plant material, genomic DNA was extracted from 2-week old seedling as described in Chapter 4. PCR reactions were prepared, and PCR products visualized on 1 % agarose gel, as mentioned in Chapter 3. A list of primers used for PCR can be found at **Appendix C, Supplementary Table S4A**.

RNA extraction, cDNA synthesis and quantitative real-time PCR (qRT-PCR)

Wild-type and *mov5.o* inflorescences were manually dissected with fine-pointed tweezers at developmental stages W2.0, W3.5, W4.0, W6.0, which correspond roughly to 25, 30, 35 and 40 Days Post Germination (DPG) in the growing conditions used. Dissected inflorescences were immediately frozen in liquid nitrogen and stored at -80 °C until further processing. RNA extraction, cDNA synthesis and qRT-PCR were completed as described in Chapter 3. Sequence of primers used for qRT-PCR are listed in **Appendix C, Supplementary Table S4B**.

Histological clearing and immunolabelling with fluorophores

Sample collection and processing for histological clearing of whole mature carpels was performed as described by Wilkinson and Tucker (2017) [43]. For immunolabelling, mature carpels were fixed in FAA solution (50 % Ethanol 100 %, 5 % acetic acid, 10 % formaldehyde 37 %, one drop of Tween-20) overnight and transferred to ethanol 70 % until further processing. Samples were dehydrated in a 70 - 100 % ethanol series and embedded in LR white resin. Samples were then sectioned using a Leica Rotary Microtome RM2265 at 1.5 μm . Immunolabelling was performed as described by Aditya *et al.* (2015) [44], on two biological replicates. Images were taken with Zeiss M1 AxioImager equipped with AxioCam 506 mono and processed using the ZEN 2012 software.

Acknowledgments

This research was supported by the APA Scholarship from the Australian Government Research Training Program Scholarship and the ACPFG Supplementary Scholarship. We are grateful to Dr. Laura Wilkinson and Mia Lou for guidance with immunolabelling and Margaret Pallotta for critical revision of the manuscript.

References

1. **Gregory FJ, Purvis ON.** Abnormal flower development in barley involving sex reversal. *Nature*. 1947;160:221–2.
2. **Barley Genetics Newsletter.** 2013;pp. 48–223.
3. **Liu H, Li G, Yang X, Kuijter HNJ, Liang W, Zhang D.** Transcriptome profiling reveals phase-specific gene expression in the developing barley inflorescence. *The Crop Journal*. 2019;1–16.
4. **Zadoks JC, Chang TT, Konzak CF.** A decimal code for the growth stages of cereals. *Weed Research*. 1974;14:415–21.
5. **Waddington SR, Cartwright PM.** A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*. 1983;51:119–30.
6. **Kawakatsu KI, Maekawa M, Izawa T, Itoh J-I, Nagato Y.** *ABERRANT PANICLE ORGANIZATION 2/RFL*, the rice ortholog of *Arabidopsis* *LEAFY*, suppresses the transition from inflorescence meristem to floral meristem through interaction with *APO1*. *The Plant Journal*. 2012;69:168–80.
7. **Bombliès K, Wang R-L, Ambrose BA, Schmidt RJ, Meeley RB, Doebley J.** Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development*. 2003;130:2385–95.
8. **Sayou C, Nanao MH, Jamin M, Posé D, Thevenon E, Grégoire L, et al.** A SAM oligomerization domain shapes the genomic binding landscape of the *LEAFY* transcription factor. *Nature Communications*. 2016;7:11222–12.
9. **Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gérard F, et al.** Structural basis for *LEAFY* floral switch function and similarity with helix-turn-helix proteins. *The EMBO Journal*. 2008;27:2628–37.
10. **Himi S, Sano R, Nishiyama T, Tanahashi T, Kato M, Ueda K, et al.** Evolution of MADS-Box gene induction by *FLO/LFY* genes. *Journal of Molecular Evolution*. 2001;53:387–93.
11. **Prasad K, Kushalappa K, Vijayraghavan U.** Mechanism underlying regulated expression of *RFL*, a conserved transcription factor, in the developing rice inflorescence. *Mechanisms of Development*. 2003;120:491–502.
12. **Lou S, Chen S, Zhao X, Chen L, Zhang J, Fu H, et al.** The far-upstream regulatory region of *RFL* is required for its precise spatial-temporal expression for floral development in rice. *Plant Mol. Biol.* 2017;93:185–95.
13. **Rao NN, Prasad K, Kumar PR, Vijayraghavan U.** Distinct regulatory role for *RFL*, the rice *LFY* homolog, in determining flowering time and plant architecture. *Proc. Natl. Acad. Sci.* 2008;105:3646–51.
14. **Schultz EA, Haughn GW.** *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell*. 1991;3:771–81.
15. **Huala E, Sussex IM.** *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. *Plant Cell*. 1992;4:901–13.
16. **Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM.** *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell*. 1992;69:843–59.
17. **Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R.** *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell*. 1990;63:1311–22.
18. **Souer E, van der Krol A, Kloos D, Spelt C, Bliet M, Mol J, et al.** Genetic control of branching pattern and floral identity during *Petunia* inflorescence development. *Development*. 1998;125:733–42.
19. **Tanahashi T, Sumikawa N, Kato M, Hasebe M.** Diversification of gene function: homologs of the floral regulator *FLO/LFY* control the first zygotic cell division in the moss *Physcomitrella patens*. *Development*. 2005;132:1727–36.
20. **Ahearn KP, Johnson HA, Weigel D, Wagner DR.** *NFL1*, a *Nicotiana tabacum* *LEAFY*-like gene, controls meristem initiation and floral structure. *Plant and Cell Physiology*. 2001;42:1130–9.
21. **Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, et al.** *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Curr. Biol.* 1997;7:581–7.
22. **DeMason DA, Schmidt RJ.** Roles of the *Uni* gene in shoot and leaf development of pea (*Pisum sativum*): phenotypic characterization and leaf development in the *uni* and *uni-tac* mutants. *International Journal of Plant Sciences*. 2001;162:1033–51.
23. **Molinero-Rosales N, Jamilena M, Zurita S, Gómez P, Capel J, Lozano R.** *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *The Plant Journal*. 1999;20:685–93.
24. **Parcy F, Nilsson O, Busch MA, Lee I, Weigel D.** A genetic framework for floral patterning. *Nature*. 1998;395:561–6.
25. **Wagner D, Sablowski RW, Meyerowitz EM.** Transcriptional activation of *APETALA1* by *LEAFY*. *Science*. 1999;285:582–4.

26. **Busch MA, Bomblies K, Weigel D.** Activation of a floral homeotic gene in *Arabidopsis*. *Science*. 1999;285:585–7.
27. **Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, et al.** A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell*. 2001;105:793–803.
28. **Lamb RS, Hill TA, Tan QK-G, Irish VF.** Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. *Development*. 2002;129:2079–86.
29. **Kang H-G, Jang S, Chung J-E, Cho Y-G, An G.** Characterization of two rice MADS box genes that control flowering time. *Molecules & Cells*. 1997;7.
30. **Pelucchi N, Fornara F, Favalli C, Masiero S, Lago C, Pè EM, et al.** Comparative analysis of rice MADS-box genes expressed during flower development. *Sex Plant Reprod*. 2002;15:113–22.
31. **Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, et al.** *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development*. 2003;130:705–18.
32. **Prasad K, Vijayraghavan U.** Double-stranded RNA interference of a rice *PI/GLO* paralog, *OsMADS2*, uncovers its second-whorl-specific function in floral organ patterning. *Genetics*. 2003;165:2301–5.
33. **Yao S-G, Ohmori S, Kimizu M, Yoshida H.** Unequal genetic redundancy of rice *PISTILLATA* orthologs, *OsMADS2* and *OsMADS4*, in lodicule and stamen development. *Plant and Cell Physiology*. 2008;49:853–7.
34. **Nonomura K-I, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, et al.** A germ cell-specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell*. 2007;19:2583–94.
35. **Causier B, Schwarz-Sommer Z, Davies B.** Floral organ identity: 20 years of ABCs. *Seminars in Cell and Developmental Biology*. 2010;21:73–9.
36. **Nonomura K-I, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, et al.** The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*. 2003;15:1728–39.
37. **Lee I, Wolfe DS, Nilsson O, Weigel D.** A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Curr Biol*. 1997;7:95–104.
38. **Chae E, Tan QK-G, Hill TA, Irish VF.** An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development*. 2008;135:1235–45.
39. **Simon R, Carpenter R, Doyle S, Coen E.** *Fimbriata* controls flower development by mediating between meristem and organ identity genes. *Cell*. 1994;78:99–107.
40. **Ingram GC, Goodrich J, Wilkinson MD, Simon R, Haughn GW, Coen ES.** Parallels between *UNUSUAL FLORAL ORGANS* and *FIMBRIATA*, genes controlling flower development in *Arabidopsis* and *Antirrhinum*. *Plant Cell*. 1995;7:1501–10.
41. **Taylor S, Hofer J, Murfet I.** *Stamina pistilloida*, the pea ortholog of *Fim* and *UFO*, is required for normal development of flowers, inflorescence and leaves. *Plant Cell*. 2001;13:31–46.
42. **Souer E, Rebocho AB, Bliet M, Kusters E, de Bruin RAM, Koes R.** Patterning of inflorescences and flowers by the F-Box protein *DOUBLE TOP* and the *LEAFY* homolog *ABERRANT LEAF AND FLOWER* of *Petunia*. *Plant Cell*. 2008;20:2033–48.
43. **Wilkinson LG, Tucker MR.** An optimised clearing protocol for the quantitative assessment of sub-epidermal ovule tissues within whole cereal pistils. *Plant Methods*. 2017;13:67–10.
44. **Aditya J, Lewis J, Shirley NJ, Tan H-T, Henderson M, Fincher GB, et al.** The dynamics of cereal cyst nematode infection differ between susceptible and resistant barley cultivars and lead to changes in (1,3;1,4)- β -glucan levels and *HvCsIF* gene transcript abundance. *New Phytol*. 2015;207:135–47.

Chapter 6



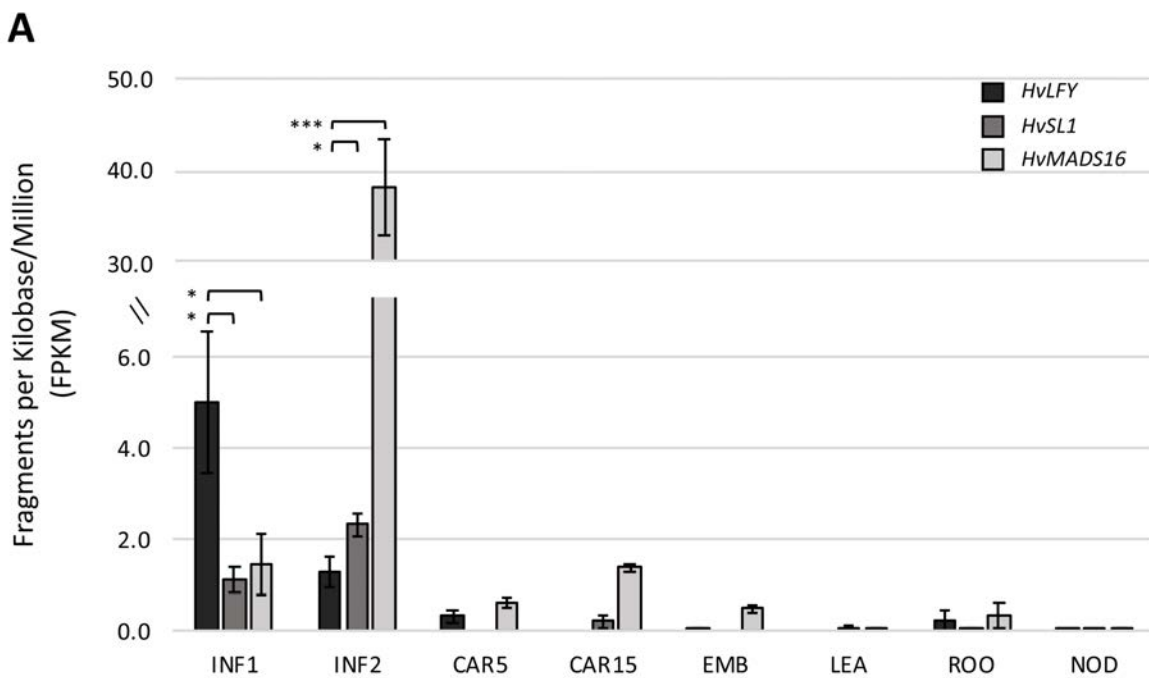
General Synthesis

Solving the jigsaw of barley flower development

Flower development results from a sophisticated balance of environmental and molecular cues. Studying floral mutants allows researchers to gain insight into the main genetic players involved in integration of these cues. In this study, characterization of multiovary mutants *mov1*, *mov2.g* and *mov5.o* and their putative underlying genes *HvMADS16*, *HvSL1* and *HvLFY*, respectively, lays the foundation for understanding floral organ development in barley. Although further genetic and molecular evidence are required to confirm the causative relationship between these barley mutants and the proposed genes, this final discussion is based around the significant findings that *mov1* carries a mutation in *HvMADS16* (gene deletion) (**Chapter 3**), *HvSL1* is mutated in *mov2.g* (gene deletion) (**Chapter 4**) and that *mov5.o* carries a mutation in *HvLFY* (SNP in a highly conserved residue) (**Chapter 5**). When considered in light of the remarkable phenotypic similarities between the rice mutants *superwoman1* (*Osspw1/Osmads16*), *stamenless1* (*Ossl1*) and *aberrant panicle organization 2/rfl* (*Osapo2/rfl*) and the respective barley mutants *mov1/Hvmads16*, *mov2.g/Hvsl1* and *mov5.o/Hvlfy*, our findings gain additional support. The results obtained also enable the formulation of a testable model that can be used in further studies of barley flower development.

Expression data from public resources (<https://ics.hutton.ac.uk/morexGenes/>) [1] show that *HvMADS16*, *HvSL1* and *HvLFY* are mainly expressed in floral and reproductive tissues in wild-type barley (cv. Morex) (**Figure 1**). This is consistent with the most prominent phenotype in

the mutant multiovary plants being seen in the flower. In addition, *HvLFY* expression appears to temporally precede expression of both *HvSL1* and *HvMADS16*, indicating that *HvLFY* may act at a relatively earlier stage of flower development. For example, although *HvLFY* is expressed at low abundance in young inflorescences (5 mm), it is significantly higher than both *HvSL1* and *HvMADS16* (**Figure 1**). As development proceeds *HvLFY* abundance decreases while both *HvSL1* and *HvMADS16* increase, so that once inflorescences are 1 – 1.5 cm in size, *HvLFY* is significantly lower than the other genes (**Figure 1**). The low FPKM values for *HvLFY* and *HvSL1* might be explained by these genes only being expressed in a small subset of cells within the entire tissue. Although *in situ* hybridisation resolved the location of *HvMADS16* in stamen and lodicule primordia, further experiments will be required to confirm the expression domain and timing of *HvLFY* and *HvSL1*.



B

Tissue	Description
INF1	Young developing inflorescences (5 mm)
INF2	Developing inflorescences (1 - 1.5 cm)
CAR5	Developing grain, bracts removed (5 DPA)
CAR15	Developing grain, bracts removed (15 DPA)
EMB	4-day embryos dissected from germinating grains
LEA	Shoots from the seedlings (10 cm shoot stage)
ROO	Roots from the seedlings (10 cm shoot stage)
NOD	Developing tillers at six leaf stage, 3 rd internode

Figure 1. (A) Expression pattern for barley genes *HvLFY* - *HORVU2Hr1G102590* (black), *HvSL1* - *HORVU3Hr1G003740* (dark grey) and *HvMADS16* - *HORVU7Hr1G091210* (light grey) in tissues at different developmental stages. Error bars represent \pm Standard Deviation. For INF1 and INF2, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**), and ≤ 0.001 (***) are shown. For each sample *n* = 3 independent biological replicates. (B) Short description for each tissue, DPA represents Days after Pollination. A more detailed description and expression values can be found at <https://ics.hutton.ac.uk/morexGenes/>.

When comparing the expression profiles obtained in this study for *HvLFY*, *HvSL1* and *HvMADS16* across all three mutants, possible interactions between the genes can be inferred (**Figure 2**). Transcription factors *HvSL1* and *HvLFY* appear to positively influence *HvMADS16* expression, as this gene seems to be down-regulated in both *mov2.g* and *mov5.o*. Specifically, *HvMADS16* expression in *mov2.g* inflorescences is lower than in wild type even before stamen primordia initiation (W3.5), while *HvMADS16* expression in *mov5.o* begins to decrease around the stage of stamen primordia appearance (W3.5) (**Figure 2**). Consistent with the hypothesis that *HvLFY* and *HvSL1* act upstream of *HvMADS16*, expression profiles of *HvSL1* and *HvLFY* do not significantly differ from wild type in *mov1* plants (**Figure 2**). However, *HvLFY* appears to act on *HvMADS16* in a *HvSL1*-independent manner, as *HvSL1* expression remains largely unchanged in *mov5.o* inflorescences. Likewise, *HvLFY* expression seems to be unaffected in *mov2.g* plants (**Figure 2**).

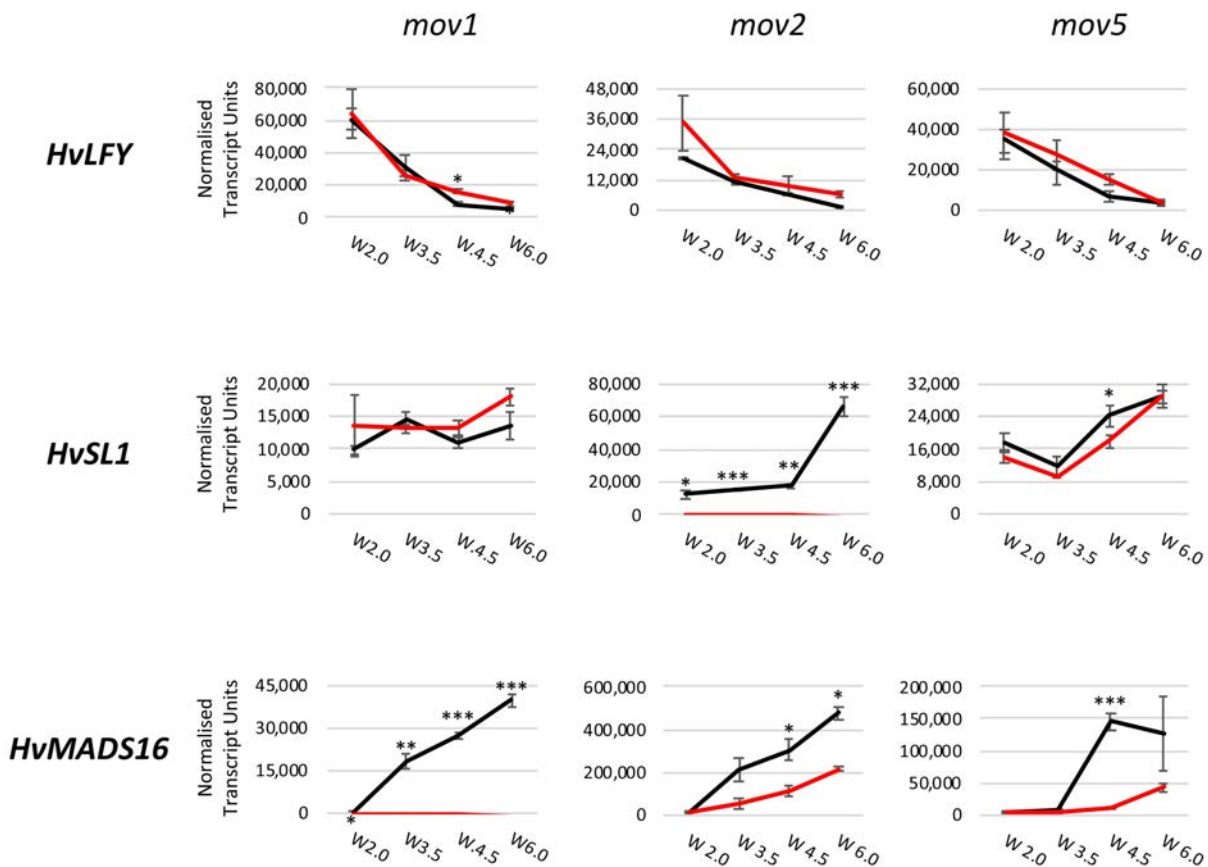


Figure 2. Transcript abundance assessed by qRT-PCR in wild-type (black) and mutant *mov1*, *mov2.g* and *mov5.o* (red) developing inflorescences at stages W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development) for: *HvLFY* (*HORVU2Hr1G102590*), *HvSL1* (*HORVU3Hr1G003740*) and *HvMADS16* (*HORVU7Hr1G091210*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**) and ≤ 0.001 (***) are shown for differences between wild type and mutant. For each sample $n = 3$ independent biological replicates.

The importance of *HvMADS16* can also be inferred from the fertility of the three mutants; carpels of *mov1* florets are completely sterile and lack a female gametophyte, while carpels produced by *mov2.g* or *mov5.o* partly retain the ability to produce a female gametophyte and viable seeds. This is in line with the observation that *HvMADS16* is fully absent in *mov1* plants, but is still expressed, despite being lower in abundance, in both *mov2.g* and *mov5.o*. This finding suggests that *HvMADS16* is not only required for correct floral organ specification, but also for correct female gametophyte development. In dicotyledons,

expression of the *HvMADS16* orthologue in *Antirrhinum* (*DEFICIENS*) has been detected in developing carpels [2], while transcripts of the *Arabidopsis* orthologue, *APETALA3*, accumulate in ovule integuments late in development [3]. In addition, *in situ* hybridisation confirmed that *HPDEF*, the orthologous gene in *Hieracium piloselloides* (tall hawkweed), is expressed in ovules during female gametophyte initiation until anthesis [4]. Early studies in rice reported the expression of B-class genes *OsMADS2* and *OsMADS4* mainly in anthers and carpels. However, recent studies state that there is no RNA detected in the gynoecium for both *OsMADS2* [5] and *OsMADS16* [6]. A different pattern is observed in maize, whereby the *HvMADS16* orthologue *SILKY1* was found to be initially expressed in the centre of the floral meristem that produces the pistil primordium but is later lost in the pistil primordium itself [7]. Preliminary transcriptomic data from our laboratory (*L. Wilkinson, M. Tucker, unpublished*) indicates that *HvMADS16* and *HvMADS2* are expressed in the later stages of developing pistil and ovule tissues, while *HvMADS4* is barely detectable (**Figure 3**).

Taken together, we speculate that *HvMADS16*, apart from being a determinant of lodicule and stamen specification, also has a role in female gametophyte development. This later role of *HvMADS16* could possibly be performed via regulation or interaction with the B-sister MADS-box proteins, *HvMADS29* and *HvMADS31*. For example, yeast two-hybrid and pull-down assays in the gymnosperm *Gnetum gnemon* have demonstrated hetero-dimerization between the B-sister and B-class orthologues [8]. However, further studies exploring this hypothesis in barley are needed.

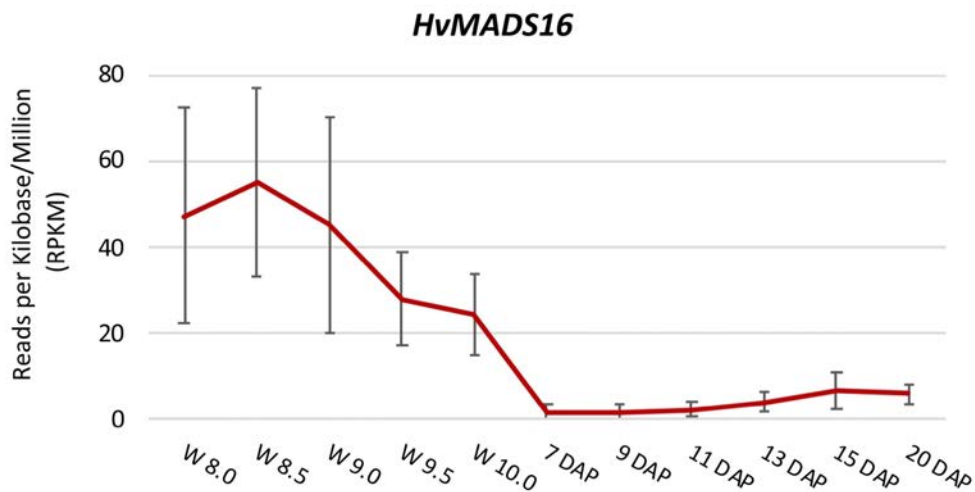


Figure 3. Transcript abundance for barley *HvMADS16* (*HORVU7Hr1G091210*) in mature carpels before and after anthesis (Days After Pollination, DAP). Stages correspond the Zadok growth scale. The transcript abundance shown is the average of values from six barley cultivars: Golden Promise (2 biological replicates for each timepoint), Salka, Wren, Forum and Gant (3 biological replicates for each timepoint). Error bars represent \pm Standard Deviation.

Based on the results described in this thesis, we present a unifying model for the interaction of *HvLFY*, *HvSL1* and *HvMADS16* in stamen development (**Figure 4**). We propose that in wild-type barley, *HvLFY* is expressed early in floral meristem differentiation, before the stamen primordia appear, and likely via interaction with an unknown F-box protein (possibly *HORVU7Hr1G108970*) positively regulates *HvMADS16* in the second (lodicules) and third (stamens) floral whorls. At presumably the same time, *HvSL1* drives independent upregulation of *HvMADS16* chiefly in whorl 3. Once *HvMADS16* is expressed in whorls 2 and 3 (**Chapter 3, Figure 6**), it forms obligate hetero-dimers with the other B-class transcription factors *HvMADS2* and *HvMADS4* (**Chapter 3, Figure 7**), probably initiating a positive regulatory feedback loop. As a result, in whorl 3 the higher-order floral quartet between B, C and E-class genes represses the expression of *HvMADS3*, *HvMADS58*, *HvMADS15* and *HvDL* that are necessary for carpel and ovule development, while promoting stamen formation. In addition, *HvLFY* also plays a role at the very early stages of inflorescence development to

establish the correct molecular context, likely via regulation of E-class genes, while *HvMADS16* possibly also functions at later stages in female gametophyte development. In multiovary mutants, this regulatory network is affected at strategic checkpoints (**Figure 4**). The most severe phenotypic effect occurs in *mov1* whereby *HvMADS16* is absent (**Chapter 3, Figure 2**). As the remaining B-class genes do not interact amongst themselves (**Chapter 3, Figure 7**), they cannot compensate for the lack of *HvMADS16*. A less pronounced phenotypic effect is observed in *mov2.g* and *mov5.o*, in which *HvMADS16* is observed to be downregulated, but is still present (**Chapter 4, Figure 6; Chapter 5, Figure 7**). In each case there is reduced expression of all B-class genes, which we interpret to be insufficient to promote stamen development. Thus, the floral quartet most likely present in whorl 3 is between C and E-class genes, promoting carpel development instead of stamens.

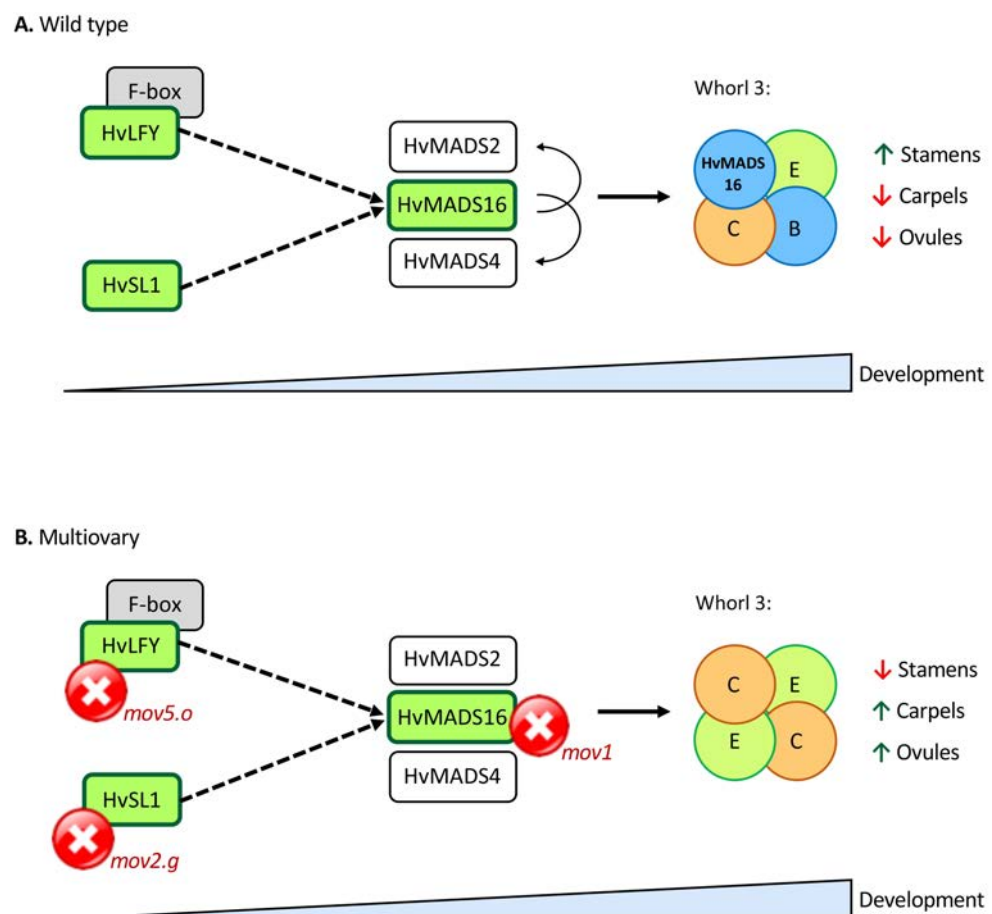


Figure 4. Model for barley stamen development. **(A)** In wild-type barley HvLFY is expressed early in floral meristems and activates HvMADS16 in whorls 2 and 3, likely via interaction with an F-box protein. Concomitantly, HvSL1 independently upregulates *HvMADS16* expression predominantly in whorl 3. In whorl 3, HvMADS16 forms hetero-dimers with the other B-class factors HvMADS2 and HvMADS4 and, together with C- and E-class MADS-box proteins, form floral quartets able to promote formation of stamens and repress carpel and ovule development. **(B)** In a multiovary context, the network is disrupted at different steps. However, all three *mov1*, *mov2.g* and *mov5.o* mutations have the effect of decreasing HvMADS16 to levels insufficient for stamen formation. The remaining B-class factors are unable to form dimers, thus the quaternary complex that predominantly forms in whorl 3 of multiovary plants is solely between C- and E-class proteins which supports carpel and ovule development. Dashed arrows indicate indirect interactions or interactions that have not been tested yet, while solid line arrows indicate direct interactions as seen in this study. Green boxed proteins represent the likely candidates for *mov1*, *mov2.g* and *mov5.o*, which the present study focused on. Grey boxed protein indicates speculative interactor based on information from other species.

Applicability of barley *mov* mutants to hybrid breeding

In a hybrid breeding context, barley *mov* mutants in which stamens have been converted into additional carpels may present the double advantage of being inherently male-sterile and being able to increase the number of receptive carpels per spike. Male-sterility ensures that any seed set is the consequence of cross-pollination, whereas an increase in the number of receptive carpels per floret increases the likelihood of successful seed set per plant from wind-borne pollination. Although having multiple carpels might be seen as a disadvantage due to potential resource competition between multiple seeds, wind-pollinated seed set from all carpels in a spike is unlikely in a hybrid seed production setting, and hence the phenotype may indeed improve average seed set. Taken together, one could imagine multiovary mutants and their wheat counterparts might contribute to a lowering of hybrid seed production costs. Although, such characteristics seem ideal, the deployment of *mov* or *mov*-like mutants in barley or wheat could be hindered by adverse pleiotropic effects.

Among the three barley *mov* mutants described in this thesis, the most detrimental phenotype is found in *mov1* (cv. Steptoe) as it is completely male and female sterile and therefore cannot be used as a parent in hybrid breeding. Interestingly, a mutation at seemingly the same locus appears to have less severe effect in the spontaneous mutant *mo5* (cv. Revelatum 1886) [9]. Indeed, *mo5* plants produced viable seeds when cross-pollinated with the cultivars Pallidum, Nutans and Nudum [10]. The difference in phenotypic severity may be a consequence of diverse genotypic background or of the type of lesion caused by the mutation. Nonetheless, this raises the possibility that negative effects of *mov1* on female fertility may be ameliorated by selecting weaker alleles.

Another mutant that would present complications for use in hybrid breeding is *mov5.o* (cv. Morex), as it partially retains florets having a variable number of stamens capable of producing viable pollen (**Chapter 5, Figure 1**). The incomplete male-sterility of *mov5.o* would likely reduce the F₁ seed purity if used in hybrid breeding. Furthermore, it was observed that *mov5.o* plants tend to possess a twisted and relatively weaker peduncle relative to wild type (**Appendix C, Supplementary Figure S1**). Whether such curled peduncle affects seed yield through disadvantageous traits is yet to be determined, but one could expect that this characteristic may render the plant susceptible to head-loss before harvest, ultimately affecting yield.

The most promising multiovary mutant among the three studied appears to be *mov2.g* (cv. Steptoe). Not only is *mov2.g* completely male-sterile and female-fertile, but florets are also able to produce multiple seeds (2 - 3) per floret (**Chapter 4, Figure 1**). Such characteristics

are of great interest to hybrid breeding since having multiple functional carpels within the same floret maximises the opportunities for successful cross-pollination.

Interestingly, when cross-pollinated *mov2.g* florets produced two seeds, both seeds appeared to be of smaller size compared to wild-type (**Figure 5**). On the other hand, when three seeds developed within the same floret, only the central seed retains a similar size to wild type, with the two lateral seeds being visibly smaller (**Figure 5**). This effect could be a consequence of physical space limitation within the floret coupled with an uneven nutrient allocation among the three seeds. Whether this is reproducible in the field is yet to be determined. However, it is worth noting that in a hybrid seed production setting it is unlikely that all pistils would be fertilised, thus reducing the competition threshold and allowing for the growth of average-sized seeds.



Figure 5. Seed size from (A) wild-type and *mov2.g* florets developing (B) double or (C) triple seeds. Scale bars: 1000 μm .

Being a recessive and self-sterile mutation, the *mov2* locus needs to be maintained in a heterozygous condition for self-seed propagation. This would necessarily introduce an additional breeding step to multiply multiovary individuals. However, genotyping the locus

at an early stage using molecular markers would facilitate trait screening, and many systems for maintenance of male-sterile plants are already in use for hybrid breeding.

Other aspects that need to be considered for the use of *mov2.g* plants in hybrid breeding are a low seed germination rate (**Appendix B, Supplementary Table S1**) and the low proportion of mutant segregants observed in this research (**Appendix B, Supplementary Table S2**). Further studies are therefore necessary to determine if these effects are caused by the *mov2* locus, which would render it deleterious for hybrid breeding, or if they are symptomatic of residual mutational load. Additionally, it was noticed that *mov2.g* individuals tend to have reduced size and vigour relative to wild-type individuals. Reduced vigour would be unappealing for hybrid breeding as it may mean that *mov2.g* plants could be more susceptible to disease and abiotic stress in the field. Furthermore, reduced vigour may affect flowering time, making matching the flowering window between female and male parents more challenging.

Ultimately, it is apparent from this study that, as seen in many instances in biology, use of multiovary mutants for hybrid breeding is likely to be a trade-off between the desired traits and disadvantageous phenotypes. Understanding the physiology of multiovary mutants provides us with the tools to regulate and minimize any unwanted effects.

Transferring knowledge to wheat

The genetic knowledge of flower development generated in this research can conceivably be transferred to wheat. This would first entail applying bioinformatics to identify the wheat orthologs of the genes identified here. As wheat is allohexaploid there are expected to be three orthologous wheat genes for each barley gene. Once identified, the expression pattern

of the candidate homeologues across tissues and through development would be assayed. To determine their function, candidate genes could be targeted and rendered non-functional via site-directed mutagenesis, for example via the use of CRISPR/Cas9 or TILLING. If there is functional conservation between wheat and barley, this is likely to lead to a similar phenotype as observed in barley, although in wheat more than one homeologue may need to be knocked out. This process could also be combined with analysis of the allelic variation between commercial wheat varieties worldwide to select lines requiring the least degree of modification.

To test this principle, the *HvSL1* gene sequence was used to identify wheat (cv. Chinese Spring) orthologues *TaSL1(3A)* (*TraesCS3A01G043300*), *TaSL1(3B)* (*TraesCS3B01G040200*) and *TaSL1(3D)* (*TraesCS3D01G037400*), located on chromosomes 3A, 3B and 3D, respectively. Transcript abundance of these wheat genes suggests that all three homeologues are expressed at a comparable abundance in floral tissues (http://bar.utoronto.ca/eplant_wheat/) [11] (**Figure 6**), possibly indicating that all three copies are functional. It would be interesting to create single knock-out mutants for each homeologue, as well as double and triple mutants. An experiment of this kind would: (1) verify if all wheat copies are functional, (2) assess if homeologues have undergone sub-functionalization and (3) provide information about functional conservation with barley. Interestingly, *TaSL1(3A)*, *TaSL1(3B)* and *TaSL1(3D)* appear to be expressed at low levels in young lemmas (Zadok stage Z59) [12], at very low levels in stigma and ovary, and not expressed in mature anthers (**Figure 6**). This observation could suggest a diverse role in flower development among wheat *SL1* copies, thus highlighting the importance of validating results when transferring knowledge across species.

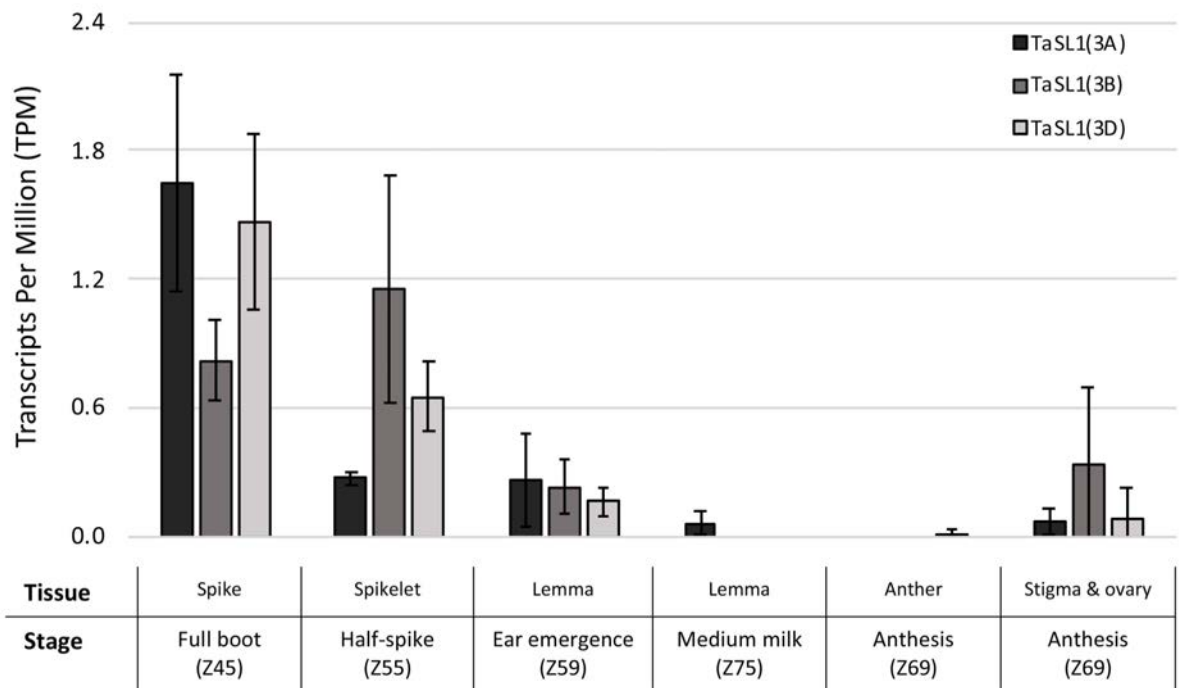


Figure 6. Expression pattern for wheat genes *TaSL1(3A)/TraesCS3A01G043300* (black), *TaSL1(3B)/TraesCS3B01G040200* (dark grey) and *TaSL1(3D)/TraesCS3D01G037400* (light grey) in reproductive and floral tissues during middle and late stage development. Stages according to the Zadok growth scale have been included. Error bars represent \pm Standard Deviation. For each sample $n = 3$ independent biological replicates. Expression values used are taken from http://bar.utoronto.ca/eplant_wheat/.

If we assume conservation of function, *TaSL1* could potentially be deployed in a hybrid wheat program in a similar context to Seed Production Technology (SPT, Pioneer Hybrid International, Inc.) to develop transgenic-free hybrid seeds. Functional analysis of *TaSL1(3A)*, *TaSL1(3B)* and *TaSL1(3D)* can inform if triple mutant stacking is necessary to obtain a completely penetrant multiovary phenotype and if, in a breeding scenario, all three loci need be mutated or backcrossed in the female parent. Although backcrossing three loci increases the breeding effort, the use of *TaSL1* may still prove advantageous. In a hybrid breeding context, use of a recessive trait such as multiovary is simpler for fertility restoration in F_1 plants, as opposed to deploying a dominant trait.

References

1. **IBSC, International Barley Genome Sequencing Consortium.** A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012;491:711–6.
2. **Schwarz-Sommer Z, Hue I, Huijser P, Flor PJ, Hansen R, Tetens F, et al.** Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal*. 1992;11:251–63.
3. **Jack T, Brockman LL, Meyerowitz EM.** The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*. 1992;68:683–97.
4. **Guerin J, Rossel JB, Robert S, Tsuchiya T, Koltunow A.** A *DEFICIENS* homologue is down-regulated during apomictic initiation in ovules of *Hieracium*. *Planta*. 2000;210:914–20.
5. **Chung Y-Y, Kim S-R, Kang H-G, Noh Y-S, Park MC, Finkel D, et al.** Characterization of two rice MADS box genes homologous to *GLOBOSA*. *Plant Science*. 1995;109:45–56.
6. **Kyozuka J, Kobayashi T, Morita M, Shimamoto K.** Spatially and temporally regulated expression of rice MADS box genes with similarity to *Arabidopsis* class A, B and C genes. *Plant and Cell Physiology*. 2000;41:710–8.
7. **Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ.** Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell*. 2000;5:569–79.
8. **Wang Y-Q, Melzer R, Theissen G.** Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of “floral quartets.” *The Plant Journal*. 2010;64:177–90.
9. **Soule JD, Skodova, Kudrna DA, A K, Kleinhofs A.** Molecular and genetic characterization of barley flower development mutants. *Barley Genetics Newsletter*. 1996;76–80.
10. **Tazhin.** The linkage of the genes *mo5** and *n* in barley. *Barley Genetics Newsletter*. 1980;10(II):69–72.
11. **Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al.** The transcriptional landscape of polyploid wheat. *Science*. 2018;361:1–12.
12. **Zadoks JC, Chang TT, Konzak CF.** A decimal code for the growth stages of cereals. *Weed Research*. 1974;14:415–21.

Conclusions



The research presented in this thesis has led to the following conclusions:

1. The barley multiovary *mov1* (cv. Steptoe) mutant involves a deletion of ~ 0.95 Mb on the long arm of chromosome 7H. The missing region contains the DEF-like B-class gene *HvMADS16*.
2. The barley multiovary *mov2* locus maps to a ~ 449 Kb region at the telomeric end of the short arm of chromosome 3H. The region of interest includes the C2H2 zinc finger transcription factor *HvSL1*, which is absent in *mov2.g* plants (cv. Steptoe).
3. The barley multiovary *mov5* locus is located on the long arm of chromosome 2H within an interval of ~ 3.8 Mb. In *mov5.o* plants (cv. Morex) the FLO/LFY transcription factor *HvLFY* contains a [C/T] transition which results in a Pro → Leu non-synonymous substitution at a highly conserved protein residue.
4. As *mov1*, *mov2.g* and *mov5.o* are positioned on different chromosomes, the three loci are not allelic.

5. The same type of mutagenesis (fast neutron irradiation) can result in different genomic lesions and varying degrees of severity.
6. The presence of *HvMADS16* is necessary for correct specification of lodicules, stamens and female gametophyte development. *HvSL1* is needed for the proper differentiation of stamens. Absence of *HvSL1* does not have a pronounced effect on female gametogenesis. *HvLFY* is necessary for proper floral organ development.
7. Both transcription factors *HvLFY* and *HvSL1* appear to act upstream of *HvMADS16*, potentially activating *HvMADS16* in an independent manner. However, *HvSL1* alone is not sufficient to directly regulate the activity of B-class genes.
8. Genes that affect reproductive organ specification act very early in inflorescence development, before and/or at the onset of organ primordia. Mutations affecting these genes cause perturbations of the ABC model.
9. Barley B-class genes encode DEF-GLO obligate hetero-dimers, as demonstrated for *Arabidopsis* and rice B-class genes.

Contributions to knowledge



The research reported in the present thesis was targeted towards the developmental characterization of barley multiovary mutants *mov1*, *mov2.g* and *mov5.o* and the identification of their underlying genes, to provide information about male-sterile mutants that could be applied to enhance cross-pollination in wheat hybrid breeding.

The contributions to knowledge of this thesis are:

1. Enhanced knowledge of flower development in barley through the identification of candidate genes necessary for correct floral organogenesis, and development of a unifying model to explain their interaction in the floral network.
2. Validation of the ABC model in barley and demonstration that, as for other species, the model undergoes species-specific modifications in barley.
3. Demonstration that the multiovary phenotype is a spectrum, depending on what checkpoint of the floral organogenesis pathway is affected. While some mutations giving rise to multiovary can result in multiple seeds, others show adverse pleiotropic effects such as complete female sterility or partial conversion of floral organs. This affects applicability and determines if the phenotype is commercially profitable.

4. Discovery that there are multiple ways to obtain a multiovary plant, which introduces the possibility of achieving a multiovary plant that maximises the gain in cross-pollination with no or little compromising pleiotropic effects.

To my knowledge, this is the first report of a detailed characterization of barley mutants *mov1*, *mov2.g* and *mov5.o* combined with a model explaining their interaction in driving flower development. Although further experiments are required to support the conclusions, this thesis represents a step forward towards our understanding of flower development in barley, with the hope that one day Science can build upon these discoveries.



Appendix A



Supplementary Material to Chapter 3

Supplementary Figure S1. Development of wild-type and *mov1* inflorescences and reproductive organs.

Supplementary Figure S2. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type and *mov1* inflorescences.

Supplementary Figure S3. *In situ* hybridization with sense probes on wild-type and *mov1* inflorescences.

Supplementary Figure S4. Sequence alignment of MADS16 proteins from *Hordeum vulgare*, *Oryza sativa*, *Zea mays*, *Antirrhinum majus* and *Arabidopsis thaliana*.

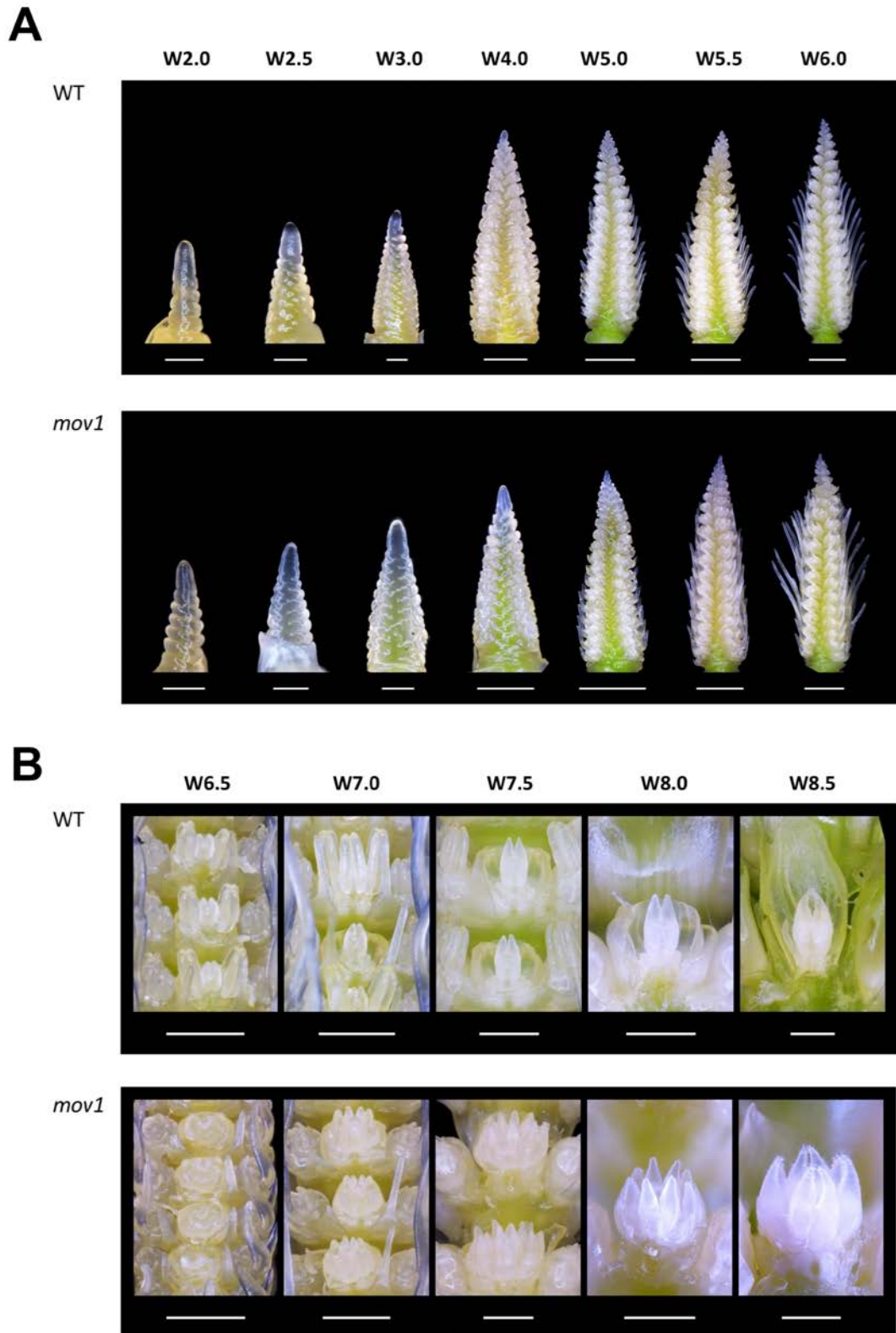
Supplementary Figure S5. Complementation and CRISPR strategy for *mov1*.

Supplementary Table S1. Segregation ratio of *mov1* by genotyping with copy number analysis and phenotyping in heterozygote plant material.

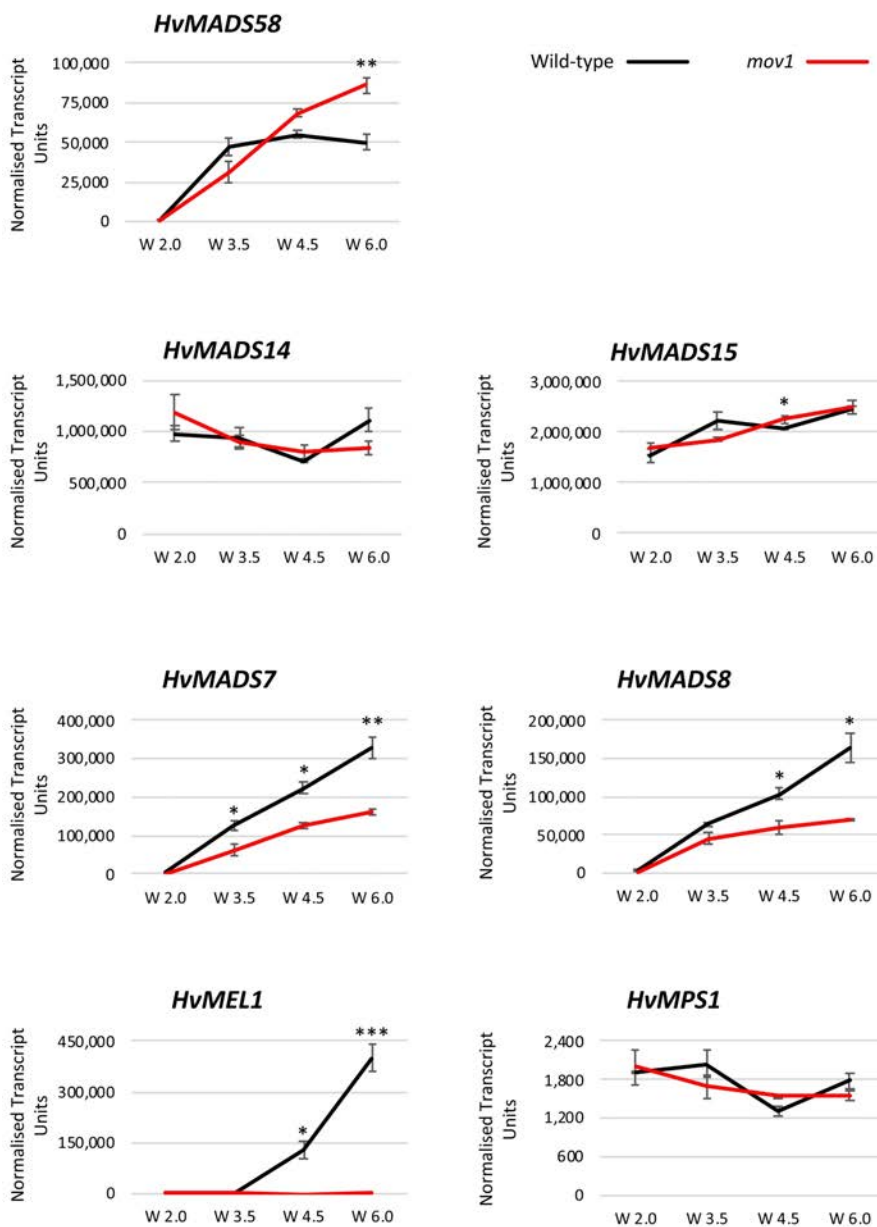
Supplementary Table S2. Annotated genes present in the *mov1* deletion (7H) and genes tested by PCR

Supplementary Table S3. Primer sequences used in this study.

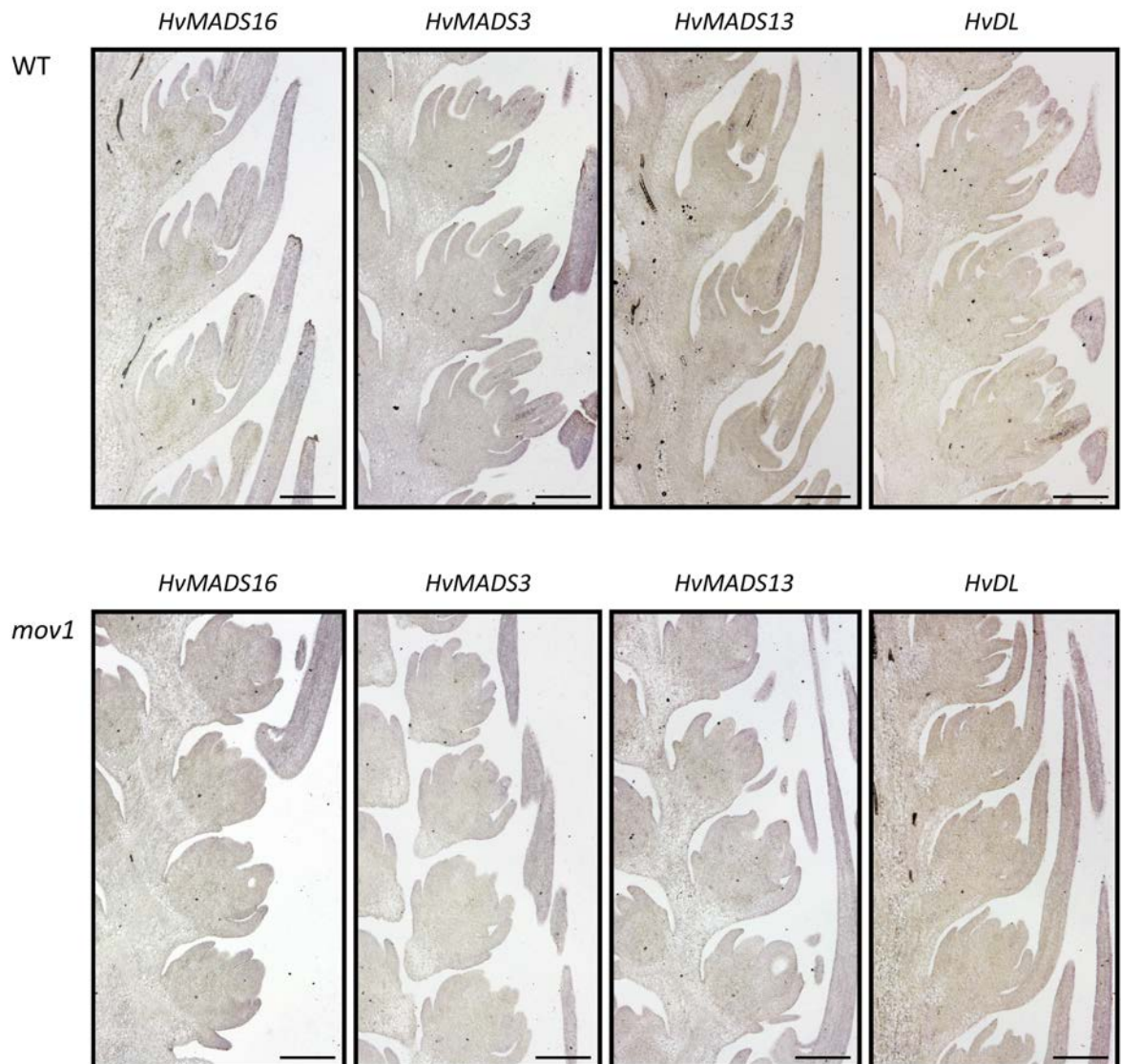
Supplementary Figure S1. Development of wild-type (WT) and *mov1* (A) inflorescences and (B) reproductive organs. Waddington stage is indicated for each developmental stage. Scale bars are: 250 μ m (W2.0 - W3.0), 500 μ m (W4.0), 1000 μ m (W5.0 - 6.0) and 500 μ m (W6.5 - 8.5).



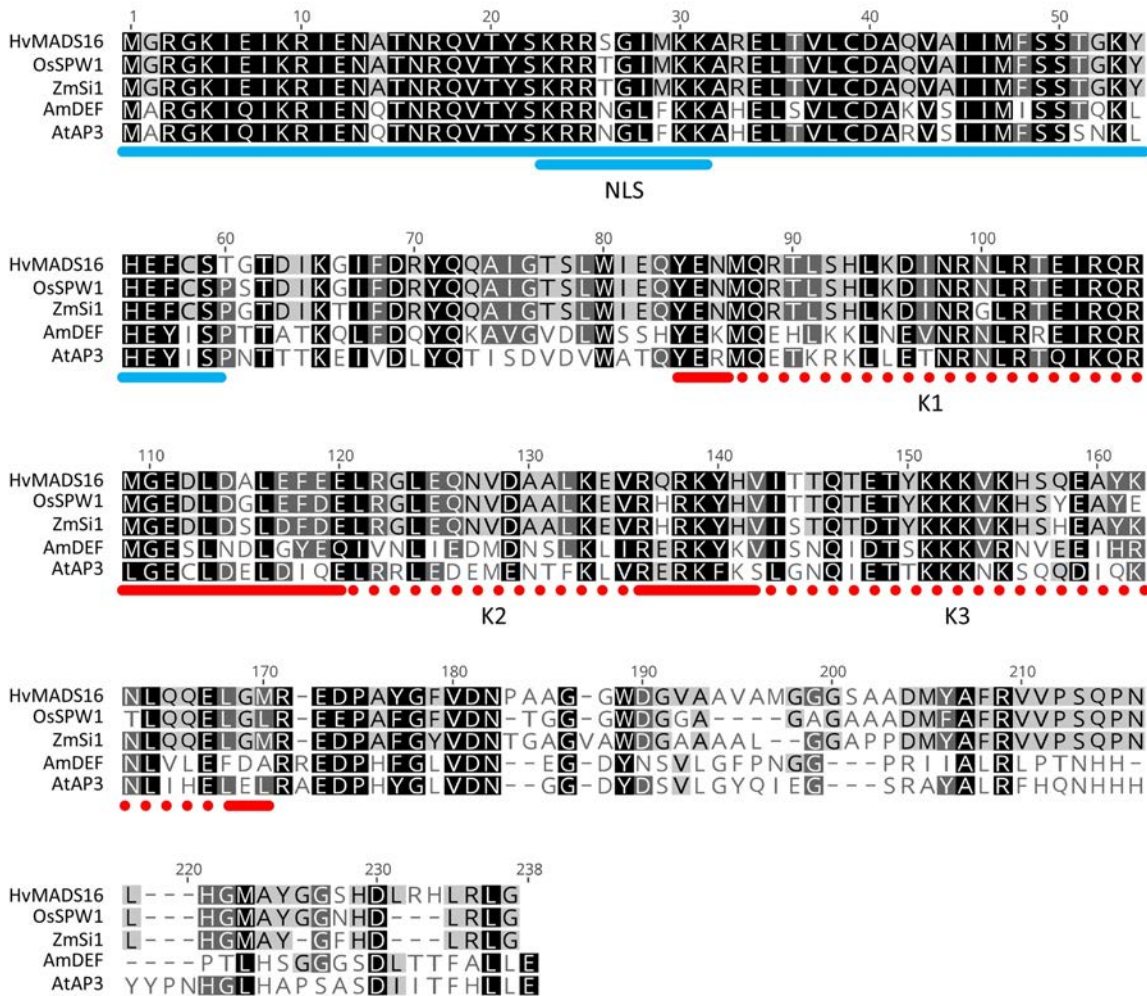
Supplementary Figure S2. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type (black) and *mov1* (red) inflorescences. Developmental stages refer to W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development) Transcript abundance is shown for C-class gene *HvMADS58* (*HORVU1Hr1G029220*); A-class genes *HvMADS14* (*HORVU5Hr1G095630*), *HvMADS15* (*HORVU2Hr1G063800*); E-class genes *HvMADS7* (*HORVU5Hr1G076400*), *HvMADS8* (*HORVU7Hr1G054220*), as well as *HvMEL1* (*HORVU5Hr1G107020*) and *HvMSP1* (*HORVU6Hr1G033670*). Error bars represent \pm Standard Error. For each developmental stage, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**) and ≤ 0.001 (***) are shown for differences between wild type and *mov1*. For each sample *n* = 3 independent biological replicates.



Supplementary Figure S3. *In situ* hybridization with sense probes on wild-type and *mov1* inflorescences. Sense probes for *HvMADS16* (HORVU7Hr1G091210), *HvMADS3* (HORVU3Hr1G026650), *HvMADS13* (HORVU1Hr1G023620) and *HvDL* (HORVU4Hr1G067780) were assayed on inflorescences at stage W6.0. Scale bars: 250 μ m.



Supplementary Figure S4. Sequence alignment of MADS16 proteins from *Hordeum vulgare* (HvMADS16 - HORVU7Hr1G091210), *Oryza sativa* (OsSPW1 - LOC_Os06g49840), *Zea mays* (ZmSi1 - LOC541799), *Antirrhinum majus* (AmDEF - Am01g03890) and *Arabidopsis thaliana* (AtAP3 - AT3G54340). The conserved MADS (blue line) and K-domain (red line) are shown, double lines indicate the conserved nuclear localization signal (NLS), dashed lines indicate the amphipatic α -helices K1, K2 and K3 which mediate interactions between MADS-box proteins. Figure contains information from Yang *et al.* (2003); Gramzow and Theissen (2010).

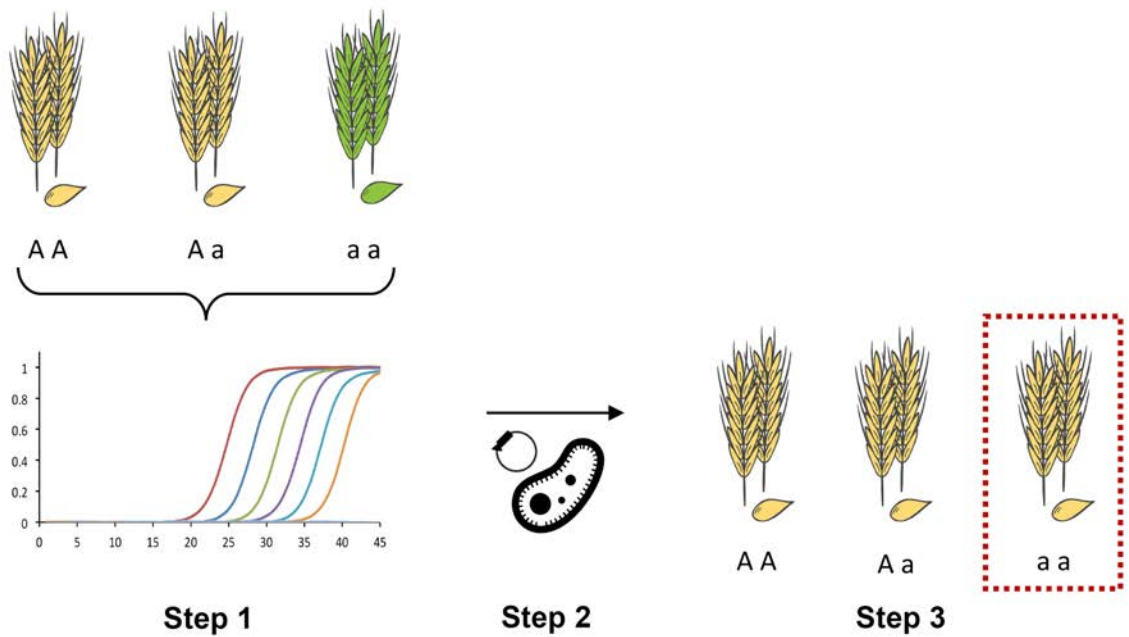


Yang Y, Fanning L, Jack T. The K domain mediates hetero-dimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant Journal*. 2003;33:47-59.

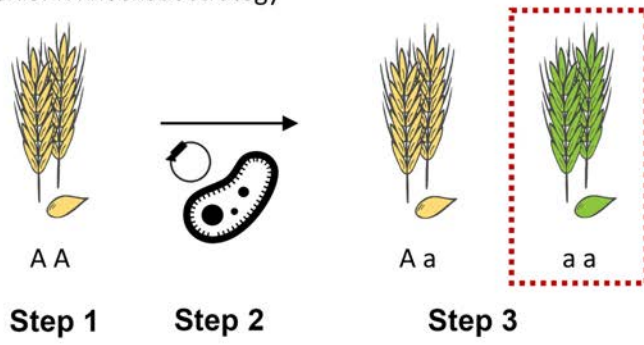
Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. *Genome Biology*. 2010;11:214.

Supplementary Figure S5. Complementation and CRISPR strategy for *mov1*. **(A)** Strategy for *mov1* complementation; *mov1* (cv. Steptoe) segregating plants are grown and copy number is performed to identify heterozygote plants at an early stage (**Step 1**), these plants can then be used as donor material (T_0) for *Agrobacterium*-mediated transformation (**Step 2**). If complementation is successful, regenerants (T_1) from the transformation should segregate genotypically, but not phenotypically (**Step 3**). **(B)** CRISPR knockout strategy; Golden Promise plants are used donor material (T_0) (**Step 1**) for *Agrobacterium*-mediated transformation (**Step 2**). Regenerants (T_1) are screened for both phenotype and genotype (**Step 3**). Yellow represents wild-type phenotype, green indicates *mov1* phenotype. **(C)** Position of the guideRNA (red); two different gRNA have been designed to target *HvMADS16* (*HORVU7Hr1G091210*) at +29 bp and +72 bp from the translational start site. **(D)** Details of the CRISPR/Cas9 vector. Vector includes a plant codon optimized Cas9 driven by the maize ubiquitin promoter (pZmUbi) and both gRNA (gRNA1 and gRNA2) driven by the rice small nuclear RNA promoters U6a (pOsU6a) and U6b (pOsU6b), respectively. NosT stands for the *Agrobacterium* nopaline synthase terminator.

A. Complementation strategy



B. CRISPR knockout strategy



C. gRNA target sites



D. CRISPR/Cas9 vector



Supplementary Table S1. Segregation ratio of *mov1* by genotyping with copy number analysis and phenotyping in heterozygote plant material.

Group #	Observed <i>mov1</i> phenotype	Observed wild-type phenotype	Total	χ^2 value	P-value	Output
1	68 (24 %)	214 (76 %)	282	0.118	0.731	ACCEPT H_0
2	10 (25 %)	30 (75 %)	40	0.000	1.000	ACCEPT H_0
3	10 (26 %)	29 (74 %)	39	0.009	0.926	ACCEPT H_0
4	9 (19%)	39 (81 %)	48	1.000	0.317	ACCEPT H_0
5	41 (24 %)	133 (76 %)	174	0.192	0.662	ACCEPT H_0

H_0 = The observed phenotypes segregate with a 3:1 ratio

H_1 = Not H_0 ; observed phenotypes segregate differently from 3:1 ratio

Degrees of freedom (DF) = 1

Supplementary Table S2. Annotated genes present in the *mov1* deletion (7H) and genes tested by PCR. All genes were present in *mov1*, except for genes shown in bold. Annotations and genomic coordinates based on the Morex reference assembly Hv_IBSC_PGSB_v2. Presence of gene *HORVU7Hr1G091190* could not be tested due to sequence repetitiveness and is therefore considered within the deletion.

Gene	Start	End	Strand	Annotation
<i>HORVU7Hr1G091170</i>	556,426,170	556,428,917	+	Protein FMP32, mitochondrial
<i>HORVU7Hr1G091180</i>	556,432,734	556,434,307	+	B-box zinc finger family protein
<i>HORVU7Hr1G091190</i>	557,017,373	557,018,262	-	40S ribosomal protein
<i>HORVU7Hr1G091200</i>	557,181,329	557,182,073	-	Undescribed protein
<i>HORVU7Hr1G091210</i>	557,244,345	557,245,839	-	MADS-box transcription factor 16
<i>HORVU7Hr1G091220</i>	557,387,804	557,392,562	+	UPF0183 protein
<i>HORVU7Hr1G091230</i>	557,392,887	557,396,669	+	Unknown protein
<i>HORVU7Hr1G091240</i>	557,397,068	557,398,785	-	Alpha-amylase-like
<i>HORVU7Hr1G091250</i>	557,426,822	557,429,314	-	Alpha-amylase-like

Supplementary Table 3A. Primer sequences and Taqman probes used for copy number analysis to genotype *mov1* plants. Fluorophore at 5' and 3' quencher are indicated for each probe.

Gene name	Gene ID	Primer sequences	Taqman probe
<i>HvCO-like</i>	<i>HORVU6Hr1G072620</i>	TGCTAACCGTGTGGCATCAC GGTACATAGTGCTGCTGCATCTG	[HEX] CATGAGCGTGTGCGTGTCTGCG [BHQ1]
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	GTTTACCTTGCCTTGTGTGCG ACGAACTGCTTTCTCAAACG	[FAM] AGTTCTCCATGCCACTGCTCAAACACCA [BHQ1]

Supplementary Table 3B. PCR primer sequences for testing the presence of barley B-class genes.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvGAPDH</i>	<i>HORVU7Hr1G074690</i>	GTGAGGCTGGTGTGATTACG	TGGTGCAGCTAGCATTGAGAC
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	GCCCCAAGATACGAACCCTTCC	GGTGGTGCAAACCTTACAGTGAGG
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	GTGTGTCCGATTGATCTACTCC	CAAGATCCCCTCTATCCTGTATCG
<i>HvMADS4</i>	<i>HORVU1Hr1G063620</i>	ACACTCCACAGAGACAAGGG	GTTGGGAAACAACCTAGCACTGG
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	CCCTCGTCCACTTTCTTCTCC	CGACACAAGGCAAGGTAAACG

Supplementary Table 3C. PCR primer sequences for testing the presence of genes upstream and downstream from *HvMADS16/HORVU7Hr1G091210* (bold) on chromosome 7H.

Gene ID	Forward primer	Reverse primer
<i>HORVU7Hr1G091170</i>	AGATATGAGATTGACAAGGTCCTGC	GGTAGAAGAATAAGGTTCCACTTGC
<i>HORVU7Hr1G091180</i>	GACTGAACTGAGCTGATGG	CAGGTGTGATACGAGTTGAAGG
<i>HORVU7Hr1G091180</i>	GAATGAGATGTTGTCGACTTGTGCG	CACATTGTAATCCCTTCGTCTCG
<i>HORVU7Hr1G091200</i>	AGTAGAAAGGGGAAATTTAGTAGCG	TGAGCATGATGATGTTGAAGGAG
<i>HORVU7Hr1G091210</i>	CCCTCGTCCACTTTCTTCTCC	CAATACACAGTCGAGCACTACG
<i>HORVU7Hr1G091220</i>	GGAATCGGAGTAGACGCAAGC	GTGGCTAACGTCGATGGACC
<i>HORVU7Hr1G091220</i>	AAACCTTGGGTCGAGTAAAGCG	TGTTGGAACAGCACCTAACACC
<i>HORVU7Hr1G091230</i>	CAGTTATTGACAGACAGAGCTCC	GGTATGGGTACAGGATGTCATC
<i>HORVU7Hr1G091240</i>	GGCCATCAGGTCCTGTTTCAG	TTTCAGCTCCGTTGTAGTGTGG
<i>HORVU7Hr1G091250</i>	AGTGAGACAATCGACAGTAGCG	TGACAGTTGAGTGAGAGTGAGC

Supplementary Table 3D. qRT-PCR primer sequences.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvGAPDH</i>	<i>HORVU7Hr1G074690</i>	GTGAGGCTGGTGTGATTACG	TGGTGCAGCTAGCATTGAGAC
<i>HvCYCLO</i>	<i>HORVU6Hr1G012570</i>	CCTGTCGTGTCGTGGTCTAAA	ACGCAGATCCAGCAGCCTAAAG
<i>HvTUB</i>	<i>HORVU1Hr1G081280</i>	AGTGTCTGTCCACCCACTC	AGCATGAAGTGGATCCTTGG
<i>HvHSP70</i>	<i>HORVU5Hr1G113180</i>	CGACCAGGGCAACCGCACAC	ACGGTGTGATGGGGTTCATG
<i>HvMADS14</i>	<i>HORVU5Hr1G095630</i>	CAGCGCGGGCAGGCGAGAG	CCAGGCTGGCCGCTGCAAC
<i>HvMADS15</i>	<i>HORVU2Hr1G063800</i>	ATATGCTACCGCCATGGAT	ATACAGCGAACCAGCATTCC
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	CCAGCATGATATCGCCTTG	TCGAGCCAGTGGTGGATAA
<i>HvMADS4 & HvKinase*</i>		ATGCCAAGATGTTCTGGTC	TTTGGCACCTTAGCCATCAT
<i>HvKinase*</i>	<i>HORVU1Hr1G063610</i>	TTCTCGTGTGTTCTGGTCA	ATGCCAAGATGTTCTGGTC
<i>HvMADS4*</i>	<i>HORVU1Hr1G063620</i>	ATGGAGCTCGGGTACCATC	CCTGCAGGTAGATGGAGCA
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	CCCAGGAGGCATACAAGAATCTGC	GCGGAAGGCGTACATGTCAGC
<i>HvMADS3</i>	<i>HORVU3Hr1G026650</i>	GCAGCAGCAGCATTACTCC	ACACATGCACGCGACAGTA
<i>HvMADS58</i>	<i>HORVU1Hr1G029220</i>	ATCATGCAGCAGCCTCAGT	GGTGTGGCCAAGCCTTAAT
<i>HvMADS13</i>	<i>HORVU1Hr1G023620</i>	TCAGCTGAACCTAGGCTGC	TTTGACAGGAATAGTTGAGTACTGGT
<i>HvMADS7</i>	<i>HORVU7Hr1G054220</i>	ACCCTCTGAGTCCCTGAA	ACGAAAGTTGCACGCAAAA
<i>HvMADS8</i>	<i>HORVU5Hr1G076400</i>	CTCAGGAGCAGATAAACAACG	GTACGCGAACCGGTTACTA
<i>HvDL</i>	<i>HORVU4Hr1G067780</i>	CCATGCAAGAGGCTGATGGACACG	GCGGCTGGTTCCTCTGCAGTCAG
<i>HvMEL1</i>	<i>HORVU5Hr1G107020</i>	TCAAGGACGTGATGTTCTATTGC	ATCCTGGCAACTTCAAAATGGT
<i>HvMSP1</i>	<i>HORVU6Hr1G033670</i>	AAGCGACACAGATGATGGAAAG	TGAAGCTTAATCTGCCGATTTC

* *HvMADS4* (*HORVU1Hr1G063620*) completely overlaps with *HvKinase* (*HORVU1Hr1G063610*). *HvMADS4* transcript abundance has thus been obtained by subtracting *HvKinase* values (primers specific to *HvKinase*) from *HvMADS4 & HvKinase* (primers designed to amplify both genes).

Supplementary Table 3E. Primer sequence for cloning of *in situ hybridization* antisense (AS) and sense (S) probes. The T7 promoter sequence is underlined.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvMADS16</i> (AS)	<i>HORVU7Hr1G091210</i>	GCAAAGGATGGGTGAAGATCTGG	<u>TAATACGACTCACTATAGGGG</u> CGGAAGGCCTA CATGTCAGC
<i>HvMADS16</i> (S)	<i>HORVU7Hr1G091210</i>	<u>TAATACGACTCACTATAGGGG</u> GCAAAGGATGGGT GAAGATCTGG	GCGGAAGGCGTACATGTCAGC
<i>HvMADS3</i> (AS)	<i>HORVU3Hr1G026650</i>	AGGTTAACATGCAGCAGCAGC	<u>TAATACGACTCACTATAGGGG</u> GGAAGATATGC AACGCGATGG
<i>HvMADS3</i> (S)	<i>HORVU3Hr1G026650</i>	<u>TAATACGACTCACTATAGGGG</u> AGGTTAACATGCA GCAGCAGC	GGAAGATATGCAACGCGATGG
<i>HvMADS13</i> (AS)	<i>HORVU1Hr1G023620</i>	ATCAGGGCCAGGAAGAATGAGC	<u>TAATACGACTCACTATAGGGG</u> CAGGTTGACTAGA ACTGATGAGCC
<i>HvMADS13</i> (S)	<i>HORVU1Hr1G023620</i>	<u>TAATACGACTCACTATAGGGG</u> ATCAGGGCCAGGA AGAATGAGC	CAGGTTGACTAGAACTGATGAGCC
<i>HvDL</i> (AS)	<i>HORVU4Hr1G067780</i>	TTCCATGCAAGAGGCTGATGG	<u>TAATACGACTCACTATAGGGG</u> CTTTGATACGCT GTATTCCTCC
<i>HvDL</i> (S)	<i>HORVU4Hr1G067780</i>	<u>TAATACGACTCACTATAGGGG</u> TTCCATGCAAGAG GCTGATGG	GCTTTGATACGCTGTATTCCTCC

Supplementary Table 3F. Primer sequences for BiFC cloning. HindIII restriction site is underlined in all forward primers, XmaI restriction site is underlined in all reverse primers.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	TTCA <u>AAGCTT</u> ACATGGGGCGCGGGAAGATCG	<u>ACCCGGG</u> ATCTAGGTGCTCCTCTGCAGATTGGG
<i>HvMADS4</i>	<i>HORVU1Hr1G063620</i>	TTCA <u>AAGCTT</u> ACATGGGGCGCGCAAGATCG	<u>ACCCGGG</u> ATCTACTTGTCTTCTGCAAGTTGGGGTG
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	TTCA <u>AAGCTT</u> ACATGGGGCGGGGAAGATCG	<u>ACCCGGG</u> TATTATCCGAGGCGCAGGTCGTG
Δ <i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	TTCA <u>AAGCTT</u> ACATGGGGCGGGGAAGATCG	<u>ACCCGGG</u> TACTTGATGTCGGTGCCGGTGC

Supplementary Table 3G. Primer sequences for *HvMADS16* (*HORVU7Hr1G091210*) CRISPR knockout. gRNA sequence is underlined.

#	Gene ID	Forward primer	Reverse primer
gRNA1	<i>HORVU7Hr1G091210</i>	<u>GATCGAGAACGCCACCAACG</u> TTTTAGAGCTAGAAAT	<u>GTTGGTGGCGTTCTCGAT</u> CCGGCAGCCAAGCCAGCA
gRNA2	<i>HORVU7Hr1G091210</i>	<u>CGGTCCGGGATCATGAAGAG</u> TTTTAGAGCTAGAAAT	<u>TCTTCATGATCCCGGACCG</u> CAACACAAGCGGCAGC

Appendix B



Supplementary material to Chapter 4

Supplementary Figure S1. Spike morphology of wild type and *mov2.g* plants.

Supplementary Figure S2. Plant morphology of F₁ plants from the *mov2.g* x Morex bi-parental population.

Supplementary Figure S3. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type and *mov2.g* inflorescences.

Supplementary Figure S4. *In silico* prediction of zinc finger TFBSs in the putative promoter (3 Kb) of barley B-class genes.

Supplementary Figure S5. Barley protoplast transfection efficiency and vector maps.

Supplementary Figure S6. Vector map for generation of CRISPR/Cas9 knockout for *HvSL1*.

Supplementary Table S1. Germination ratio of heterozygous plant material segregating for the *mov2.g* phenotype

Supplementary Table S2. Segregation ratios of *mov2.g* in growing material through genotyping by copy number of *HvSL1* and among F₂ lines from the *mov2.g* (cv. Steptoe) x Morex bi-parental population.

Supplementary Table S3. Annotated genes present in the mapped critical interval for *mov2* between flanking markers chr3H_9748112 and chr3H_10289104.

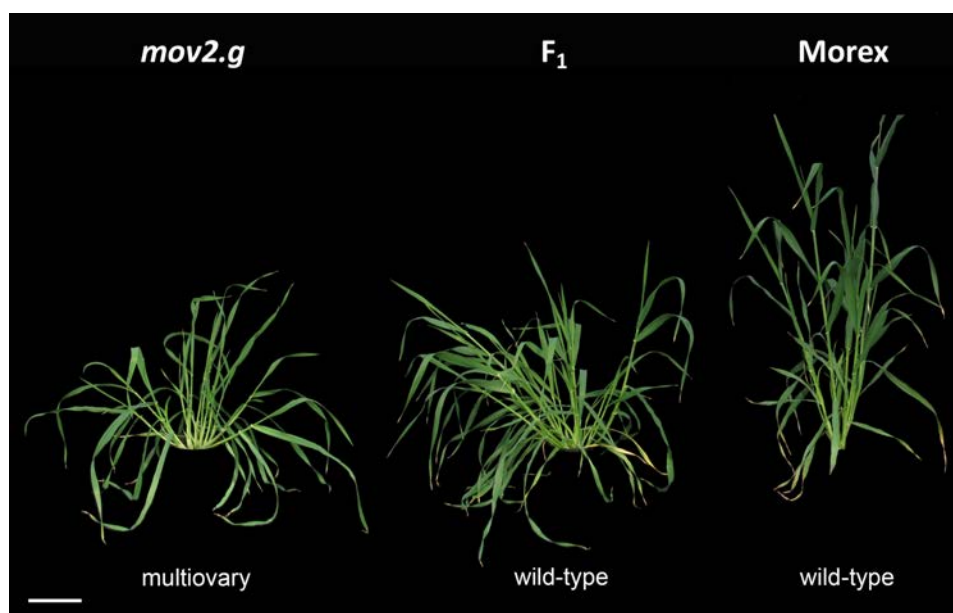
Supplementary Table S4. Sequence of KASP™ markers on chromosomes 3H used to map the *mov2* locus.

Supplementary Table S5. Primer sequences used in this study.

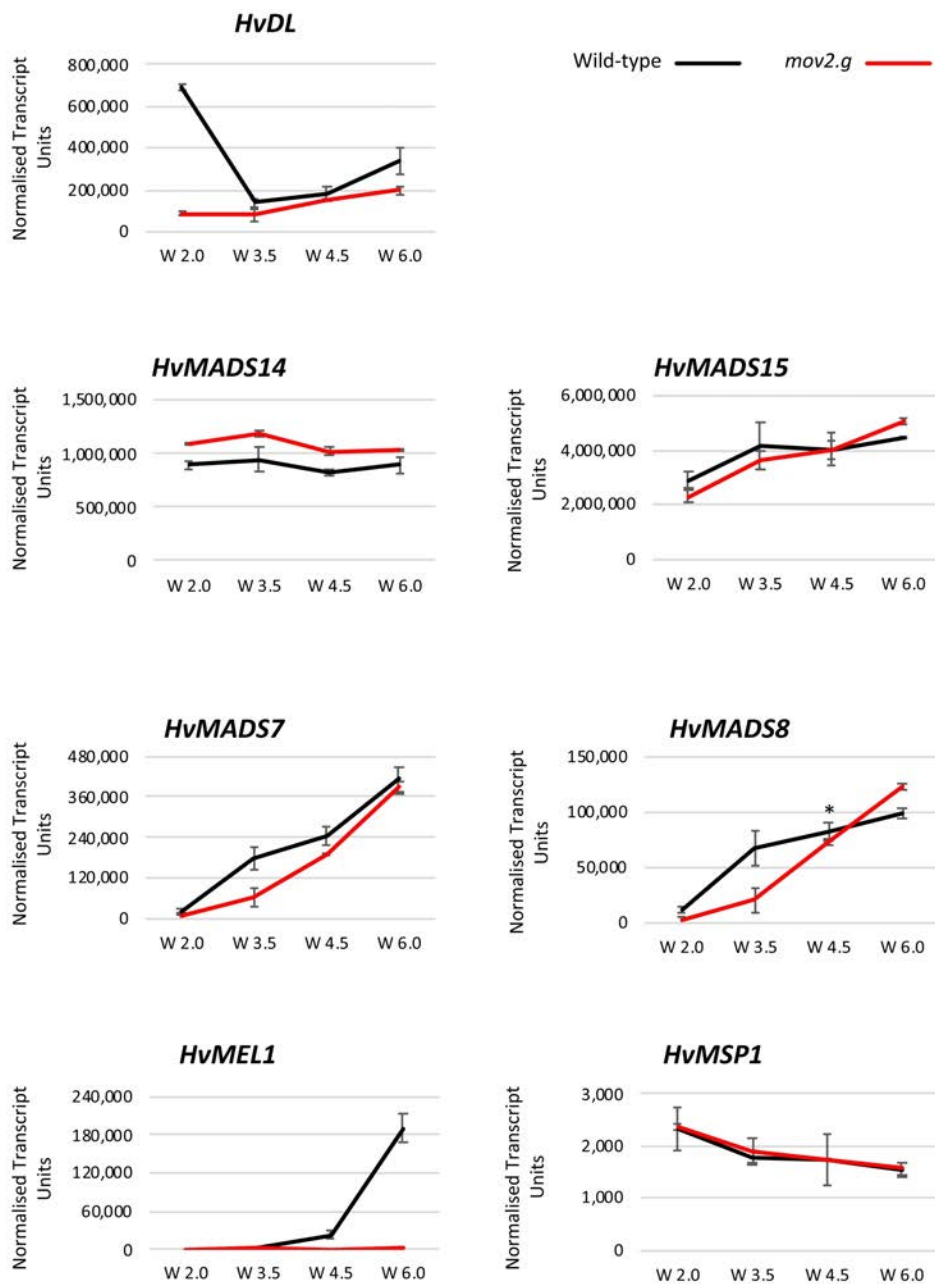
Supplementary Figure S1. Spike morphology of (A) wild type and (B - D) *mov2.g* plants. Scale bars: 1 cm.



Supplementary Figure S2. Plant morphology of F_1 (*mov2.g* x Morex) plants from the *mov2.g* (cv. Steptoe) x Morex bi-parental population. Scale bar: 20 cm. Floret phenotype is indicated.

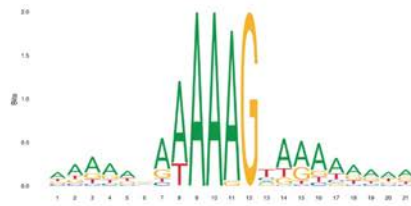


Supplementary Figure S3. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type (black) and *mov2.g* (red) inflorescences. Developmental stages refer to W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development). Transcript abundance is shown for *HvDL* (*HORVU4Hr1G067780*); A-class genes *HvMADS14* (*HORVU5Hr1G095630*), *HvMADS15* (*HORVU2Hr1G063800*); E-class genes *HvMADS7* (*HORVU5Hr1G076400*), *HvMADS8* (*HORVU7Hr1G054220*), as well as *HvMEL1* (*HORVU5Hr1G107020*) and *HvMSP1* (*HORVU6Hr1G033670*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**), and ≤ 0.001 (***) are shown for differences between wild type and *mov2.g*. For each sample *n* = 3 independent biological replicates.



Supplementary Figure S4. *In silico* prediction of zinc finger TFBSs in the putative promoter (3 Kb) of barley B-class genes. **(A)** *Arabidopsis thaliana* C2H2 zinc finger input motif (Matrix ID MA1277.1) used to predict Transcription Factor Binding Sites (TFBSs) in barley B-class genes. **(B)** Predicted TFBSs in putative promoter (3 Kb) of *HvMADS2* (HORVU3Hr1G091000), *HvMADS4* (HORVU1Hr1G063620) and *HvMADS16* (HORVU7Hr1G091210) using a 75 % similarity threshold to input motif. Motifs identified with an 80 % similarity threshold are indicated with *. TFBSs were predicted using JASPAR 7th release (<http://jaspar.genereg.net>). **(C)** Graphical representation of TFBSs position in the 3 Kb putative promoter. Motifs identified with an 80 % similarity threshold are indicated with *.

A. C2H2 zinc finger input motif



B. Barley TFBSs prediction

pHvMADS2

#	Score	Start	End	Strand	Sequence
1	9.16	1,510	1,530	-	AAATATGAAAACCTGAATAAT
2	8.91	2,284	2,304	-	GCCTATAAAAAGCGGAGAAAT

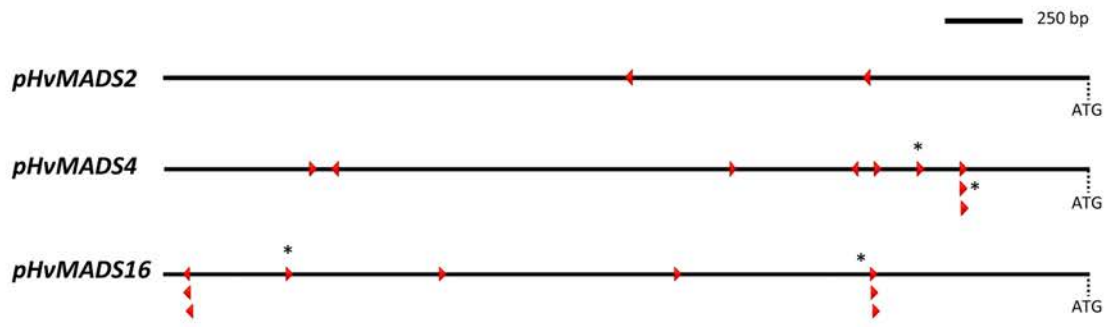
pHvMADS4

#	Score	Start	End	Strand	Sequence
* 1	13.39	2,570	2,590	+	GAGAGAAAAAGAGGAGGAAA
* 2	12.57	2,424	2,444	+	ACGAACAAAAGGCAGACAAA
7	10.87	2,280	2,300	+	TGAATGCAAAAGCAATGCAAA
3	10.76	388	408	+	TCACTAAAAAGTTGCATACA
6	10.44	2,205	2,225	-	GCCGAATAAAAGTAAAAATGT
9	9.58	2,572	2,592	+	GAGAAAAAGAGGAGGAAAAGA
5	9.44	1,796	1,816	+	CAAAGAGAAGAGGAAAAGATT
8	9.01	2,569	2,589	+	AGAGAGAAAAAGAGGAGGAA
4	8.95	461	481	-	TAGCAAAAACGAAGAAGATG

pHvMADS16

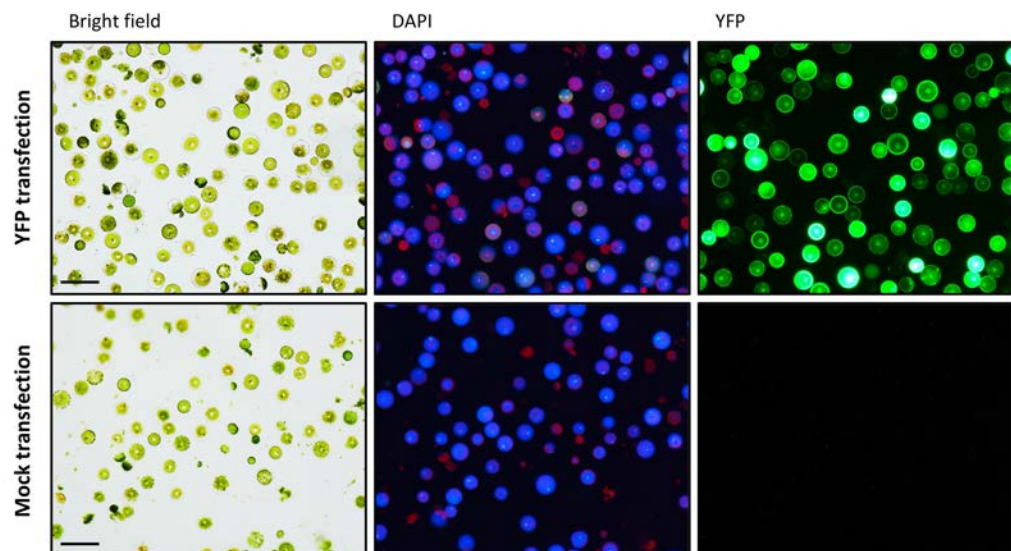
#	Score	Start	End	Strand	Sequence
* 1	12.66	378	398	+	TAGATAAAAAGTAAGGTTCA
* 2	11.90	2,279	2,299	+	TCGACACAAAAGAAGAAAAGA
3	10.40	2,287	2,307	+	AAAGAAGAAAAGAGAGGGGGA
4	9.73	877	897	+	GGATGCAAAAAGTTATGAAGC
5	9.24	1,642	1,662	+	AAAATCTTAAGGTTAAAGAAT
6	9.07	52	72	-	ATAGTCAAAAATATAAAAAC
7	8.96	44	64	-	AAATATAAAAACACTAACAATT
8	8.85	2,282	2,302	+	ACACAAAAGAAGAAAAGAGAG
9	8.66	43	63	-	AATATAAAAACACTAACAATTC

C. Position of TFBSs

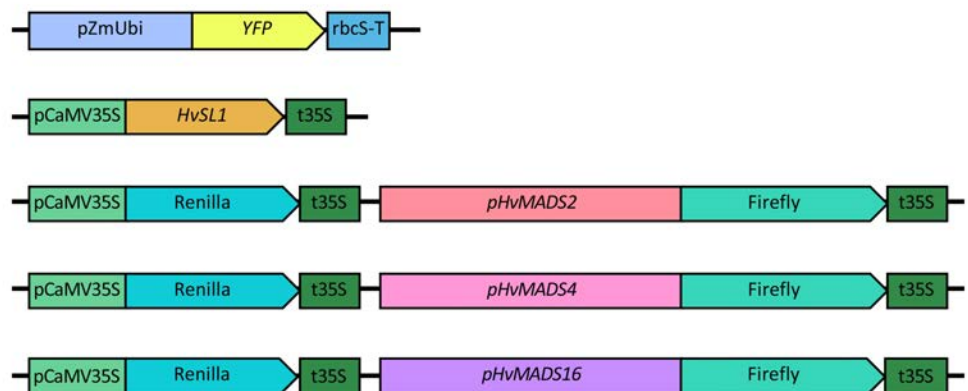


Supplementary Figure S5. Barley protoplast transfecion efficiency and vector maps. **(A)** For each sample, efficiency of protoplast transfecion was calculated by averaging the number of protoplasts expressing YFP in three representative images. **(B)** Schematics of constructs used in the Dual Luciferase assay (DLR). A YFP-containing construct was used as control for protoplast efficiency. Constitutive expression of YFP was driven by the maize (*Zea mays*) ubiquitin promoter (pZmUbi). Constitutive expression of *HvSL1* (*HORVU3Hr1G003740*) was obtained by using the Cauliflower Mosaic Virus 35S promoter (pCaMV35S). Putative promoters (3 Kb) of B-class genes (*pHvMADS2*, *pHvMADS4* and *pHvMADS16*) were cloned in the pGreenII-0800 LUC backbone which contains Firefly (*Photinus pyralis*) and Renilla luciferase (*Renilla reniformis*). Constitutive expression of Renilla luciferase acts as internal control for the DLR assay. *rbcS-T* represents the terminator of the ribulose-1,5-bisphosphate carboxylase gene from *Chrysanthemum morifolium*; t35S indicates the CaMV35S terminator. Scaler bars: 100 μ m.

A. Protoplast transfecion efficiency

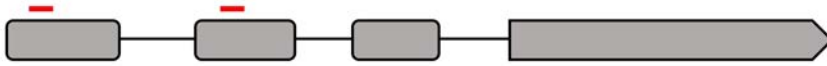


B. DLR assay constructs



Supplementary Figure S6. Vector map for generation of CRISPR/Cas9 knockout for *HvSL1*. **(A)** Position of the guideRNA (red); two different gRNA designed to target *HvSL1* (*HORVU3Hr1G003740*) at +39 bp and +269 bp from the translational start site. **(B)** Details of the CRISPR/Cas9 vector. Vector includes a plant codon optimized Cas9 driven by the maize ubiquitin promoter (pZmUbi) and both gRNA (gRNA1 and gRNA2) driven by the rice small nuclear RNA promoters U6a (pOsU6a) and U6b (pOsU6b), respectively. NosT stands for the *Agrobacterium* nopaline synthase terminator.

A. gRNA target sites



B. CRISPR/Cas9 vector



Supplementary Table S1. Germination ratios of heterozygous plant material segregating for the *mov2.g* phenotype.

Group #	Germinated seeds	Total seeds	Germination ratio
1	80	116	69 %
2	352	680	52 %
3	124	210	59 %
4	94	250	38 %
5	334	681	49 %

Supplementary Table S2A. Segregation ratios of *mov2.g* in growing material through genotyping by copy number of *HvSL1*.

Group #	Mutant (-/-)	Heterozygote (A/-)	Wild type (A/A)	Total
1	20 (6 %)	211 (60 %)	121(34 %)	352
2	13 (10 %)	68 (54 %)	44 (35 %)	125
3	6 (6 %)	51 (54 %)	37 (39 %)	94
4	24 (7 %)	189 (58 %)	112 (34 %)	325

Supplementary Table S2B. Observed segregation ratios of *mov2.g* phenotype among F₂ lines of the *mov2.g* (cv. Steptoe) x Morex bi-parental population.

F ₁ line	Observed <i>mov2.g</i> phenotype in F ₂	Observed wild-type phenotype in F ₂	Total	χ^2 value	P-value	Output
V144_XE8-2	18 (20 %)	70 (80%)	88	0.970	0.325	ACCEPT H ₀
V144_XE8-3	19 (22 %)	68 (78 %)	87	0.464	0.496	ACCEPT H ₀
V144_XE8-5	12 (14 %)	76 (86 %)	88	6.061	0.014	REJECT H ₀
V144_XE8-8	5 (6 %)	83 (94 %)	88	17.515	0.000	REJECT H ₀

H₀ = The observed phenotypes segregate with a 3:1 ratio

H₁ = Not H₀; observed phenotypes segregate differently from 3:1 ratio

Degrees of freedom (DF) = 1

Supplementary Table S3. Annotated genes present in the mapped *mov2* critical interval between flanking markers chr3H_9748112 and chr3H_10289104. Annotations and genomic coordinates are based on the Morex reference assembly Hv_IBSC_PGSB_v2, gene order is based on Morex Scaffold_1432 (Dr. Martin Mascher, IPK Gatersleben, Germany). *HvSL1* (*HORVU3Hr1G003740*) is indicated in bold, genes tested by PCR are indicated with *.

Gene	Start	End	Strand	Annotation
<i>HORVU3Hr1G003690</i> *	9,737,630	9,749,764	+	N.A
<i>HORVU3Hr1G003700</i>	9,753,370	9,754,129	-	Undescribed protein
<i>HORVU3Hr1G003710</i>	9,756,062	9,756,300	-	Undescribed protein
<i>HORVU3Hr1G003720</i>	9,757,456	9,764,144	-	Unknown function
<i>HORVU3Hr1G003730</i>	9,774,072	9,774,466	-	Undescribed protein
<i>HORVU3Hr1G003760</i> *	9,942,084	9,945,779	+	Protein of unknown function (DUF1666)
<i>HORVU3Hr1G003750</i>	9,937,908	9,938,930	-	Undescribed protein
<i>HORVU3Hr1G003740</i> *	9,908,524	9,910,297	-	Zinc finger protein 6
<i>HORVU3Hr1G003780</i>	10,031,768	10,032,424	-	Undescribed protein
<i>HORVU3Hr1G003770</i>	10,030,637	10,031,653	-	Undescribed protein
<i>HORVU3Hr1G003800</i>	10,168,437	10,169,618	-	Undescribed protein
<i>HORVU3Hr1G003810</i> *	10,169,902	10,172,670	+	Disease resistance protein
<i>HORVU3Hr1G003820</i> *	10,173,750	10,180,522	+	Synaptotagmin A
<i>HORVU3Hr1G003830</i>	10,180,811	10,181,062	-	Undescribed protein
<i>HORVU3Hr1G003840</i>	10,200,846	10,209,364	+	BnaA07g10090D protein
<i>HORVU3Hr1G003850</i>	10208187	10213828	-	Nuclease S1
<i>HORVU3Hr1G003870</i>	10,259,028	10,259,724	+	Endonuclease 2
<i>HORVU3Hr1G003790</i>	10,058,110	10,058,378	+	Undescribed protein
<i>HORVU3Hr1G003860</i>	10,224,976	10,227,068	+	Endonuclease 2
<i>HORVU3Hr1G003880</i>	10,288,324	10,289,509	-	Chromosome 3B, genomic scaffold

Supplementary Table S4. Sequence of KASP™ markers on chromosomes 3H used to map the *mov2* locus. Flanking markers are indicated in bold. Marker order is based on the genetic map.

Marker name	Forward primer Allele 1	Forward primer Allele 2	Reverse Common primer	SNP
chr3H_28805649 ¹	GAAGGTGACCAAGTTCATGCTCGACGGTTCATTCTGCAC	GAAGTCGGAGTCAACGGATTGCTCGACGGTTCATTCTGCAA	TCTTGATGCTGACATGATATGTTCTTGAT	[T/A]
chr3H_1006543 ^{1,2}	GAAGGTGACCAAGTTCATGCTAAGCTTGATTCCACATGACCAATTT	GAAGTCGGAGTCAACGGATTCTAAGCTTGATTCCACATGACCAATTT	ACTCTCCCATGGCCGACCTT	[G/A]
chr3H_1367441 ²	GAAGGTGACCAAGTTCATGCTCCGGATTCTCAAGAGCTCT	GAAGTCGGAGTCAACGGATTCTCCGGATTCTCAAGAGCTCC	GCCTCGGAGTGGGGAGGTT	[T/C]
chr3H_3865263 ²	GAAGGTGACCAAGTTCATGCTGAAACCATATACCATAGCAGCAGCAA	GAAGTCGGAGTCAACGGATTAAACCATATACCATAGCAGCAGCAG	CAAATGCTTTACTATAACGGCGGCAT	[A/G]
chr3H_7767159 ³	GAAGGTGACCAAGTTCATGCTCGGCGAGGACGAGACA	GAAGTCGGAGTCAACGGATTGATCGGCGAGGACGAGACG	CCCACCGTAGACTCCGATA	[A/G]
chr3H_7767871 ^{3,4}	GAAGGTGACCAAGTTCATGCTAAGCTTGATGCTCTTACCAATGGT	GAAGTCGGAGTCAACGGATTGATGCTCTTACCAATGGC	TTCAGCTGAGCTTGAATGGGACTT	[A/G]
chr3H_8709612 ³	GAAGGTGACCAAGTTCATGCTGTGATGGACCGCCCTCGT	GAAGTCGGAGTCAACGGATTGATGGACCGCCCTCGC	CTCCCTCCCAATGCACACCGAT	[A/G]
chr3H_8787424 ³	GAAGGTGACCAAGTTCATGCTATGAAAGTTTCTAGATGATGCAAGGC	GAAGTCGGAGTCAACGGATTGATGAAAGTTTCTAGATGATGCAAGGA	CATTTTCCAAGTTTCTGCCACAGTTTT	[G/T]
chr3H_744266 ³	GAAGGTGACCAAGTTCATGCTGTAACAAATATCTACTCTGC	GAAGTCGGAGTCAACGGATTGCTGTAACAAATATCTACTCTGT	GGAATTCGACCCCTGACCATGTTA	[C/T]
chr3H_8969648 ³	GAAGGTGACCAAGTTCATGCTCCCGTCTCCGTGCGCGG	GAAGTCGGAGTCAACGGATTCCCGTCTCCGTGCGCGC	CTCCAGATGCTGCTGCTCGAA	[C/G]
chr3H_9095799 ³	GAAGGTGACCAAGTTCATGCTACGCGCAGGTGATGCACCG	GAAGTCGGAGTCAACGGATTGACGCGCAGGTGATGCACCA	GCCCTCCACGGCTCTCTT	[G/A]
chr3H_9356192 ⁴	GAAGGTGACCAAGTTCATGCTGTTGATCGATCGACGCCG	GAAGTCGGAGTCAACGGATTGCTGTTGATCGATCGACGCCA	GCCCTTTCTCCGGGCATCAT	[C/T]
chr3H_9347808 ⁴	GAAGGTGACCAAGTTCATGCTATAGAGAATGGAGTAGTCTATAC	GAAGTCGGAGTCAACGGATTGCTATAGAGAATGGAGTAGTCTATAT	CACAACAGGTAGGATAGTAGAAACTATA	[G/A]
chr3H_9748112 ⁴	GAAGGTGACCAAGTTCATGCTATCAAACGCAATCAAGGTACTTTAC	GAAGTCGGAGTCAACGGATTATCAAACGCAATCAAGGTACTTTAG	CCCTGAAATGAATAACCTTTTTTAGGGAA	[C/G]
chr3H_10289104 ⁴	GAAGGTGACCAAGTTCATGCTGGCGCGGAGGCTCTG	GAAGTCGGAGTCAACGGATTGGCGCGGAGGCTCTC	CCCCGTCAAAGCTCCCAAGAA	[G/C]
chr3H_10601795 ⁴	GAAGGTGACCAAGTTCATGCTGTTGGGGCTTACTATGTGCCA	GAAGTCGGAGTCAACGGATTGTTGGGGCTTACTATGTGCCG	CCACACAGTGAACACRTTCAAGGAT	[A/G]
chr3H_10797294 ⁴	GAAGGTGACCAAGTTCATGCTATGTTGGGACGCTTCTCCTC	GAAGTCGGAGTCAACGGATTGATGTTGGGACGCTTCTCCTA	TTTGAGCAGTAGCAGTGCAGCCAT	[C/A]
chr3H_11039299 ³	GAAGGTGACCAAGTTCATGCTCCAGGTCTTCGAGTGCCCC	GAAGTCGGAGTCAACGGATTCCAGGTCTTCGAGTGCCCCG	GCCCGCAGGACTTGCAGGTTT	[C/G]
chr3H_11335959 ³	GAAGGTGACCAAGTTCATGCTCCGGCTACGAGTACCCC	GAAGTCGGAGTCAACGGATTCCGGCTACGAGTACCCC	GTTACCGGGGCTGCTGGTT	[G/C]
chr3H_11617707 ³	GAAGGTGACCAAGTTCATGCTACTTGCTCAATAATCACAGCTCTC	GAAGTCGGAGTCAACGGATTACTTGCTCAATAATCACAGCTCTA	AGGGTATTGAAATGATGATGGATCTCAT	[G/T]
chr3H_11702941 ³	GAAGGTGACCAAGTTCATGCTGACAGCCGTTTCTGCGCCACA	GAAGTCGGAGTCAACGGATTGACAGCCGTTTCTGCGCCACG	CCGGTTGTCTACGCTATATGAT	[A/G]
chr3H_17951104 ²	GAAGGTGACCAAGTTCATGCTACTTGTCTGTTCTGTCTCCTC	GAAGTCGGAGTCAACGGATTACTTGTCTGTTCTGTCTCCTT	TTGTACGGGGGAGGCACTAA	[G/A]
chr3H_28161638 ²	GAAGGTGACCAAGTTCATGCTCTTCCGCCCTGAGTTTG	GAAGTCGGAGTCAACGGATTCTTCCGCCCTGAGTTTA	GAAGTCGACGCTGTTGAAGTCTT	[C/T]
chr3H_28805649 ²	GAAGGTGACCAAGTTCATGCTCGACGGTTCATTCTGCAC	GAAGTCGGAGTCAACGGATTGCTCGACGGTTCATTCTGCAA	TCTTGATGCTGACATGATATGTTCTTGAT	[G/T]

¹ markers used to confirm heterozygosity of F₁ plants

² markers used to confirm *mov2* position

³ markers used for mapping with F₂ segregants

⁴ markers used for mapping with F₃ recombinants

Supplementary Table S5A. Primer sequences and Taqman probes used for copy number analysis to genotype *mov2.g* plants. Fluorophore at 5' and 3' quencher are indicated for each probe.

Gene name	Gene ID	Forward primer	Taqman probe
<i>HvCO-like</i>	<i>HORVU6Hr1G072620</i>	TGCTAACCGTGTGGCATCAC GGTACATAGTCTGCTGCATCTG	[HEX] CATGAGCGTGTGCGTGTCTGCG [BHQ1]
<i>HvSL1</i>	<i>HORVU3Hr1G003740</i>	GGAGGAGGAGGATTCAGGGGAGG GCGTCGTGCTGTAGAGGTAGTGG	[FAM] TCGGAGACCCAAGCCACCACCCATT [BHQ1]

Supplementary Table S5B. PCR primer sequences for testing the presence of genes upstream and downstream from *HvSL1/HORVU3Hr1G003740* (bold) on chromosome 3H.

Gene ID	Forward primer	Reverse primer
<i>HORVU3Hr1G003690</i>	GTTACTTTACCCITTCGATGTTCC	ACGAAGTAGTGCGTCCCGAAG
<i>HORVU3Hr1G003760</i>	CTAGCTAGCGAGCGCATTATACC	TGGGAGGTCGATCTCATCAGTGC
<i>HORVU3Hr1G003740</i>	CCAAACCAACACTTTAAGACTGC	TCTTATGGGGAGTAAAAAGGACC
<i>HORVU3Hr1G003810</i>	GGGTGTAATCTGGTTGCTAATCC	ATCGACGTATCCTGATTCATTCC
<i>HORVU3Hr1G003820</i>	GTTTAGCAGTACGCATGAGACCC	AGCTAGTAGGGAGTCTTGGAGG

Supplementary Table S5C. qRT-PCR primer sequences.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvGAPDH</i>	<i>HORVU7Hr1G074690</i>	GTGAGGCTGGTCTGATTACG	TGGTGCAGCTAGCATTGAGAC
<i>HvCYCLO</i>	<i>HORVU6Hr1G012570</i>	CCTGTCTGTCTGCGGTCTAAA	ACGCAGATCCAGCAGCCTAAA
<i>HvTUB</i>	<i>HORVU1Hr1G081280</i>	AGTGCTCTGCCACCACTC	AGCATGAAGTGGATCCTTGG
<i>HvHSP70</i>	<i>HORVU5Hr1G113180</i>	CGACCAGGGCAACCGCACCCAC	ACGGTGTGATGGGGTTCATG
<i>HvSL1</i>	<i>HORVU3Hr1G003740</i>	GGAGGAGGAGGATTCAGGGGAGG	GCGTCGTGCTGTAGAGGTAGTGG
<i>HvMADS14</i>	<i>HORVU5Hr1G095630</i>	CAGCGGCGGCAGGCGAGAG	CCAGGCTGGCCGCTGCAAC
<i>HvMADS15</i>	<i>HORVU2Hr1G063800</i>	ATATGCCTACCGCATGGAT	ATACAGCGAACCGCATTCC
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	CCAGCATGATATCGCCTTG	TCGAGCCAGTGGTGGATAA
<i>HvMADS4 & HvKinase*</i>		ATGCCAAGATGTTCTGGTTC	TTTGGCACCCTAGCCATCAT
<i>HvKinase*</i>	<i>HORVU1Hr1G063610</i>	TTCTCGTGTGTTCTGGTCA	ATGCCAAGATGTTCTGGTTC
<i>HvMADS4*</i>	<i>HORVU1Hr1G063620</i>	ATGGAGCTCGGGTACCATC	CCTGCAGGTAGATGGAGCA
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	CCCAGGAGGCATACAAGAATCTGC	GCGGAAGCGTACATGTCAGC
<i>HvMADS3</i>	<i>HORVU3Hr1G026650</i>	GCAGCAGCAGCATTACTCC	ACACATGCACGCGACAGTA
<i>HvMADS58</i>	<i>HORVU1Hr1G029220</i>	ATCATGCAGCAGCCTCAGT	GGTGTGGCCAAGCCTTAAT
<i>HvMADS13</i>	<i>HORVU1Hr1G023620</i>	TCAGCTGAACCTAGGCTGC	TTTGACAGGAATAGTTGAGTACTGGT
<i>HvMADS7</i>	<i>HORVU7Hr1G054220</i>	ACCCTCTGAGTCCCTGAA	ACGAAAGTTGCACGCAAAA
<i>HvMADS8</i>	<i>HORVU5Hr1G076400</i>	CTCAGGAGCAGATAAACAACG	GTACGCGAACGCGGTACTA
<i>HvDL</i>	<i>HORVU4Hr1G067780</i>	CCATGCAAGAGGCTGATGGACACG	GCGGCTGGTTCCTCTGCAGTCAG
<i>HvOSH1</i>	<i>HORVU4Hr1G009730</i>	TGGGAGATGCACTACAAGTGG	ATGTAGAAGGCGGCTTAGG
<i>HvMEL1</i>	<i>HORVU5Hr1G107020</i>	TCAAGGACGTGATGTTCTATTGC	ATCCTGGCAACTCAAATGGT
<i>HvMSP1</i>	<i>HORVU6Hr1G033670</i>	AAGCGACACAGATGATGAAAG	TGAAGCTTAATCTGCCGATTTC

* *HvMADS4* (*HORVU1Hr1G063620*) completely overlaps with *HvKinase* (*HORVU1Hr1G063610*). *HvMADS4* transcript abundance has thus been obtained by subtracting *HvKinase* values (primers specific to *HvKinase*) from *HvMADS4* & *HvKinase* (primers designed to amplify both genes).

Supplementary Table S5D. Primer sequences for Dual Luciferase assay cloning. Apal restriction site is underlined in all forward primers, SacII restriction site is underlined in all reverse primers.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	AT <u>GGGCC</u> CACTGAACCAACCCGAAAGCATATG	TAC <u>CGCGG</u> ATCCAGGAAGGGTTCGTATCTTCG
<i>HvMADS4</i>	<i>HORVU1Hr1G063620</i>	AT <u>GGGCC</u> CAGTAAACGCTTCCACCCGACG	TAC <u>CGCGG</u> ATCTCAACAAGTACTCGCCGG
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	AT <u>GGGCC</u> CAATTCGGCAAAACCCAAATCTAAACC	TAC <u>CGCGG</u> ATCGATCCACCTCCTCCGTCC

Supplementary Table S5E. Primer sequences for *HvSL1* (*HORVU3Hr1G003740*) CRISPR knockout. gRNA sequence is underlined.

#	Gene ID	Forward primer	Reverse primer
gRNA1	<i>HORVU3Hr1G003740</i>	<u>CCTGCACTCGTACACCTTGC</u> GTTTTAGAGCTAGAAAT	<u>GCAAGGTGTACGAGTGCAGG</u> CGGCAGCCAAGCCAGCA
gRNA2	<i>HORVU3Hr1G003740</i>	<u>CTTGCTTGCCGTACTCCTCG</u> TTTTAGAGCTAGAAAT	<u>GAGGAGTACGGCAAGCAAGCA</u> ACACAAGCGGCAGC

Appendix C



Supplementary material to Chapter 5

Supplementary Figure S1. Examples of curled peduncle in *mov5.o* plants

Supplementary Figure S2. Representative images of selected 2-week old F₂ plants from the *mov5.o* (cv. Morex) x Steptoe bi-parental population showing stunted growth.

Supplementary Figure S3. Sequence alignment of FLO/LFY proteins from *Hordeum vulgare*, *Oryza sativa*, *Zea mays*, *Antirrhinum majus* and *Arabidopsis thaliana*.

Supplementary Figure S4. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type and *mov5.o* inflorescences.

Supplementary Table S1. Observed segregation ratios of the *mov5.o* phenotype among F₂ lines of the *mov5.o* (cv. Morex) x Steptoe bi-parental population.

Supplementary Table S2. Annotated genes present in the mapped critical interval for *mov5* between flanking markers chr2H_544617620 and chr2H_742203860.

Supplementary Table S3. Sequence of KASP™ markers used to map the *mov5* locus.

Supplementary Table S4. Primer sequences used in this study.

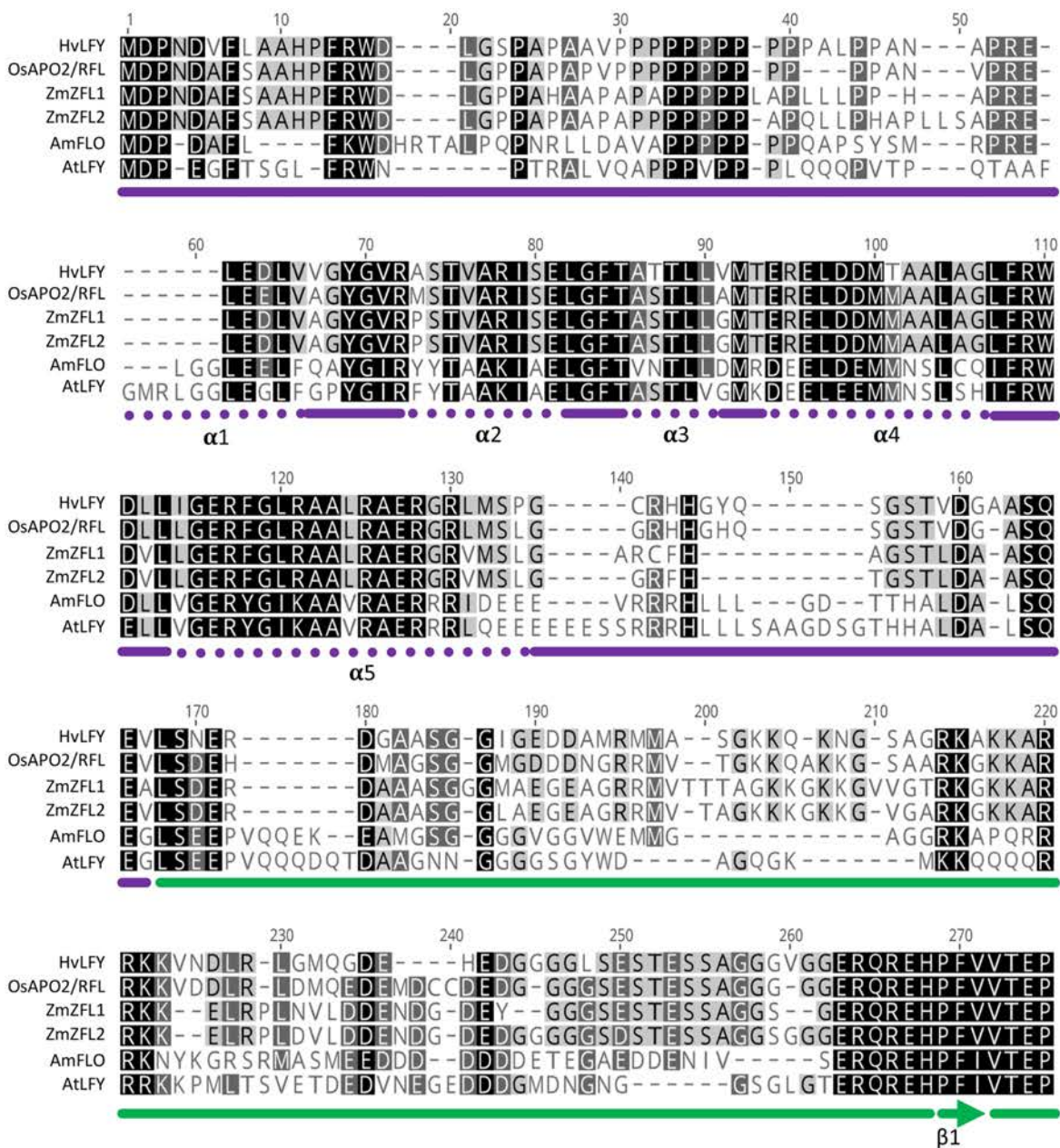
Supplementary Figures S1. Examples of curled peduncle in *mov5.o* plants. Scale bar: 1 cm.

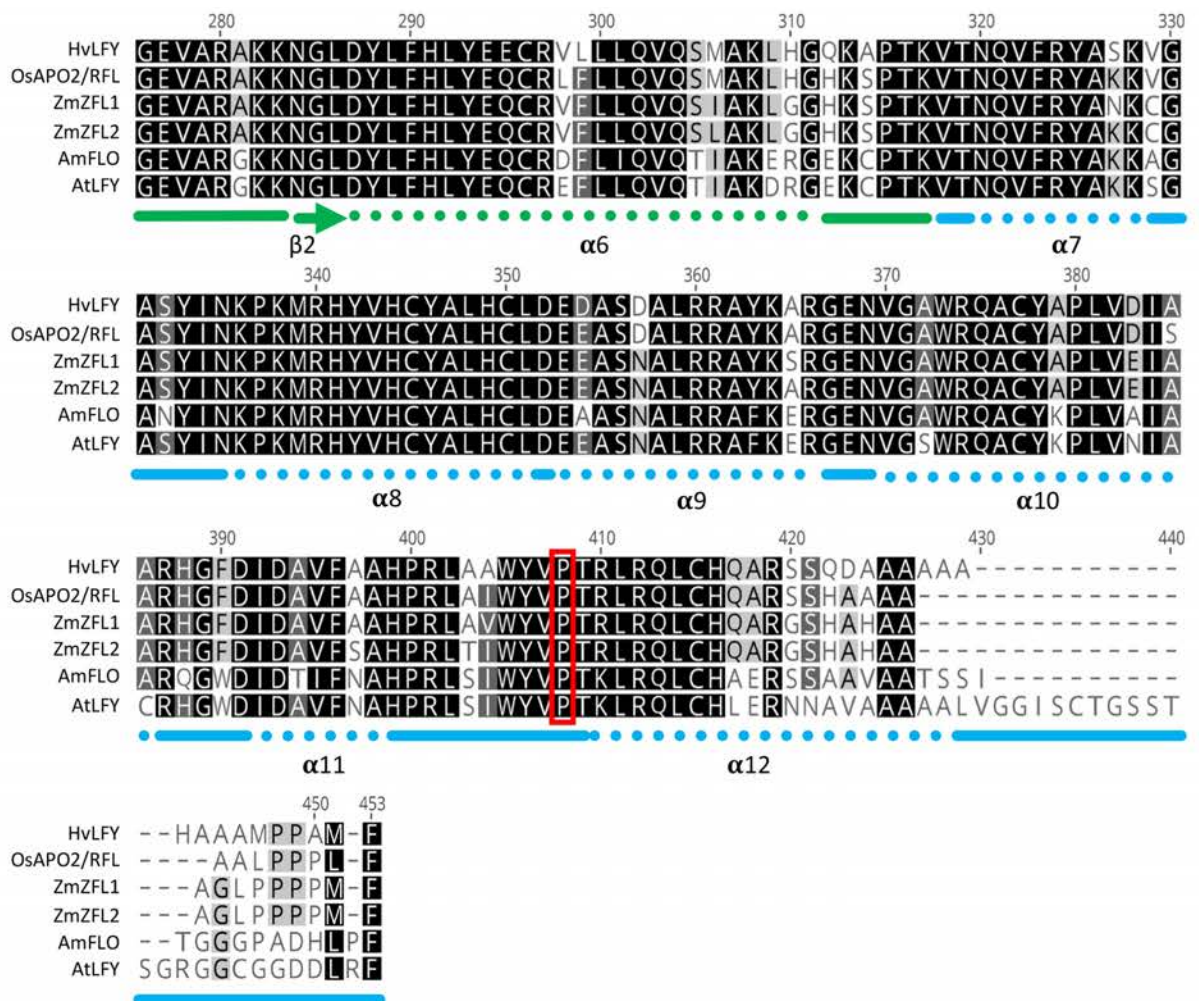


Supplementary Figures S2. Representative images of selected 2-week old F₂ plants from the *mov5.o* (cv. Morex) x Steptoe bi-parental population showing stunted growth. Scale bar: 1cm.



Supplementary Figure S3. Sequence alignment of FLO/LFY proteins from *Hordeum vulgare* (HvLFY - HORVU2Hr1G102590), *Oryza sativa* (OsAPO2/RFL - LOC_Os04g51000), *Zea mays* (ZmZFL1 - LOC542098 and ZmZFL2 - LOC103645994), *Antirrhinum majus* (AmFLO - Am06g19060) and *Arabidopsis thaliana* (AtLFY - AT5G61850). The protein residues encoded by the first (purple line), second (green line) and third (blue line) exons are indicated. Secondary structure elements are shown for the N-terminal domain (α 1-5) and for the DNA-binding domain (β 1-2, α 6-12), using dashed lines for α -helices and arrow for β -strands. The red box highlights the conserved Pro residue mutated in *mov5.o* plants. Figure contains information from Sayou *et al.* (2016) and Hamès *et al.* (2008).

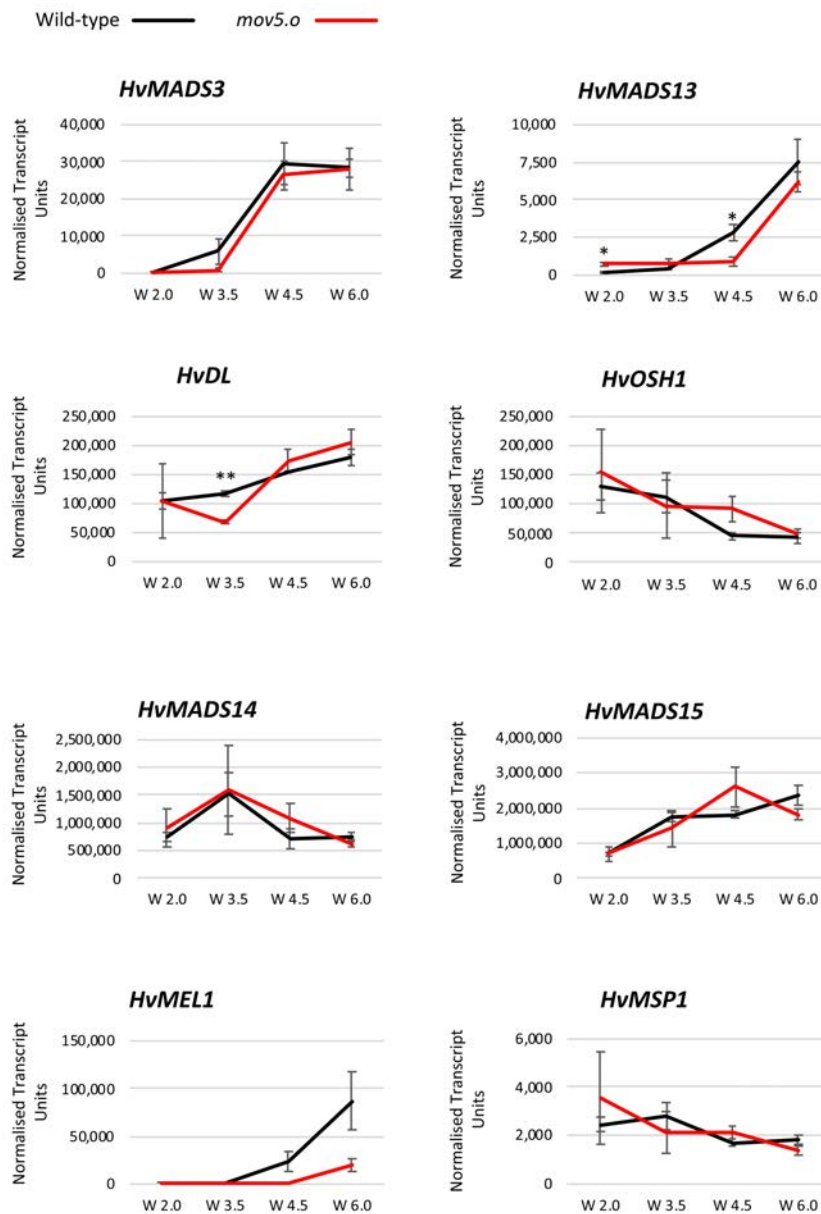




Sayou C, Nanao MH, Jamin M, Posé D, Thevenon E, Grégoire L, *et al.* A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nature Communications*. 2016;7:11222–12.

Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gérard F, *et al.* Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *The EMBO Journal*. 2008;27:2628–37.

Supplementary Figure S4. Transcript abundance of floral development genes, as detected by RT-qPCR in wild-type (black) and *mov5.o* (red) inflorescences. Developmental stages refer to W2.0 (double ridge), W3.5 (stamen primordia), W4.5 (carpel primordium) and W6.0 (stamen and carpel development). Transcript abundance is shown for C-class gene *HvMADS3* (*HORVU3Hr1G026650*); D-class gene *HvMADS13* (*HORVU1Hr1G023620*); *HvDL* (*HORVU4Hr1G067780*) and *HvOSH1* (*HORVU4Hr1G009730*); A-class genes *HvMADS14* (*HORVU5Hr1G095630*), *HvMADS15* (*HORVU2Hr1G063800*); as well as *HvMEL1* (*HORVU5Hr1G107020*) and *HvMSP1* (*HORVU6Hr1G033670*). Error bars represent \pm Standard Error. For each timepoint, two-tailed T-test *P*-values ≤ 0.05 (*), ≤ 0.005 (**), and ≤ 0.001 (***) are shown for differences between wild type and *mov5.o*. For each sample *n* = 3 independent biological replicates.



Supplementary Table S1. Observed segregation ratios of the *mov5.o* phenotype among F₂ lines of the *mov5.o* (cv. Morex) x Steptoe bi-parental population.

F ₁ line	Observed <i>mov5.o</i> phenotype in F ₂	Observed wild-type phenotype in F ₂	Total	χ^2 value	P-value	Output
V5317_S3-2	25 (28 %)	63 (72 %)	88	0.545	0.460	ACCEPT H ₀
V5317_S3-4	15 (17 %)	73 (83 %)	88	2.970	0.085	ACCEPT H ₀
V5317_S3-5	16 (21 %)	59 (79 %)	75	0.538	0.463	ACCEPT H ₀
V5317_S3-6	13 (18 %)	59 (82 %)	72	1.852	0.174	ACCEPT H ₀

H₀ = The observed phenotypes segregate with a 3:1 ratio

H₁ = Not H₀; observed phenotypes segregate differently from 3:1 ratio

Degrees of freedom (DF) = 1

Supplementary Table S2. Annotated genes present in the mapped critical interval for *mov5* between flanking markers chr2H_544617620 and chr2H_742203860. Annotations and genomic coordinates are based on the Morex reference assembly Hv_IBSC_PGsb_v2. *HvLFY* (*HORVU2Hr1G102590*) is indicated in bold, genes tested by PCR are indicated with *.

Gene	Start	End	Strand	Annotation
<i>HORVU2Hr1G102130</i>	695,896,102	695,898,815	+	Hypoxia-responsive family protein
<i>HORVU2Hr1G102140</i> *	695,905,369	695,918,422	-	Nucleolar pre-ribosomal-associated protein 1
<i>HORVU2Hr1G102150</i>	695,907,443	695,907,987	-	Undescribed protein
<i>HORVU2Hr1G102160</i>	695,955,837	695,956,130	-	Peroxidase
<i>HORVU2Hr1G102170</i>	695,957,305	695,957,622	-	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
<i>HORVU2Hr1G102180</i>	696,105,155	696,107,635	+	COBW domain-containing protein 1
<i>HORVU2Hr1G102190</i>	696,109,724	696,112,387	+	WD-40 repeat family protein
<i>HORVU2Hr1G102200</i>	696,113,153	696,117,753	+	Clathrin interactor EPSIN 1
<i>HORVU2Hr1G102210</i>	696,118,526	696,119,631	+	Undescribed protein
<i>HORVU2Hr1G102220</i>	696,121,090	696,121,633	+	Undescribed protein
<i>HORVU2Hr1G102230</i>	696,161,428	696,161,899	-	Unknown function
<i>HORVU2Hr1G102240</i>	696,230,841	696,231,956	-	Undescribed protein
<i>HORVU2Hr1G102250</i>	696,232,110	696,232,926	+	Undescribed protein
<i>HORVU2Hr1G102260</i>	696,239,400	696,241,577	-	Cytochrome P450, family 718
<i>HORVU2Hr1G102270</i>	696,240,138	696,241,577	+	Unknown function
<i>HORVU2Hr1G102280</i>	696,253,418	696,254,029	-	Undescribed protein
<i>HORVU2Hr1G102290</i>	696,251,647	696,253,600	+	Undescribed protein
<i>HORVU2Hr1G102300</i> *	696,257,846	696,260,039	+	P-loop containing nucleoside triphosphate hydrolases superfamily protein
<i>HORVU2Hr1G102310</i>	696,278,497	696,281,523	-	Undescribed protein
<i>HORVU2Hr1G102320</i>	696,293,510	696,294,994	-	Unknown protein
<i>HORVU2Hr1G102330</i>	696,296,443	696,298,467	+	Wound-induced basic protein
<i>HORVU2Hr1G102340</i>	696,308,004	696,308,485	+	Undescribed protein
<i>HORVU2Hr1G102350</i>	696,373,179	696,373,343	+	Retrotransposon protein, putative, unclassified
<i>HORVU2Hr1G102360</i>	696,374,857	696,375,615	+	Undescribed protein
<i>HORVU2Hr1G102370</i>	696,384,061	696,386,592	-	P-loop containing nucleoside triphosphate hydrolases superfamily protein
<i>HORVU2Hr1G102380</i>	696,582,483	696,584,350	+	HXXXD-type acyl-transferase family protein
<i>HORVU2Hr1G102390</i>	696,585,683	696,586,172	-	Undescribed protein
<i>HORVU2Hr1G102400</i>	696,679,576	696,681,060	+	Scramblase-related
<i>HORVU2Hr1G102410</i> *	696,704,953	696,717,024	+	TPR and ankyrin repeat-containing protein 1
<i>HORVU2Hr1G102420</i>	696,725,450	696,725,671	-	Myosin-J heavy chain
<i>HORVU2Hr1G102430</i>	696,725,747	696,725,986	-	RNA-directed DNA polymerase (reverse transcriptase)-related family protein
<i>HORVU2Hr1G102440</i>	696,726,715	696,726,867	-	Undescribed protein
<i>HORVU2Hr1G102450</i>	696,727,773	696,728,892	-	Unknown function
<i>HORVU2Hr1G102460</i>	696,780,511	696,783,093	+	Glycosyltransferase family 61 protein
<i>HORVU2Hr1G102470</i>	696,814,282	696,831,996	+	Undescribed protein
<i>HORVU2Hr1G102480</i>	696,824,849	696,831,942	-	Undescribed protein
<i>HORVU2Hr1G102490</i>	696,822,457	696,823,275	+	F-box family protein
<i>HORVU2Hr1G102500</i>	696,874,935	696,880,571	+	Tudor/PWWP/MBT superfamily protein
<i>HORVU2Hr1G102510</i>	696,897,305	696,900,942	+	Protein kinase family protein
<i>HORVU2Hr1G102520</i>	696,904,256	696,904,561	-	Undescribed protein
<i>HORVU2Hr1G102530</i> *	696,904,801	696,907,731	-	Protein kinase family protein
<i>HORVU2Hr1G102540</i>	696,904,801	696,920,678	+	Undescribed protein

Supplementary Table S2. *Continued*

Gene ID	Start	End	Strand	Annotation
<i>HORVU2Hr1G102550</i>	696,910,217	696,912,401	-	Retrotransposon protein, putative, unclassified
<i>HORVU2Hr1G102560</i>	696,912,535	696,913,220	-	Retrotransposon protein, putative, unclassified
<i>HORVU2Hr1G102570</i>	696,990,165	696,992,387	+	General transcription factor 2-related zinc finger protein
<i>HORVU2Hr1G102580</i>	696,999,051	697,004,597	+	Protein kinase family protein
<i>HORVU2Hr1G102590</i> *	697,165,877	697,169,123	-	Floricaula/leafy homolog
<i>HORVU2Hr1G102610</i> *	697,191,404	697,193,301	-	60S ribosomal protein L12
<i>HORVU2Hr1G102620</i>	697,328,577	697,332,674	+	Unknown function
<i>HORVU2Hr1G102630</i>	697,328,750	697,330,304	-	Undescribed protein
<i>HORVU2Hr1G102640</i>	697,518,389	697,522,933	+	Undescribed protein
<i>HORVU2Hr1G102650</i>	697,379,522	697,380,518	+	Undescribed protein
<i>HORVU2Hr1G102660</i> *	697,379,522	697,380,600	-	Unknown function
<i>HORVU2Hr1G102670</i>	697,388,779	697,389,088	+	Unknown function
<i>HORVU2Hr1G102680</i>	697,393,607	697,393,997	+	Undescribed protein
<i>HORVU2Hr1G102690</i>	697,402,513	697,403,044	-	Undescribed protein
<i>HORVU2Hr1G102700</i>	697,405,242	697,405,876	+	Undescribed protein
<i>HORVU2Hr1G102710</i>	697,576,254	697,580,169	-	Protein NRT1/ PTR family 8.3
<i>HORVU2Hr1G102720</i>	697,623,190	697,623,877	+	Protein NRT1/ PTR family 8.3
<i>HORVU2Hr1G102730</i>	697,736,651	697,738,756	-	Protein NRT1/ PTR family 8.3
<i>HORVU2Hr1G102740</i> *	697,856,647	697,864,487	-	WRKY DNA-binding protein 35
<i>HORVU2Hr1G102750</i>	698,009,411	698,009,722	+	Unknown function
<i>HORVU2Hr1G102760</i>	698,010,610	698,010,892	+	Unknown function
<i>HORVU2Hr1G102770</i>	698,011,045	698,011,323	+	Retrotransposon protein, putative, unclassified
<i>HORVU2Hr1G102780</i>	698,144,749	698,149,032	+	Secretory carrier-associated membrane protein 6
<i>HORVU2Hr1G102790</i> *	698,160,555	698,166,624	-	Leucine-rich repeat receptor-like protein kinase family protein
<i>HORVU2Hr1G102800</i>	698,219,974	698,225,810	-	Elongation factor 1-alpha
<i>HORVU2Hr1G102810</i> *	698,231,394	698,234,766	-	26S proteasome non-ATPase regulatory subunit 13 homolog B
<i>HORVU2Hr1G102820</i>	698,242,715	698,243,199	-	Undescribed protein
<i>HORVU2Hr1G102830</i>	698,292,865	698,293,458	-	Unknown function
<i>HORVU2Hr1G102840</i>	698,313,955	698,318,954	-	Vacuolar cation/proton exchanger 1b
<i>HORVU2Hr1G102850</i>	698,352,819	698,353,021	+	Undescribed protein
<i>HORVU2Hr1G102860</i> *	698,441,459	698,444,405	+	MIP18 family protein FAM96A
<i>HORVU2Hr1G102870</i>	698,442,509	698,442,746	-	Undescribed protein
<i>HORVU2Hr1G102880</i>	698,444,733	698,447,746	+	Protein of unknown function (DUF760)
<i>HORVU2Hr1G102890</i>	698,450,482	698,452,969	+	Oligopeptide transporter 4
<i>HORVU2Hr1G102900</i>	698,616,999	698,618,771	+	Zinc finger protein 862
<i>HORVU2Hr1G102910</i>	698,621,895	698,622,620	-	Protein of unknown function (DUF674)
<i>HORVU2Hr1G102920</i>	698,666,842	698,670,298	+	Subtilisin-like protease
<i>HORVU2Hr1G102930</i> *	698,730,196	698,732,521	+	High mobility group B protein 6
<i>HORVU2Hr1G102940</i>	698,774,176	698,775,582	-	FAR1-related sequence 5
<i>HORVU2Hr1G102950</i>	698,779,804	698,788,619	+	Polyphenol oxidase, chloroplastic
<i>HORVU2Hr1G102960</i>	698,784,613	698,785,579	-	Unknown function
<i>HORVU2Hr1G102970</i>	698,868,231	698,868,725	+	F-box/RNI-like superfamily protein
<i>HORVU2Hr1G102980</i>	698,868,826	698,870,090	+	Chromosome 3B, genomic scaffold, cultivar Chinese Spring
<i>HORVU2Hr1G102990</i>	698,871,780	698,872,135	-	Ubiquinol oxidase
<i>HORVU2Hr1G103000</i>	698,877,139	698,881,499	-	Polyphenol oxidase, chloroplastic
<i>HORVU2Hr1G103010</i> *	698,900,914	698,903,541	-	Autophagy 8E
<i>HORVU2Hr1G103020</i>	698,910,153	698,916,534	+	Pectin lyase-like superfamily protein

Supplementary Table S2. *Continued*

Gene ID	Start	End	Strand	Annotation
<i>HORVU2Hr1G103030</i>	698,910,235	698,911,454	-	Undescribed protein
<i>HORVU2Hr1G103040</i>	699,028,407	699,034,603	+	Polyphenol oxidase, chloroplastic
<i>HORVU2Hr1G103050</i>	699,045,987	699,047,435	+	Undescribed protein
<i>HORVU2Hr1G103060</i>	699,045,341	699,045,560	+	Undescribed protein
<i>HORVU2Hr1G103070</i>	699,052,486	699,053,443	+	Fatty acid desaturase 2
<i>HORVU2Hr1G103080</i>	699,151,247	699,151,516	+	Unknown function
<i>HORVU2Hr1G103090</i>	699,184,255	699,185,119	+	Undescribed protein
<i>HORVU2Hr1G103100</i> *	699,184,532	699,188,117	-	Signal recognition particle 14 kDa protein
<i>HORVU2Hr1G103110</i>	699,192,383	699,195,437	+	Undescribed protein
<i>HORVU2Hr1G103120</i>	699,193,182	699,195,429	-	Brefeldin A-inhibited guanine nucleotide-exchange protein 1
<i>HORVU2Hr1G103130</i>	699,225,050	699,226,419	-	Aminomethyltransferase
<i>HORVU2Hr1G103140</i>	699,250,044	699,250,760	+	Unknown function
<i>HORVU2Hr1G103150</i>	699,286,827	699,287,546	+	Pectinesterase inhibitor domain containing protein
<i>HORVU2Hr1G103160</i>	699,290,789	699,291,430	+	Undescribed protein
<i>HORVU2Hr1G103170</i>	699,312,615	699,315,791	-	DNA polymerase III subunit epsilon
<i>HORVU2Hr1G103180</i> *	699,321,924	699,325,619	+	L-lactate dehydrogenase
<i>HORVU2Hr1G103190</i>	699,415,076	699,415,604	+	bZIP transcription factor TRAB1
<i>HORVU2Hr1G103200</i>	699,419,184	699,419,761	+	Undescribed protein
<i>HORVU2Hr1G103210</i>	699,436,544	699,437,695	-	Beta-lactamase domain-containing protein 2
<i>HORVU2Hr1G103220</i>	699,442,362	699,447,731	-	Polyamine oxidase 2
<i>HORVU2Hr1G103230</i>	699,476,154	699,477,081	-	Undescribed protein
<i>HORVU2Hr1G103240</i>	699,507,826	699,508,845	-	bZIP transcription factor 27
<i>HORVU2Hr1G103250</i>	699,549,790	699,549,903	+	Protein-tyrosine phosphatase mitochondrial 1
<i>HORVU2Hr1G103260</i>	699,554,903	699,555,225	-	Undescribed protein
<i>HORVU2Hr1G103270</i>	699,724,494	699,725,057	+	Undescribed protein
<i>HORVU2Hr1G103280</i>	699,725,317	699,725,965	+	Peroxidase superfamily protein

Supplementary Table S3A. Sequence of KASP™ markers used to confirm heterozygosity of F₁ plants derived from a cross between mov5.o (cv. Morex) and Steptoe.

Marker name	Forward primer Allele 1	Forward primer Allele 2	Reverse Common primer	SNP
chr3H_1006543	GAAGGTGACCAAGTTCATGCTAAGCTTGATTTCCACATGACCAATTTT	GAAGTTCGGAGTCAACGGATTCTAAGCTTGATTTCCACATGACCAATTTT	ACTCTCCATGGCTGACCTT	[G/A]
chr3H_7767159	GAAGGTGACCAAGTTCATGCTGATCGGCGAGGACGAGACA	GAAGTTCGGAGTCAACGGATTGATCGGCGAGGACGAGACG	CCCACCGTCGAGACTCCGATA	[A/G]
chr3H_8787424	GAAGGTGACCAAGTTCATGCTATGTAAGTTTCTAGATGATGCAAGGC	GAAGTTCGGAGTCAACGGATTGATGTAAGTTTCTAGATGATGCAAGGA	CAITTTCCAAGTTTCTGCCACAGTTT	[G/T]
chr3H_11702941	GAAGGTGACCAAGTTCATGCTGACAGCCGTTTCGTCGCCACA	GAAGTTCGGAGTCAACGGATTGACCCGTTTCGTCGCCACG	CCGCTTGTCTACGCTGATATGAT	[A/G]
chr3H_14011512	GAAGGTGACCAAGTTCATGCTGAGCCGCTGATTTTGACGAGA	GAAGTTCGGAGTCAACGGATTGACCCGCTGATTTTGACGAGG	GAACCGCATAACGAAAATCATATGAA	[T/C]
chr3H_28805649	GAAGGTGACCAAGTTCATGCTGACGGTTCATTCTGCAC	GAAGTTCGGAGTCAACGGATTGCTGACGGTTCATTCTGCAC	TCTTGATGTCGATGATATGTTCTTGAT	[G/T]

Supplementary Table S3B. Sequence of KASP™ markers used to identify as pre-screen to identify the chromosome of interest, indicative marker on chromosome 2H is shown in bold. Marker order is based on the genetic map.

Marker name	Forward primer Allele 1	Forward primer Allele 2	Reverse Common primer	SNP
chr1H_13590934	GAAGGTGACCAAGTTCATGCTACCTTACATGGCTTCTGAGCAA	GAAGTTCGGAGTCAACGGATTACCTTACATGGCTTCTGAGCAG	GGTCTRAAGTTGATCCATGCTTATT	[A/G]
chr1H_46538882	GAAGGTGACCAAGTTCATGCTGATACAGTCTGAAGACCTGTGA	GAAGTTCGGAGTCAACGGATTGATACAGTCTGAAGACCTGTCTC	TATAACTCGGTGACCTGGACAA	[A/C]
chr1H_402308685	GAAGGTGACCAAGTTCATGCTAAAACCTCTGATTGCTGTTAAAGAGTACT	GAAGTTCGGAGTCAACGGATTACTCTGATTGCTGTTAAAGAGTACTG	TATGAACCTTTCTCTTCTGCTGCCAA	[A/C]
chr1H_498205652	GAAGGTGACCAAGTTCATGCTGAGGACAGTCTGTAACCTCTTAA	GAAGTTCGGAGTCAACGGATTAGGACAGTCTGTAACCTCTTAG	CTCAATGAAGCTGATCTGGAGGTT	[A/G]
chr1H_533708596	GAAGGTGACCAAGTTCATGCTGATTTCAAATACCACTAAGCACC	GAAGTTCGGAGTCAACGGATTCAAATACCACTAAGCACC	GCATAGTTCTTCCAGTGAACAAGAGAAT	[A/C]
chr2H_21582729	GAAGGTGACCAAGTTCATGCTCAGTATTACTCTTTGCGCCG	GAAGTTCGGAGTCAACGGATTACTCTCTTTGCGCCG	ACGGTAACAGCTAATATAAGTAATTCGACAA	[C/G]
chr2H_544617620	GAAGGTGACCAAGTTCATGCTGAGCGATGTTGAAGCTGCCCAA	GAAGTTCGGAGTCAACGGATTAGCGATGTTGAAGCTGCCCAA	CGAGAGGCCAGTCTGGAAACAAT	[A/G]
chr2H_659264767	GAAGGTGACCAAGTTCATGCTACAGGAGGCTGATCATGATCGTT	GAAGTTCGGAGTCAACGGATTACAGGAGGCTGATCATGATCGTC	CCCGCTTCTCGGAGGCTGTT	[A/G]
chr2H_742203860	GAAGGTGACCAAGTTCATGCTGCTCGATCTGCTACCTGAT	GAAGTTCGGAGTCAACGGATTCTGATCTGCTACCTGAT	TTGAGTATAACTGCCTCAAGCTATGTT	[A/G]
chr3H_9748112	GAAGGTGACCAAGTTCATGCTATCAACGCCAATCAAGTTACTTTAC	GAAGTTCGGAGTCAACGGATTATCAACGCCAATCAAGTTACTTTAG	CCCTGAATGAATAACCTTTTTTAGGGAA	[C/G]
chr3H_10289104	GAAGGTGACCAAGTTCATGCTGCGCGGGAGGCTCTG	GAAGTTCGGAGTCAACGGATTGCGCGGGAGGCTCTC	CCCCGTCAAAGCTCCCAAGAA	[G/C]
chr3H_55590274	GAAGGTGACCAAGTTCATGCTGAAATGGTGCATAGGTGGCAA	GAAGTTCGGAGTCAACGGATTGAAATGGTGCATAGGTGGCAG	GAGCTCACGGTGGCTGACACT	[A/G]
chr3H_605214250	GAAGGTGACCAAGTTCATGCTGCGCCACGCATCAGCA	GAAGTTCGGAGTCAACGGATTGCGCCACGCATCAGCG	TTGCGAAAGGTCAAGCAAGTTAGCTAAA	[A/G]
chr3H_638623189	GAAGGTGACCAAGTTCATGCTGAAGTAATACACGGCGCGCAGA	GAAGTTCGGAGTCAACGGATTGAAGTAATACACGGCGCGCAGG	GTTTCTGAACAGAAACCTCAACATGGTT	[A/G]
chr4H_9579405	GAAGGTGACCAAGTTCATGCTATTATTAGATTGAGAATCGAATGACCT	GAAGTTCGGAGTCAACGGATTGATTGAGAATCGAATGACCT	GGTAACCAACCTCTCTGCTGCAA	[A/G]
chr4H_471656963	GAAGGTGACCAAGTTCATGCTGATCAGCCTATCCTCGCACCT	GAAGTTCGGAGTCAACGGATTATCAGCCTATCCTCGCACCC	AAGTAGGGCTAACATAGTGTGATCAA	[A/G]
chr4H_581788124	GAAGGTGACCAAGTTCATGCTAATCGCAGAGGAAACCAAGACAAA	GAAGTTCGGAGTCAACGGATTATCGCAGAGGAAACCAAGACAAA	CGAGATCGCCGTGGAACCT	[A/G]
chr4H_623749491	GAAGGTGACCAAGTTCATGCTCCTCAGTTGATACCGCTGGGT	GAAGTTCGGAGTCAACGGATTCTCAGTTGATACCGCTGGGG	TGCGGTGCCGAGAAATATACAGTA	[A/G]
chr5H_9866403	GAAGGTGACCAAGTTCATGCTAAACACATACGAAATCTTGCTATTGGATT	GAAGTTCGGAGTCAACGGATTACATACGAAATCTTGCTATTGGATC	CGCCCTGTTAAAGGAGGCTTTTA	[A/G]
chr5H_144294247	GAAGGTGACCAAGTTCATGCTCCACAGGAGCTTGCAAGCTGAT	GAAGTTCGGAGTCAACGGATTCCACAGGAGCTTGCAAGCTGAC	CTGGAAGCTGTGACCAAGCAA	[A/G]
chr5H_483243877	GAAGGTGACCAAGTTCATGCTGCTGCTGAAAGATTGCCAGTGAAA	GAAGTTCGGAGTCAACGGATTGCTGCTGAAAGATTGCCAGTGAAAT	AGATATCTCAACACTTATCCCAAGTT	[A/T]
chr5H_542621417	GAAGGTGACCAAGTTCATGCTCCCAACCAACATCCACAATA	GAAGTTCGGAGTCAACGGATTCCCAACCAACATCCACAATG	ATCTGCGGGCAGCTCCACAT	[A/G]
chr5H_603612830	GAAGGTGACCAAGTTCATGCTTACAGCATAGGACTCAAACCTG	GAAGTTCGGAGTCAACGGATTACAGCATAGGACTCAAACCCG	CAACTACTCGTCTTACTGACTGAGAT	[A/G]
chr5H_638951179	GAAGGTGACCAAGTTCATGCTGCTGCGGCAAGGCG	GAAGTTCGGAGTCAACGGATTGCTGCTGCGGCAAGGCCC	GATGTGAAGACGAGCTGATGTTA	[C/G]
chr6H_34582878	GAAGGTGACCAAGTTCATGCTATCCAGATCCGCGCCAC	GAAGTTCGGAGTCAACGGATTATCCAGATCCGCGCCACG	CGTTAGATCGCAAGCAACAACAAATA	[C/G]
chr6H_463863740	GAAGGTGACCAAGTTCATGCTGCAACCCGCAAGCAGCTGT	GAAGTTCGGAGTCAACGGATTCAACCCGCAAGCAGCTGCTC	CCACAACAGGTGGTGTGTACAAAT	[A/G]
chr6H_541830535	GAAGGTGACCAAGTTCATGCTACAAAAGTTTGGGTCTAAACAAGTAAT	GAAGTTCGGAGTCAACGGATTCAAAAAGTTTGGGTCTAAACAAGTAAC	GTGGTTATTGGTGGATGGGACTGTA	[A/G]
chr7H_12860024	GAAGGTGACCAAGTTCATGCTGAGGCTGAGGACGAGGAG	GAAGTTCGGAGTCAACGGATTGAGGCTGAGGACGAGGAC	GTCTGCACTAGTTCTTGCCATCAT	[C/G]
chr7H_95639213	GAAGGTGACCAAGTTCATGCTCATGTACCAGGAGCTGTTCT	GAAGTTCGGAGTCAACGGATTATGTACCAGGAGCTGTTCT	GGCTCATCCAGTACTTCTCAGTT	[A/G]
chr7H_626516365	GAAGGTGACCAAGTTCATGCTCCACAAGTAAGGACGAGGAGCT	GAAGTTCGGAGTCAACGGATTCAACAAGTAAGGACGAGGAGCC	GGTACTATAAAGTTTACAGCCGAACAAA	[A/G]
chr7H_645016449	GAAGGTGACCAAGTTCATGCTGCTGTTGGTCTGCTGCT	GAAGTTCGGAGTCAACGGATTGCTGCTGTTGGTCTGCTGCTC	CCACAATACTGCTGAAGTCCATTGAT	[A/G]

Supplementary Table S3C. Sequence of KASP™ markers on chromosome 2H used to map the *mov5* locus, flanking markers are indicated in bold. Marker order is based on the genetic map.

Marker name	Forward primer Allele 1	Forward primer Allele 2	Reverse Common primer	SNP
chr2H_615045226	GAAGGTGACCAAGTTCATGCTAGAATCCTCAGGTCATACACCTA	GAAGTCGGAGTCAACGGATTAATCCTCAGGTCATACACCTG	CTTAGCAACTATGTTGCATATGTTGGGTA	[A/G]
chr2H_631976772	GAAGGTGACCAAGTTCATGCTGCAAGAACTATGGAAGCTAAGGTAAT	GAAGTCGGAGTCAACGGATTAACAAAGAACTATGGAAGCTAAGGTAAC	TGCGCCGACGATGGTGAAGGAT	[A/G]
chr2H_659264767	GAAGGTGACCAAGTTCATGCTACAGGAGGCTGTACATGATCGTT	GAAGTCGGAGTCAACGGATTACAGGAGGCTGTACATGATCGTC	CCCGGCTTCTCGGAGGCTGTT	[A/G]
chr2H_663236849	GAAGGTGACCAAGTTCATGCTGCCAATTTAGCACATCAACAAAAGTC	GAAGTCGGAGTCAACGGATTGCCAATTTAGCACATCAACAAAAGTC	AGGCAGGGATCTCTTCGGAATCAA	[C/G]
chr2H_672027508	GAAGGTGACCAAGTTCATGCTATTGATGGTACCGGTGCTTGCTT	GAAGTCGGAGTCAACGGATTGATGGTACCGGTGCTTGCTC	CAAAGGCTTTGAGTTGCAATTTGCACAA	[A/G]
chr2H_677023646	GAAGGTGACCAAGTTCATGCTTGAACACCCCTGTGGATTCTGAT	GAAGTCGGAGTCAACGGATTGAACACCCCTGTGGATTCTGAC	GCGAGAAGCTGTACTGTATTGGAA	[T/C]
chr2H_685623966	GAAGGTGACCAAGTTCATGCTATATAAAGGCATCCCGATGGAGAC	GAAGTCGGAGTCAACGGATTATATAAAGGCATCCCGATGGAGAT	GCCTTGTGGCCCTCTCGAGAT	[C/T]
chr2H_694119660	GAAGGTGACCAAGTTCATGCTATAAATTCGAATTTGAGATCTTCCTC	GAAGTCGGAGTCAACGGATTATAAATTCGAATTTGAGATCTTCCTG	CAGCACCTCAATTTGAGAATCATAAGGTT	[C/G]
chr2H_695593770	GAAGGTGACCAAGTTCATGCTCCGCTCTACCTCG	GAAGTCGGAGTCAACGGATTAACCTCTCCGCTCTACCTCA	TGCCACGTGTTGCGGATGTT	[G/C]
chr2H_695884392	GAAGGTGACCAAGTTCATGCTCAACACGACAGATGAGGATGG	GAAGTCGGAGTCAACGGATTCAACACGACAGATGAGGAGTGA	GGCGTCAGATCGTCATAGCGAT	[G/A]
chr2H_697042015 *	GAAGGTGACCAAGTTCATGCTAGCTGCGGAGCCTGGTGA	GAAGTCGGAGTCAACGGATTGCTGCGGAGCCTGGTGG	GCGCCTGCCCGCTGGTA	[C/T]
chr2H_699725902	GAAGGTGACCAAGTTCATGCTCAAGCACATAAACTTTGTCCAACATGTT	GAAGTCGGAGTCAACGGATTCAAGCACATAAACTTTGTCCAACATGTA	GTGGTAGAAAAATGGTTGGTCTGTTGTAA	[A/T]
chr2H_699753442	GAAGGTGACCAAGTTCATGCTCCATGAAATTTAGAAAAACAGAGGCATT	GAAGTCGGAGTCAACGGATTATGAAATTTAGAAAAACAGAGGCATC	TAGCCATGGCATAATTTGGTTCACTTT	[A/G]
chr2H_700117606	GAAGGTGACCAAGTTCATGCTGGCTCATCAAGACCTTCCTCG	GAAGTCGGAGTCAACGGATTGGCTCATCAAGACCTTCCTCA	CGTAAATATCGCTGAGTGAACGGTATA	[C/T]
chr2H_700214504	GAAGGTGACCAAGTTCATGCTATACGTGAATTTACTGACGGGAG	GAAGTCGGAGTCAACGGATTATACGTGAATTTACTGACGGGAA	TAGATCGGTTCAGCTCAAGCGCTTT	[C/T]
chr2H_700396591	GAAGGTGACCAAGTTCATGCTGCAAAATGTACAGCAATCAATACATCAGA	GAAGTCGGAGTCAACGGATTGCAAAATGTACAGCAATCAATACATCAGC	CCGGCGAGACCATTCTGTCTAAAT	[A/C]
chr2H_702156095	GAAGGTGACCAAGTTCATGCTGTTCAAGGACCTTCCTCA	GAAGTCGGAGTCAACGGATTGTTCAAGGACCTTCCTCC	TAAGTAAAGATAAATACAGACTCCCATCAA	[A/C]
chr2H_704891286	GAAGGTGACCAAGTTCATGCTGCAATTAATACTTACTTACGTTACGT	GAAGTCGGAGTCAACGGATTGCAATTAATACTTACTTACGTTACGC	GTGCGACATGCAGCATATAGCAGTA	[A/G]
chr2H_718628790	GAAGGTGACCAAGTTCATGCTTGACCCGTAGAAATGAACG	GAAGTCGGAGTCAACGGATTGACCCGTAGAAATGAAC	CGAAGGATTGACCGCTCCAT	[C/G]
chr2H_718962633	GAAGGTGACCAAGTTCATGCTTGAGACCTACCCCGTGTCTAT	GAAGTCGGAGTCAACGGATTGAGACCTACCCCGTGTCTCAA	AGCGGTCAGTCTGGGCTGAT	[T/A]
chr2H_722462781	GAAGGTGACCAAGTTCATGCTATTTATTTTCCCGGAAGAAGAAAT	GAAGTCGGAGTCAACGGATTATTTATTTTCCCGGAAGAAGAAATC	CGCTGAAGGTATGAAGGACCTGAAA	[A/G]
chr2H_732633470	GAAGGTGACCAAGTTCATGCTATAGCAACAGCCAGCCAGTATT	GAAGTCGGAGTCAACGGATTAGCAACAGCCAGCCAGTATC	CGCACAACTATATGTTGCCTCTGGTT	[T/C]
chr2H_742203860	GAAGGTGACCAAGTTCATGCTGCCTCGATCTCGTCTACCTGAT	GAAGTCGGAGTCAACGGATTCTCGATCTCGTCTACCTGAC	TTGAGTATAACTGCTCAAGCTATGTT	[A/G]

* KASP™ marker within *HvLFY* that completely co-segregates with the *mov5.o* phenotype.

Supplementary Table S4A. PCR primer sequences for testing the presence of genes within the mapped critical interval for *mov5* on chromosome 2H; *HvLFY* (*HORVU2Hr1G102590*) is indicated in bold.

Gene ID	Forward primer	Reverse primer
<i>HORVU2Hr1G102140</i>	ATTCTGCCTAATCGTGTGGTGC	TGCCAAAGCTGGTCATTCACG
<i>HORVU2Hr1G102300</i>	GCTAGTGGTGATGCAGATCAGG	GATTAGCTCGGTGCGTTGG
<i>HORVU2Hr1G102410</i>	GAGTCTCATCTTCAGTCCCAAGC	GCAGGACATGCACAATCTTACG
<i>HORVU2Hr1G102530</i>	AGATACACAAACACGCATCTGC	TTGGAGATGGTATCACAACCTGG
<i>HORVU2Hr1G102530</i>	CAGTAAATCAGGGAACGTGAGG	GAGAGAGTCCATCTGTATTCTTACC
<i>HORVU2Hr1G102590</i>	CGAATAAATGCCGTGCAGTGC	CATCTGGGCTTGTTGATGTAGC
<i>HORVU2Hr1G102610</i>	TCCCTTCGTTTCATGTCAAGAGC	CATAACTCTTCGTGCCGAACC
<i>HORVU2Hr1G102660</i>	GTGCTTCTTCTCCTGTATCTCC	GCTCCAACCCCAACGG
<i>HORVU2Hr1G102740</i>	CTTGACGGCCTACTTTACC	TCTAGGATATGAACGGCCTTAGG
<i>HORVU2Hr1G102790</i>	GACATCTCTCACTCACCCACC	GGCGTTTGACCTAAATCCAACC
<i>HORVU2Hr1G102810</i>	CATCATATAAGATCGCTTGCAAGC	CGTCAGCGGGAGCATCTG
<i>HORVU2Hr1G102860</i>	GACCTGAGTAGGACTGGAAACC	AAACACGCCGACAACCTAAACG
<i>HORVU2Hr1G102930</i>	CTCTCTCCTCGGACCGATCC	CGGGCACAGTCTTCTCTCG
<i>HORVU2Hr1G103010</i>	TTCGTCTCGTCTTCTCGTGC	GCTGAAGGTCAACAATGGTCG
<i>HORVU2Hr1G103100</i>	GTCATGCTTTATCCCGTTCATCG	AGGTATTCACAGGCTTCATCTCG
<i>HORVU2Hr1G103180</i>	ATGTGAGTCTGCTGGTTCC	ATGACAACAAGACTCGTCTGATCC

Supplementary Table S4B. qRT-PCR primer sequences.

Gene name	Gene ID	Forward primer	Reverse primer
<i>HvGAPDH</i>	<i>HORVU7Hr1G074690</i>	GTGAGGCTGGTGCTGATTACG	TGGTGCAGCTAGCATTTGAGAC
<i>HvCYCLO</i>	<i>HORVU6Hr1G012570</i>	CCTGTCGTGTCGTCCGGTCTAAA	ACGCAGATCCAGCAGCCTAAAG
<i>HvTUB</i>	<i>HORVU1Hr1G081280</i>	AGTGTCTGTCCACCCACTC	AGCATGAAGTGGATCCTTGG
<i>HvHSP70</i>	<i>HORVU5Hr1G113180</i>	CGACCAGGGCAACCGACCAC	ACGGTGTTGATGGGGTTCATG
<i>HvLFY</i>	<i>HORVU2Hr1G102590</i>	CATCTTGGGCTTGTGATGTAGC	CTGCTCCAGGTGCAGTCC
<i>HvMADS14</i>	<i>HORVU5Hr1G095630</i>	CAGCGGCGGCAGGCGAGAG	CCAGGCTGGCCGTGCAAC
<i>HvMADS15</i>	<i>HORVU2Hr1G063800</i>	ATATGCCTACCGCCATGGAT	ATACAGCGAACCAAGCATTCC
<i>HvMADS2</i>	<i>HORVU3Hr1G091000</i>	CCAGCATGATATCGCCTTG	TCGAGCCAGTGGTGGATAA
<i>HvMADS4 & HvKinase*</i>		ATGCCAAGATGTTCTGGTC	TTTGGCACCTTAGCCATCAT
<i>HvKinase*</i>	<i>HORVU1Hr1G063610</i>	TTCTCGTGTTTGTCTGGTCA	ATGCCAAGATGTTCTGGTC
<i>HvMADS4*</i>	<i>HORVU1Hr1G063620</i>	ATGGAGCTCGGGTACCATC	CCTGCAGGTAGATGGAGCA
<i>HvMADS16</i>	<i>HORVU7Hr1G091210</i>	CCCAGGAGGCATACAAGAATCTGC	GCGGAAGGCGTACATGTCAGC
<i>HvMADS3</i>	<i>HORVU3Hr1G026650</i>	GCAGCAGCAGCATTACTCC	ACACATGCACGCGACAGTA
<i>HvMADS58</i>	<i>HORVU1Hr1G029220</i>	ATCATGCAGCAGCCTCAGT	GGTGTGGCCAAGCCTTAAT
<i>HvMADS13</i>	<i>HORVU1Hr1G023620</i>	TCAGTGAACCTAGGCTGC	TTTGACAGGAATAGTTGAGTACTGGT
<i>HvMADS7</i>	<i>HORVU7Hr1G054220</i>	ACCCTCTGAGTCCCTGAA	ACGAAAGTTGCACGCAAAA
<i>HvMADS8</i>	<i>HORVU5Hr1G076400</i>	CTCAGGAGCAGATAAACAACG	GTACGCGAACCGGTACTA
<i>HvDL</i>	<i>HORVU4Hr1G067780</i>	CCATGCAAGAGGCTGATGGACACG	GCGGCTGGTTCCTCTGCAGTCAG
<i>HvOSH1</i>	<i>HORVU4Hr1G009730</i>	TGGGAGATGCACTACAAGTGG	ATGTAGAAGGCGGCTTAGG
<i>HvMEL1</i>	<i>HORVU5Hr1G107020</i>	TCAAGGACGTGATGTTCTATTGC	ATCCTGGCAACTTCAAATGGT
<i>HvMSP1</i>	<i>HORVU6Hr1G033670</i>	AAGCGACACAGATGATGGAAAG	TGAAGCTTAATCTGCCGATTTCC

* *HvMADS4* (*HORVU1Hr1G063620*) completely overlaps with *HvKinase* (*HORVU1Hr1G063610*). *HvMADS4* transcript abundance has thus been obtained by subtracting *HvKinase* values (primers specific to *HvKinase*) from *HvMADS4* & *HvKinase* (primers designed to amplify both genes).

Appendix D



Statement of Authorship

Publication Status	Accepted at <i>The Plant Journal</i> on September 2019

By signing the Statement of Authorship, each author certifies that:

- i. each author's contribution to the manuscript is accurate; and
- ii. permission is granted to include the manuscript in the thesis

Principal Authors

Name of Principal Author	Dr. Yong Jia		
	genetic mapping, phylogeny, gene duplication and selection analyses, protein modelling, barley seed sampling, extraction, qRT-PCR, mRNA <i>in situ</i> hybridization, anthocyanin extraction, EMSA and data mining. Analyzed and interpreted the results. Wrote the manuscript.		
Declaration	I am jointly the primary author of this manuscript. I hereby certify that the Statement of Authorship is accurate.		

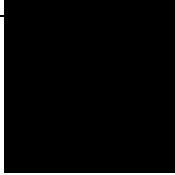
Name of Principal Author	Caterina Selva		
Contribution to the Paper	Designed and performed the following experiments: genetic mapping, syntenic analysis, wheat seed sampling, RNA extraction and semi-quantitative RT-PCR. Analysed and interpreted the results. Wrote the manuscript.		
Percentage of Contribution	45 %		
Certification	I am jointly the primary author of this manuscript. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	21/09/2019

Co-author contributions

Name of Co-author	Dr. Yujuan Zhang		
Contribution to the Paper	Assistance in experimental procedures, especially in seed sampling, RNA extraction, qRT-PCR, mRNA <i>in situ</i> hybridization and anthocyanin extraction in barley. I hereby certify that the Statement of Authorship is accurate.		
Signature	n.a.	Date	21/09/2019

Name of Co-author	Bo Li		
Contribution to the Paper	Assistance in experimental procedures, especially in regard to gene sequencing. I hereby certify that the Statement of Authorship is accurate.		
Signature		Date	21/09/2019

tribution to the Paper	sistance in experimental procedures. I hereby certify the Statement of Authorship is accurate.
	<input type="text"/>

tribution to the Paper	sistance with glasshouse trials. I hereby certify that the Statement of Authorship is accurate.
ature	 21/09/2019

tribution to the Paper	sistance in experimental procedures, especially performing genomic DNA extraction. I hereby certify that the Statement of Authorship is accurate.
	<input type="text"/> 05/9/20

tribution to the Paper	
tribution to the Paper	sistance in experimental procedures, especially performing genomic DNA extraction. I hereby certify that the Statement of Authorship is accurate.
	<input type="text"/>

tribution to the Paper	sistance in retrieving in-house genomic data. I hereby certify that the Statement of Authorship is accurate.
	<input type="text"/>

ame of Co-author	Dr. Cong Tan		
gnature		Date	21/09/2019

ame of Co-author	Dr. Tefera Angessa		
ontribution to the Paper	Assistance in experimental procedures, especially in performing genomic DNA extraction. I hereby certify that the Statement of Authorship is accurate.		

ontribution to the Paper	Assistance in experimental procedures. I hereby certify that the Statement of Authorship is accurate.		

ontribution to the Paper	Supervised designing of the experiments. Evaluated and edited the manuscript. I hereby certify that the Statement of		
gnature		Date	21/09/2019

tribution to the Paper	pervised designing of the experiments. Evaluated ited the manuscript. I hereby certify that the Stateme			
iture		<table border="1"> <tr> <td>Date</td> <td>21/09/2019</td> </tr> </table>	Date	21/09/2019
Date	21/09/2019			

Uncovering the evolutionary origin of blue anthocyanins in cereal grains

Yong Jia^{*,1,3}, Caterina Selva^{*,2}, Yujuan Zhang³, Bo Li⁴, Lee Anne McFawn⁵, Sue Broughton⁵, Xiaohu Zhang^{1,3}, Sharon Westcott⁵, Penghao Wang³, Cong Tan^{1,3}, Tefera Angessa⁵, Yanhao Xu⁴, Ryan Whitford², Chengdao Li^{1,3,4,5}

* These authors contributed equally to the work.

¹ Western Barley Genetic Alliance, Murdoch University, Western Australia, 6150, Australia

² School of Agriculture, Food and Wine, Adelaide University, South Australia, 5064, Australia

³ State Agricultural Biotechnology Centre (SABC), School of Veterinary and Life Sciences, Murdoch University, Western Australia, 6150, Australia

⁴ Hubei Collaborative Innovation Centre for Grain Industry, Yangtze University, Hubei Jingzhou 434025, China

⁵ Department of Primary Industry and Regional Development, Government of Western Australia, South Perth WA 6155, Australia

This work is supported by the Grains Research and Development Corporation (GRDC) project “Improved adaption of barley to acid soils” (project number: UMU00046) and by the University of Adelaide, Australia.

Corresponding author: Chengdao Li

Email address: C.Li@murdoch.edu.au

Summary

Functional divergence after gene duplication plays a central role in plant evolution. Among cereals, only barley, wheat and rye accumulate delphinidin-derived (blue) anthocyanins in the aleurone layer of grains, but not rice, maize and sorghum. The underlying genetic basis for this natural occurrence remains elusive. Here, we mapped the barley *Blx1* locus involved in blue aleurone to a ~ 1.13 Mb genetic interval on chromosome 4HL, thus identifying a tri-genic cluster named MbHF35 (containing *HvMYB4H*, *HvMYC4H* and *HvF35H*). Sequence and expression data supported the role of these genes in conferring blue-coloured (blue aleurone) grains. Synteny analyses across monocot species showed that MbHF35 has only evolved within distinct *Triticeae* lineages, as a result of dispersed gene duplication. Phylogeny analyses revealed a shared evolution pattern for MbHF35 in *Triticeae*, suggesting that these genes have co-evolved together. We also identified a Pooideae-specific flavonoid 3',5'-hydroxylase (F3'5'H) lineage, termed here Mo_F35H2 which has higher amino acid similarity with eudicot F3'5'Hs, demonstrating a scenario of convergent evolution. Indeed, selection tests identified 13 amino acid residues in Mo_F35H2 which underwent positive selection, possibly driven by protein thermostability selection. Furthermore, exploring the barley germplasm there's evidence to suggest that *HvMYB4H* and *HvMYC4H* have undergone human selection. Collectively, our study favours blue aleurone as a recently evolved trait resulting from environmental adaptation. Our findings provide an evolutionary explanation for the absence of blue anthocyanins in other cereals and highlight the importance of gene functional divergence for plant diversity and environmental adaptation.

Significance statement

Our findings provide a genetic and evolutionary explanation why only barley, wheat and rye grains can develop blue colour in nature, but not rice, maize and sorghum.

Keywords: blue anthocyanin, blue aleurone, barley, convergent evolution, environmental adaptation, domestication, gene duplication, flavonoid 3',5'-hydroxylase, bHLH, MYB.

Introduction

Anthocyanins are ubiquitous plant secondary metabolites and play important roles in diverse critical aspects of a plant's life cycle [1,2]. This provides a perfect model to study the interaction of plants with the environment. By absorbing different spectra of visible light, anthocyanins are responsible for the colour of most flowers, fruits and seeds [3]; thus participating in reproduction by attracting insect pollinators and seed dispersers [4,5]. They also protect against UV radiation and strong light which has facilitated the establishment of vascular plants on land from the original marine environments [2,6]. As defence mechanisms, anthocyanins act as potent antioxidants [7] and some also demonstrate antiviral, antibacterial and fungicidal activities co-operating in plant-pathogen interactions [5,8-12]. In the health sector, the antioxidant capacities of anthocyanins have recently spurred increasing interest for use in cancer, diabetes and inflammation prevention [13-15].

A MBW protein complex consisting of: R2R3-MYB transcription factors (TF), bHLH/MYC TF and WD40 proteins is required for the strict spatio-temporal regulation of anthocyanin production [16]. This complex is responsible for activating most of the biosynthetic genes of the anthocyanin pathway (**Figure S1**). Among the structural biosynthetic enzymes, two

enzymes belonging to the cytochrome P450 family catalyse the hydroxylation pattern of the flavonoid B-ring backbone, them being flavonoid 3'-hydroxylase (F3'H) and flavonoid 3'5'-hydroxylase (F3'5'H) [17,18] (**Figure S1**). Depending on the hydroxylation pattern of the flavonoid B-ring anthocyanins can be divided into: pelargonidin (4'-hydroxylated), cyanidin (3'4'-hydroxylated) and delphinidin (3'4'5'-hydroxylated); which produce red/orange, dark red and purple/blue pigments, respectively [19].

Mature cereal grains may develop different colours (yellow, purple, red, blue, black and grey) due to the different pigments accumulated. Cyanidin-derived (red) anthocyanin is dominant in both the red- and black- coloured rice [20,21]. The red colour in the aleurone layer of maize is associated with F3'H activity, which contributes to the production of cyaniding-derived anthocyanin as well [22]. Despite their abundance in plants, delphinidin-derived (blue) anthocyanins are absent in maize and rice grains [21]. However, accumulation of blue anthocyanins in the aleurone layer of grains can be seen for barley (*Hordeum vulgare*), wheat (*Triticum aestivum*) and rye (*Secale cereale*) [21,23-26]. Blue aleurone is a controversial trait: it is advantageous in agronomy, not only as a source of anthocyanin-enriched foods [21,25], but also as a visual trait to follow gene flow in hybrid breeding [24,27,28]. On the other hand, blue aleurone is an undesired characteristic in malting barley varieties used for brewing beer. It is therefore important to understand the genetics of blue aleurone in order to introduce this trait only in appropriate breeding programs. In hexaploid wheat, blue aleurone derives from introgression of chromosomal segments from wild relatives and behaves as a single dominant *Ba1* locus [29]. In contrast to wheat, the genetic basis of the blue aleurone trait in barley is more complex. It is associated by five complementary loci located on chromosomes

4H (*Blx1*, *Blx3*, *Blx4*) and 7H (*Blx2*, *Blx5*) [23,30]. However, the genes underlying these loci in both wheat and barley remain to be characterized.

In the present study, the region associated with blue-coloured barley grains was mapped to the long arm of chromosome 4H, predicted to be distant ~ 26.1 Mb from the acid soil tolerance gene *HvMATE* [31]. Three linked genes *HvMYB4H*, *HvMYC4H* and *HvF35H* were isolated and further characterized as candidates for *Blx1*. The evolutionary perspective taken suggests an explanation as to why blue anthocyanins has only been observed in the grains of some specific cereal crops. From a wider perspective, the identification of genes underlying the *Blx1* locus can have practical implications for both breeding and in the design of food products with positive health benefits.

Results

Anthocyanins accumulate in the aleurone layer during wheat and barley seed development

Barley caryopses were sampled at 14, 21, 28, 35 and 42 Days Post Anthesis (DPA) (**Figure 1A**). No significant colour differences were observed between blue (Halcyon) and white (Sloop) aleurone varieties at 14 and 21 DPA. A blue/green hue emerged in Halcyon seeds at 28 DPA and reached its highest intensity at 35 DPA. At maturity, though less distinct, colour differences remained evident between Halcyon and Sloop. Mature blue aleurone grains contain significantly higher total anthocyanin content (TAC) than white barley. Moreover, the most abundant anthocyanins in blue aleurone barley were characterised as delphinidin- and malvidin-derived anthocyanins, typically associated with a blue coloration (**Figure S2 and Appendix S1**).

Developing wheat caryopses with contrasting aleurone colour were also sampled at 12, 15, 19, 22, 26 and 30 DPA (**Figure 1B**). No observable colour differences were present at 12 and 15 DPA. The blue pigmentation was evident around the embryo at 19 DPA in Sebesta Blue 3 and was observed to expand in a basal-apical direction within the aleurone as the seeds developed. Sebesta Blue 3 seeds at maturity remained darker than Halcyon seeds and were very distinct from Sebesta seeds, which appeared golden.

Genetic mapping of the *B/x1* locus in barley

To identify the locus responsible for blue aleurone, genetic mapping in three barley populations: MN607/Vlamingh (265 lines), Sloop/Halcyon (186 lines) and Sahara/08S917N-260 (556 lines) was performed. The mapping identified critical region on the long arm of

chromosome 4H (4HL) between markers JY93 and JY810, spanning an interval of only ~ 1.13 Mb (**Figure 1C**). To further confirm the mapping results, the interval found was further verified in a barley Clipper/Sahara population (**Data S1**).

According to the public barley reference genome (Version r1) only seven candidate genes were annotated within this interval (**Table 1**). Of these seven gene, a gene encoding a putative flavonoid 3', 5'-hydroxylase (HvF35H, **Figure 1D**) and a MYB (HvMYB4H, **Figure 1E**) transcription factor (TF) were selected for further analyses based on their predicted function in the anthocyanin biosynthesis. Recently, a basic helix-loop-helix (bHLH) TF *ThMYC4E* was shown to be critical for the blue aleurone trait in wheat [32]. Homology search for *ThMYC4E* within the barley genome identified intergenic sequences with strong similarity (> 85 %, **Table 1, Figure 1F**) within the mapped interval on chromosome 4H. This putative gene, designated *HvMYC4H*, is located between *HvMYB4H* and *HvF35H* forming a tri-genic cluster termed here the MbHF35 (MYB-bHLH-F3'5'H) cluster.

Genotyping of Bowman near isogenic line (NIL) BW063 (*Blx1*) showed that the interval between flanking markers JY93 and JY810 is also substituted in BW063 compared to Bowman (**Figure 1C**). BW063 contains the dominant allele for *Blx1* and has a blue aleurone phenotype, consistent with the MbHF35 cluster underlying the *Blx1* locus.

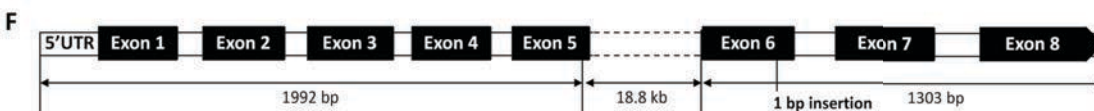
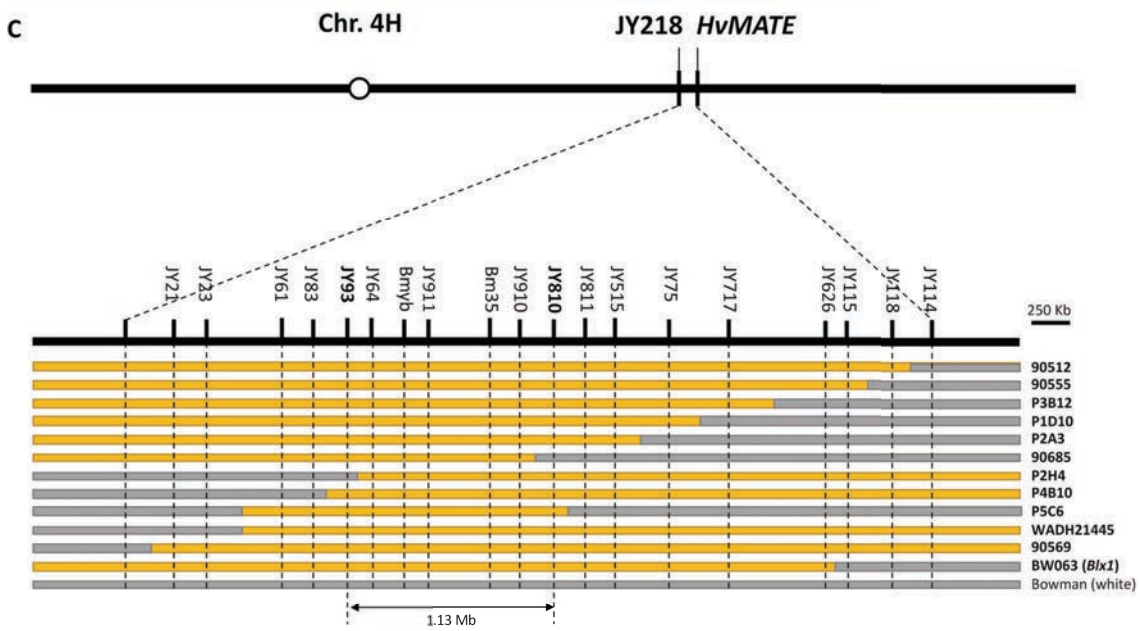
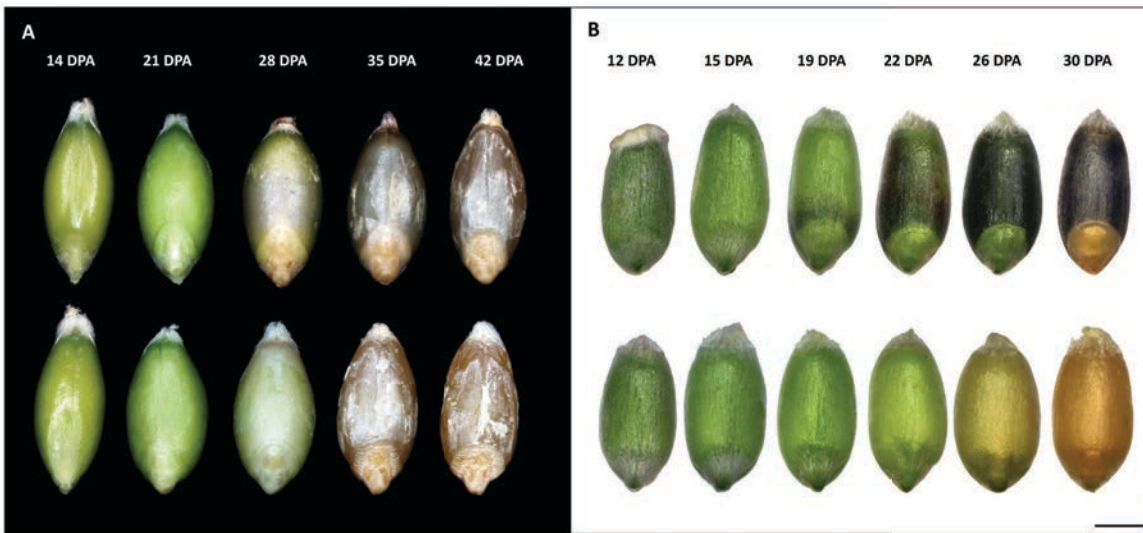


Figure 1. Developing grains with blue/white aleurone and genetic mapping. **(A):** Barley blue aleurone variety Halcyon (top row) and white aleurone variety Sloop (bottom row) at 14, 21, 28, 35 and 42 Days Post Anthesis (DPA). For wheat **(B):** Wheat blue aleurone variety Sebesta Blue 3 (top row) and white aleurone variety Sebesta (bottom row) at 12, 15, 19, 22, 26 and 30 DPA. Scale bar: 1,000 μm . **(C)** Fine mapping of the blue aleurone locus and representation of recombinant lines. Line IDs are displayed on the right-hand side, bold indicates lines having blue aleurone phenotype. Gene structures displayed for *HvF35H* **(D)**, *HvMYB4H* **(E)** and *HvMYC4H* **(F)**. Sequence variants identified between blue and white haplotypes are indicated.

Table 1. List of genes within the mapped interval. Bold genes indicate candidates in the MbHF35 cluster. *HvMYC4H* is not annotated in the reference genome assembly Hv_IBSC_PGSD_v2.

Gene ID	Start (bp)	End (bp)	Functional Annotation	PFAM ID
HORVU4Hr1G063710	533,483,898	533,488,202	Calmodulin-lysine N-methyltransferase	PF10294
HORVU4Hr1G063730	533,491,828	533,494,412	Pathogenesis-related thaumatin superfamily protein	PF00314
HORVU4Hr1G063760 (<i>HvMYB4H</i>)	533,783,874	533,785,088	MYB domain protein 7	PF00249
HORVU4Hr1G063770	533,863,136	533,864,493	Callose synthase 5	PF02364
Un-annotated (<i>HvMYC4H</i>)	534,044,159	534,066,299	<i>ThMYC4E</i> homolog	PF00010, PF14215
HORVU4Hr1G063780 (<i>HvF35H</i>)	534,190,658	534,192,411	Cytochrome P450 superfamily protein	PF00067
HORVU4Hr1G063790	534,228,837	534,232,658	Endo-14-beta-xylanase Z	PF00331
HORVU4Hr1G063810	534,320,796	534,327,020	Acyl-CoA-binding domain 3	PF00887

Genes within the MbHF35 cluster are significantly higher expressed in blue cultivars

To dissect the dynamics of the MbHF35 cluster, RT-qPCR was performed on developing seeds of Halcyon (b), BW063 (b), Sloop (w), and Bowman (w). Henceforth, (b) will be used to denote blue aleurone varieties, while (w) will be used to indicate cultivars showing a white aleurone. As shown in **Figure 2A-2B**, the overall expression of *HvF35H* in the developing seeds of blue

aleurone varieties Halcyon (b) and BW063 (b) was significantly higher at most time points compared to their white aleurone counterparts Sloop (w) and Bowman (w). Generally, expression of *HvF35H* at 14 and 21 DPA was very low for all varieties analysed and showed steep upregulation at 28 DPA. Peak expression of *HvF35H* in blue aleurone varieties (28 DPA) temporally precedes highest expression in the white aleurone varieties Sloop (w) and Bowman (w), which was detected at 42 DPA and 35 DPA, respectively. *HvF35H* levels were barely detectable or not expressed at all in young leaves and stems.

Similarly to *HvF35H*, also *HvMYC4H* tended to be significantly higher expressed in blue cultivars (**Figure 2C-2D**). As seen for *HvF35H*, *HvMYC4H* levels were generally low in the early stages of seed development and rose at 28 DPA, with the sharpest increase present in Halcyon (b) and BW063 (b). While *HvMYC4H* levels remained high in these blue aleurone varieties, with peaks at 42 DPA for Halcyon (b) and 35 DPA for BW063 (b), expression dropped or tended to remain low for Sloop (w) and Bowman (w). In addition, some *HvMYC4H* transcription was also detected in leaf tissues for Halcyon (b) and BW063 (b), but not in Sloop (w) and Bowman (w). Only a very low expression of *HvMYC4H* was observed in the stem.

The transcriptional profile for *HvMYB4H* was quite distinct (**Figure 2E-2F**). Unlike *HvF35H* and *HvMYC4H*, *HvMYB4H* levels in Halcyon (b) increased gradually from the initial stages of seed development and peaked at 28 DPA. Expression remained significantly lower or barely detectable in white varieties Sloop (w) and Bowman (w). On the other hand, expression levels tended to remain high in BW063 (b) when compared to Bowman (w), not only in seeds but also in leaf and stem.

Semi quantitative RT-PCR showed that also for the genes in the wheat MbHF35 cluster, transcript abundance was higher in blue wheat cultivar (Sebesta Blue 3) than that in the white cultivar Sebesta (**Figure S3 and Appendix S2**)

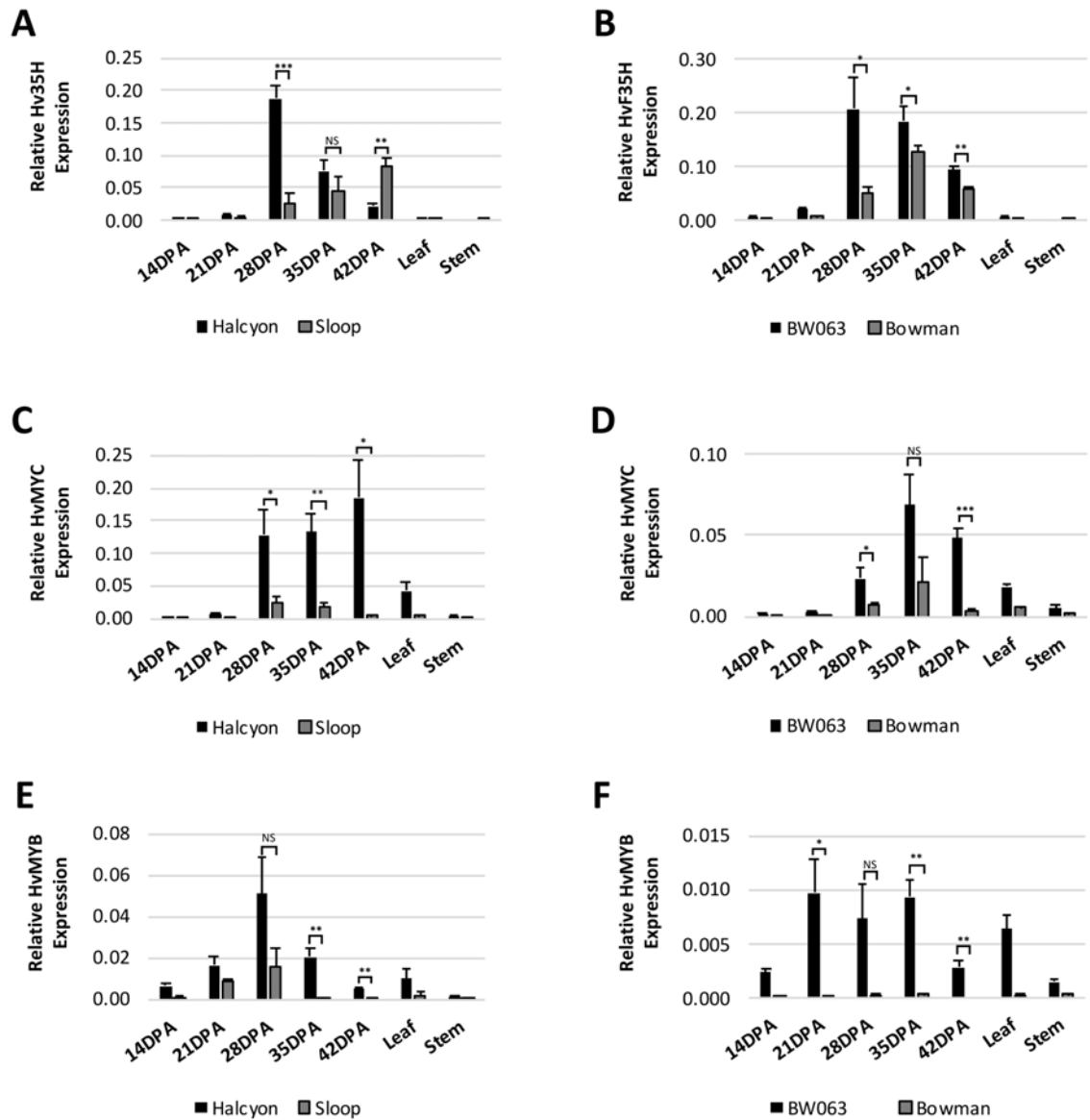


Figure 2. Transcriptional profiles of barley MbHF35 cluster genes. (**A-B**) qRT-PCR analyses for HvF35H; (**C-D**) qRT-PCR analyses for HvMYC4H and (**E-F**) qRT-PCR analyses of HvMYB4H. For each sample, n=3 for 2 technical replicates. Error bars represent \pm STD. Two-tailed t-test *P*-values ≤ 0.05 (*), ≤ 0.01 (**), ≤ 0.001 (***) and not significant (NS) are shown for differences between blue (Halcyon, BW063) and white (Sloop, Bowman) aleurone cultivars.

HvF35H, *HvMYB4H* and *HvMYC4H* are specifically expressed in aleurone cells

To validate the RT-qPCR results and determine the expression domain of *HvF35H*, *HvMYB4H* and *HvMYC4H* in developing grains, mRNA *in situ* hybridisation was performed on Halcyon (b) seeds at 28 DPA (**Figure S4**). For *HvF35H*, a clear signal was detected in the aleurone cells, while no signal was detected in the husk and seed coat. Notably, for both test and control samples a slight signal was also detected in the starchy cell layer adjacent to the aleurone layer, but with no significant difference observable between sense and antisense RNA probes. This suggests *HvF35H* is not expressed in the endosperm. Likewise, no signal was detected in the embryo except for the tipping point which directly adjoins the aleurone cell layer. However, this signal was relatively weak compared to aleurone cells. Interestingly, *HvF35H* signal was noticeably stronger in the aleurone cells of the abaxial layer compared to those on the adaxial side of the seed. For both *HvMYB4H* and *HvMYC4H* (**Figure S4**), similar expression profiles as *HvF35H* were observed. Strong signals of *HvMYB4H* and *HvMYC4H* expression were detected in the aleurone cells on the abaxial side of the seed, while relatively weaker on the adaxial side. Moderate signals were also found in the embryo cells adjoining the aleurone cells but not in the rest part of the embryo. In addition, no signal could be detected in the starchy endosperm, husk and seed coat cells.

Distinct haplotypes for the MbHF35 cluster define blue and white coloured barley

To identify genetic variation that may underlie the blue aleurone trait, the candidate genes *HvF35H*, *HvMYC4H* and *HvMYB4H* were Sanger sequenced from selected blue and white aleurone lines. For *HvF35H*, the genetic region 1078 bp upstream of the putative start codon

was also considered. Alignment of *HvF35H* sequences of blue and white cultivars identified 2 distinct haplotypes. Ten conserved SNPs, resulting in 2 non-synonymous and 4 synonymous substitutions were identified (**Figure 3A**), as well as a 1 bp insert/deletion (indel) and a 7 bp indel. Interestingly, the reference cultivar Morex (w) shared the same haplotype as blue barley with the exception of a 6 bp indel in the putative promoter (**Data S2**).

Sequencing of *HvMYB4H* revealed 10 SNPs between blue and white barley, including 2 non-synonymous mutations in the coding region and 8 SNPs in the 3'UTR (**Figure 3B**). A 4 bp insertion in blue barley was identified in the 3'UTR. In addition, a 6 bp deletion and a 2 bp deletion in the 3'UTR were only identified in Morex (w) and Sloop (w), respectively.

Mapping of the Halcyon (b) *HvMYC4H* cDNA sequence to the barley reference genome allowed to reconstruct the *HvMYC4H* gene structure (**Figure 1F**). *HvMYC4H* is composed of 8 exons, with a 18,846 bp intron between exon 5 and exon 6. Sequence alignment of the *HvMYC4H* genes between blue and white barley (excluding the long intron) identified four synonymous SNPs, one non-synonymous SNP within exons (**Figure 3C**), and several indels. Of these indels, a 1 bp insertion at position +2504 in exon 6 of white aleurone barleys caused a pre-mature termination in *HvMYC4H* translation. Sequencing of additional cultivars Dayton (b), Henley (b), Gardiner (w) and 08S917N-260 (w) showed consistency in the blue and white haplotypes identified (**Data S2**).

In addition to BW063 (b), possessing the dominant *Blx1* allele, we also sequenced BW064 and BW065 which contain the recessive alleles *blx3* and *blx4*, respectively. Results showed that both BW064 and BW065 have the blue haplotype for the three identified candidate genes (**Data S2**), further confirming the mapped locus as candidate for *Blx1*, and suggesting that *blx3* and *blx4* reside elsewhere in the genome.

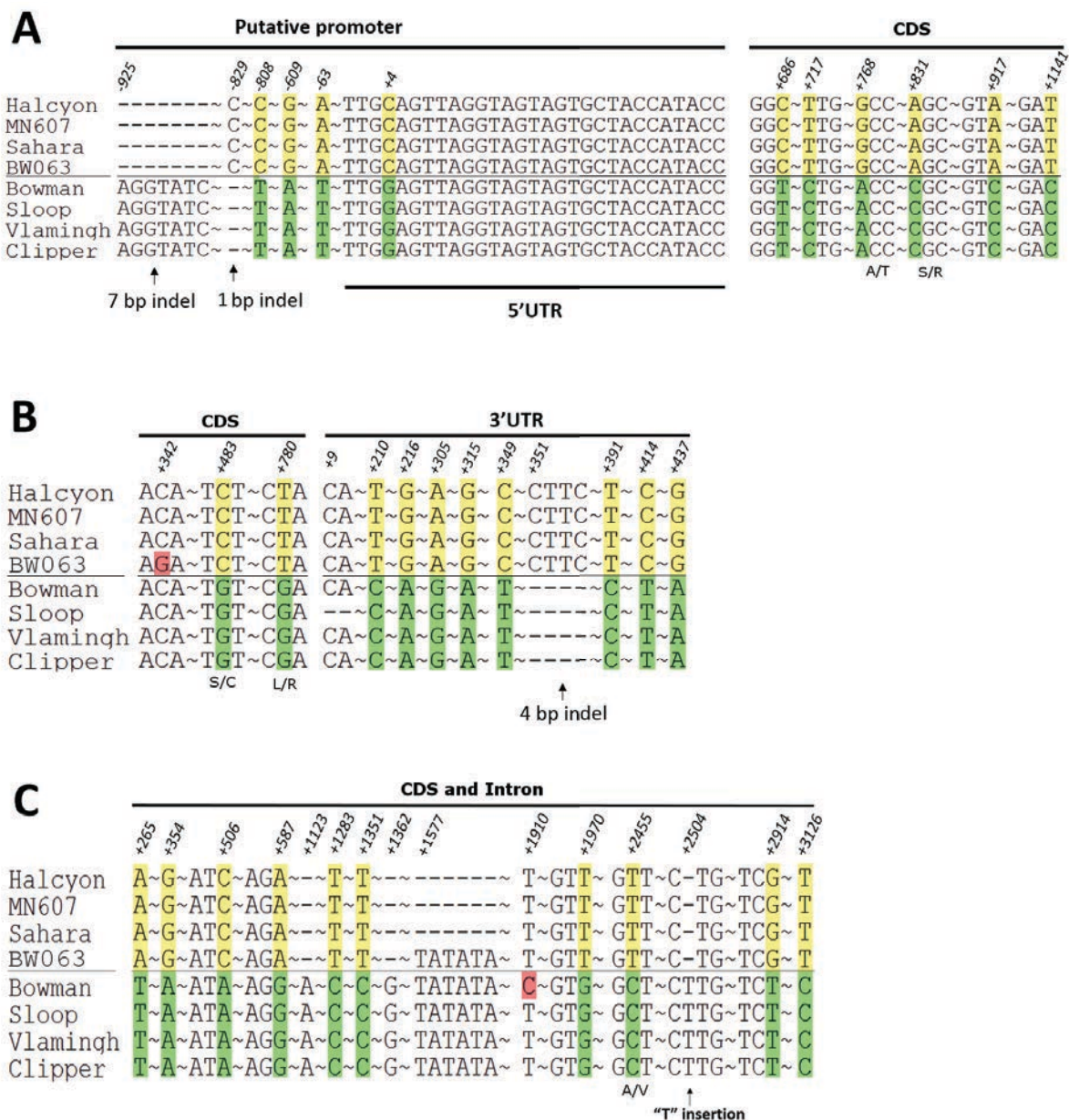


Figure 3. Sequence variations for *HvF35H*, *HvMYB4H* and *HvMYC4H* in selected barley varieties. Displaying the sequence variations of *HvF35H* (A), *HvMYB4H* (B) and *HvMYC4H* (C). Variety names are indicated on the left. The top cluster represents blue aleurone varieties, whilst the lower cluster represents white aleurone varieties. For the variations in CDS region, the three letter genetic codes were displayed. Amino acid substitutions resulting from non-synonymous SNP were indicated under the genetic code. For variations in intron region and 3'/5' UTRs, only the sequence at the variation sites were displayed. Positions were numbered with the transcription start site (TSS) as 1. For the 3'UTR numbers correspond to genomic position of variants starting from the stop codon.

HvMYB4H interacts with the putative promoter region of *HvF35H*

To understand the relationship between the genes of the MbHF35 cluster, the putative promoter binding sites of *HvF35H* were predicted using bioinformatics tools. As listed in **Table S1**, 13 MYB and 2 bHLH TFs were predicted to bind to the putative promoter region of *HvF35H*. The DNA fragments identified as potential target sites for these MYBs are mainly located at three regions: -911 to -882 (PR1), -360 to -340 (PR2) and -1 to +15 (PR3). Only PR1 and PR3 were selected for protein interaction test with HvMYB4H using electrophoretic mobility shift assay (EMSA) since the previously mentioned Sanger sequencing detected no genetic variation in PR2 (**Figure 3A**). Thus, recombinant HvMYB4H protein was expressed and purified using the *E. coli* expression system (**Figure S5**). Short DNA fragments encompassing PR1 and PR3 were cloned from Sloop (w) and Halcyon (b). EMSA tests showed that HvMYB4H could bind PR3 from both blue and white barley (**Figure S5**), however, the binding interaction with the blue *HvF35H* allele appears slightly stronger than interaction with the white allele. Weaker interaction was also detected for PR1 in Sloop (w). Taken together, these results indicate HvMYB4H is likely to regulate *HvF35H* with PR3 as the preferred target binding sites.

Synteny analyses indicate the presence of the MbHF35 cluster in barley, wheat D sub-genome and rye

Micro-synteny for the genomic region surrounding the candidate genes was assessed. Overall, the genes neighbouring the barley MbHF35 cluster were conserved and organized in syntenous blocks in other species (**Figure 4; Data S3**). As expected, synteny and conservation was highest for species belonging to the Triticeae tribe. Interestingly, direct orthologues for the MbHF35 cluster could only be found in members of the Pooideae

subfamily, and specifically only in some members of the Triticeae tribe (barley, sub-genome D of bread wheat, *A. tauschii* and rye). Noteworthy, the MbHF35 orthologues in rye were found at the same genetic position on chromosome 7R (**Data S3**), although fine genome annotation is not yet available for this species. In both *T. turgidum ssp. dicoccoides* and *T. urartu*, an orthologue was identified for *HvF35H* while orthologues of *HvMYC4H* and *HvMYB4H* were absent (**Data S3**). Interestingly, the MbHF35 clusters in barley and *A. tauschii* are located at two different genetic locations, distanced by ~ 47.36 Mb using barley chromosome 4H as the reference (**Figure S6**).

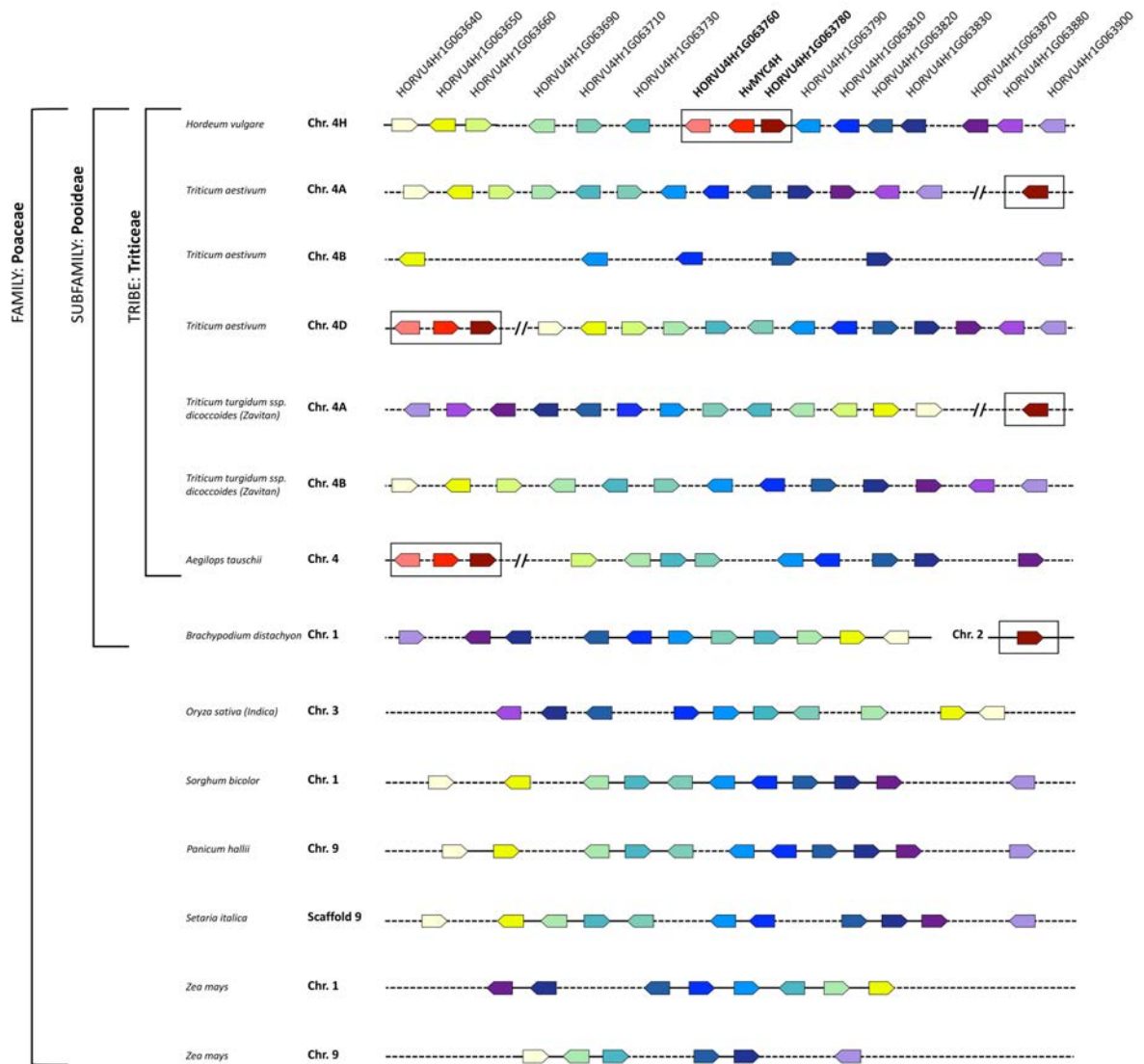


Figure 4. Shared synteny across *Poaceae* for genes surrounding the mapped interval. The species, taxonomic group and corresponding chromosome are indicated on the left. Gene names are indicated for barley. Genes are depicted by polygons; the direction of the polygons indicate the transcriptional orientation of the gene. Polygons of the same colour represent homologues across species. The black line between genes indicates contiguous genes, dashed lines indicate the presence of non-syntenic genes, whilst double slanted lines indicate gene clusters that are not physically proximal on the chromosome. The black box defines the genes within the MbHF35 cluster. Since no genome browser could be found for wheat and *T. turgidum* ssp. *dicoccoides*, genes are shown based on their relative position, but evenly distributed along a dashed line.

HvF35H belongs to a distinct F3'5'H subclade within monocots

F3'5'H and F3'H are close homologs, leading to the production of blue and red anthocyanins, respectively [33]. To investigate the evolutionary history of F3'5'H homologs, firstly, a comprehensive Neighbour Joining (NJ) tree based on the amino acid sequence alignment was constructed (**Figure S7**). In the NJ phylogeny, the F3'5'H lineage grouped into three distinct subclades with strong bootstrapping support: one Eudicot F3'5'H branch (Eu_F35H) and two Monocot F3'5'Hs branches (Mo_F35H1 and Mo_F35H2) (**Figure S7**). *HvF35H* fell within the Mo_F35H2 subclade. Noteworthy, this subclade tended to group with Eu_F35H with strong support (0.91; **Figure S7**), suggesting a closer relationship than with Mo_F35H1. This is supported by the amino acid sequence similarity calculation (**Data S4**): 54.0 % (\pm 3.2) between Eu_F35H and Mo_F35H2 compared to 50.79 % (\pm 4.6) between Mo_F35H2 and Mo_F35H1. To verify this observation, a separate Maximum Likelihood (ML) phylogeny was constructed. Also in this case Eu_F35H and Mo_F35H2 grouped together (0.70; **Figure S8**).

To further explore the evolutionary history of plant F3'5'H, a Bayesian phylogenetic tree based on the CDS sequence alignment was constructed (**Figure 5A**). Overall, the phylogeny has strong topology support. The three subclades (Eu_F35H, Mo_F35H1 and Mo_F35H2) can still be clearly recognized. In contrast to the NJ and ML phylogenies, Mo_F35H2 now grouped with Mo_F35H1, conforming to the known species phylogeny. A closer inspection shows that Mo_F35H1 sequences are present in all monocot species considered, while Mo_F35H2 diverged from Mo_F35H1 and are only found in specific members of the Pooideae family. Among these, the F3'5'Hs within the MbHF35 clusters in barley, wheat, *A. tauschii* and rye form a distinct branch from the other Mo_F35H2 homologs.

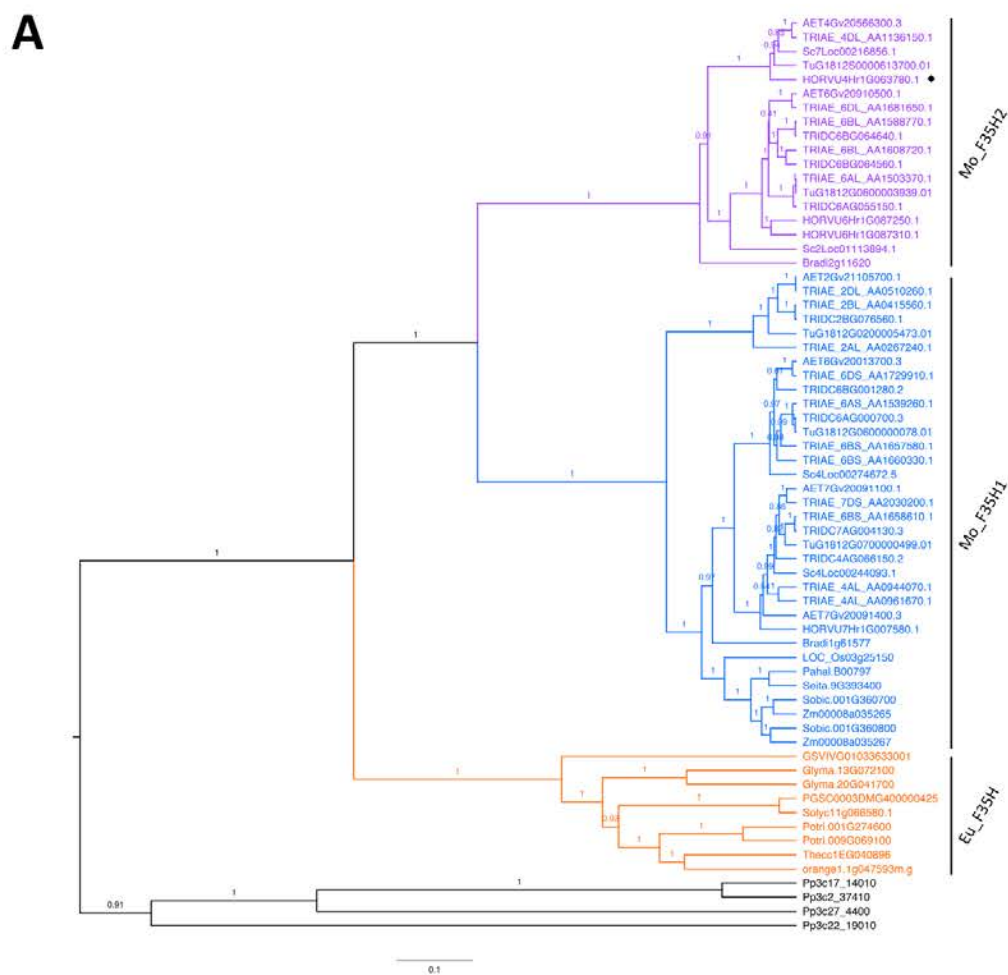
Phylogeny analyses reveal a shared evolutionary pattern for MYC, MYB and F3'5'H gene families in Triticeae

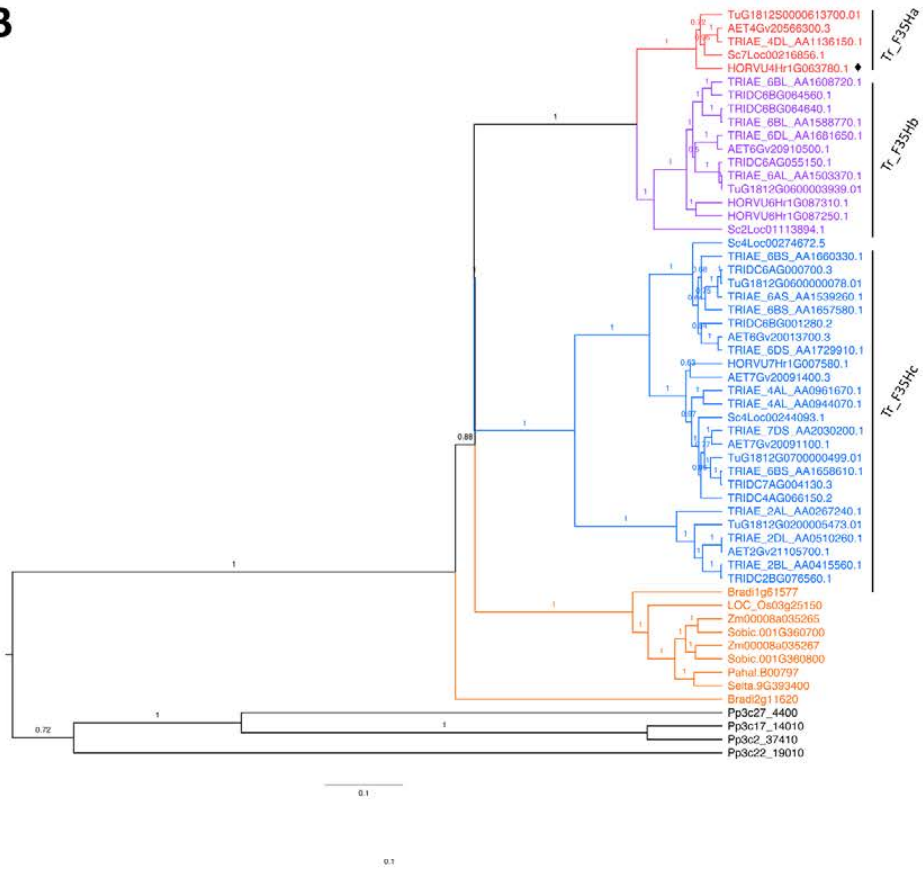
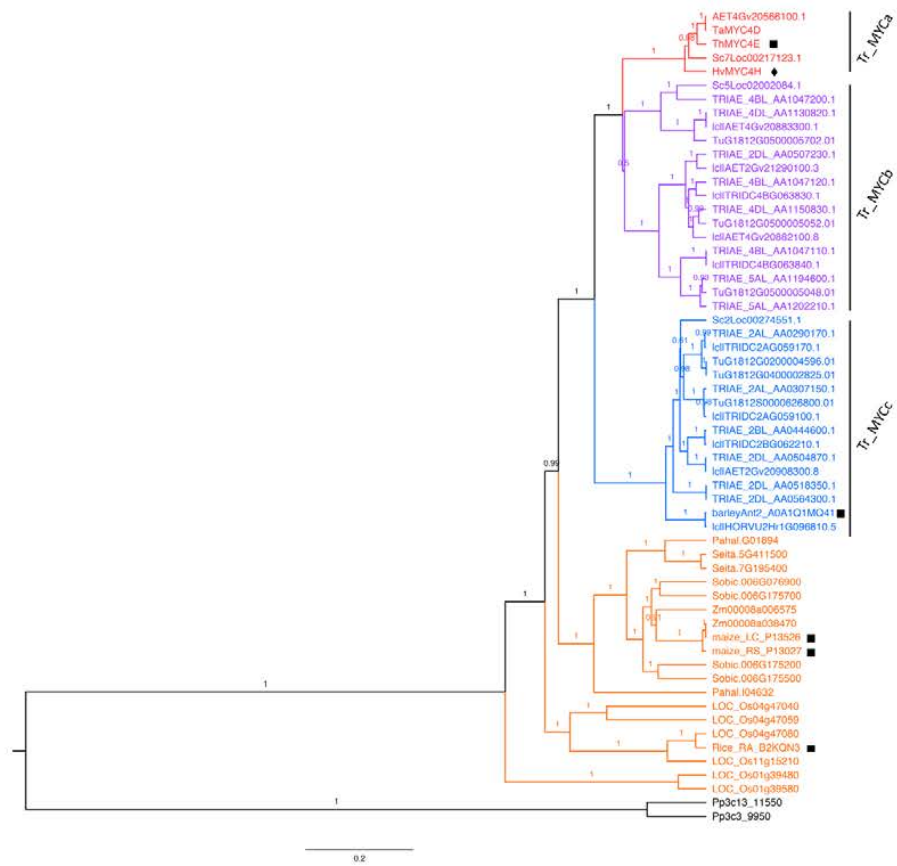
Due to the presence of a Mo_F35H2 lineage, the evolutionary relationship of the *Triticeae* tribe F3'5'Hs was resolved in a separate phylogeny, whereby Triticeae F3'5'Hs were fixed as a monophyletic group. As shown in **Figure 5B**, the HvF35H orthologues from barley, wheat, *A. tauschii*, rye and *T. urartu* were grouped in a single lineage (Tr_F35Ha). Tr_F35Ha grouped with another subclade Tr_F35Hb. Together, these subclades comprise the F35H2 genes identified in the Triticeae tribe. The remaining F3'5'H genes evolved into a separate lineage, Tr_F35Hc, that corresponds to the Mo_F35H1 group.

To similarly investigate the evolutionary origin of HvMYC4H, genuine orthologues of *HvMYC4H* were identified by a comprehensive NJ tree (**Figure S9**). *ThMYC4E*, *HvMYC4H* and *TaMYC4D* (wheat D-genome orthologue of *HvMYC4H*) grouped with the previously characterized rice *RA* [34], maize *RS* [35], maize *LC* [36] and barley *Ant2* [37] to form an independent branch (**Figure S9**). This branch was subsequently used for CDS-based Bayesian phylogeny construction. As shown in **Figure 5C**, the rice *RA* gene diverged first, followed by the Panicoideae MYCs. The most recently-evolved branches correspond to the Triticeae MYCs, which evolved into 3 major subclades: Tr_MYCa, Tr_MYCb and Tr_MYCc (**Figure 5C**). In subclade Tr_MYCa, the genes: *ThMYC4E*, *HvMYC4H*, *TaMYC4D*, *AET4Gv20566100.1* (*A. tauschii*) and *Sc7Loc00217123.1* (rye) formed a single lineage separated from the other wheat MYC homologues (Tr_MYCb). This indicates that Tr_MYCa diverged relatively late and has evolved independently in the Triticeae subfamily. Similar to that seen for F3'5'H Bayesian phylogeny, Tr_MYCa sequences were only present in barley, wheat, *A. tauschii* and rye.

The evolutionary history of *HvMYB4H* was analysed using a similar method to *HvMYC4H*. Genuine *HvMYB4H* orthologues were identified by a comprehensive NJ phylogeny (**Figure**

S10). As shown in **Figure 5D**, Panicoideae species and rice MYBs diverged first, with the previously characterized maize *C1* gene within this group. Then, Triticeae MYB homologs evolved into three major subclades: Tr_MYBa, Tr_MYBb and Tr_MYBc. These subclades matched well with the Tr_MYCa, Tr_MYCb and Tr_MYCc subclades. Notably, the Tr_MYBa clade covers the MYB genes of MbHF35 clusters in barley, wheat, *A. tauschii* and rye, and grouped with the Tr_MYBb subclade. In contrast, Tr_MYBc seemed to be more divergent from Tr_MYBa and Tr_MYBb.



B**C**

D

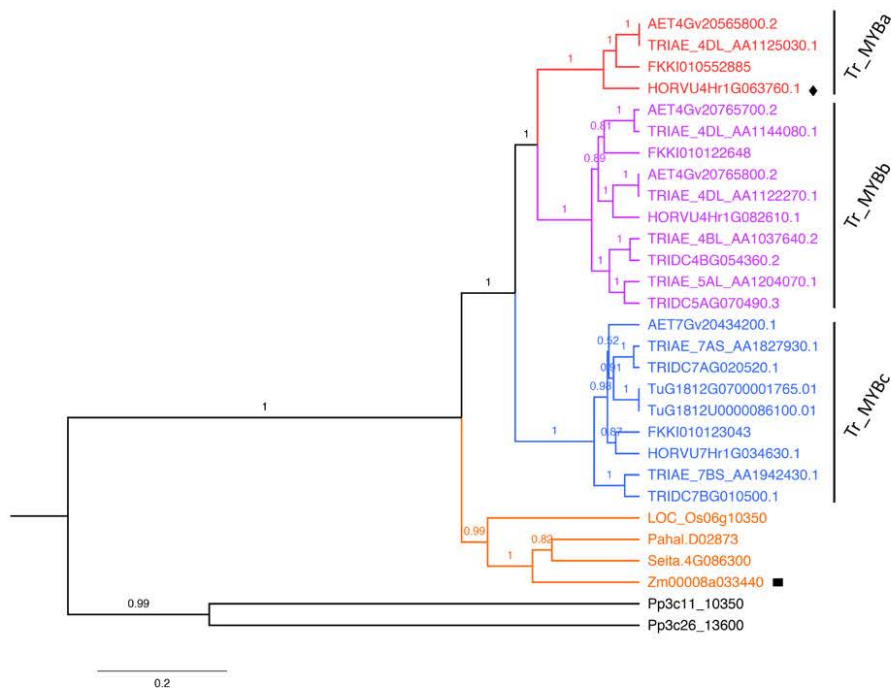


Figure 5. Phylogenetic analyses of the F3'5'H gene family in plants and of the MbHF35 cluster genes in monocots. **(A)** Bayesian phylogeny of monocot and eudicot F3'5'Hs based on CDS sequence alignment. Remote F3'5'H homologs from *Physcomitrella patens* were included as outgroup. **(B)** Bayesian phylogeny of monocot F3'5'H gene family. **(C)** Bayesian phylogeny of monocot *HvMYC4H* homologues. Previously characterised rice *RA*, maize *LC*, maize *RS*, barley *Ant2* and wheat *ThMYC4E* are indicated by ■. **(D)** Bayesian phylogeny of monocot *HvMYB4H* homologues based on CDS sequence alignment. Previously characterized maize *C1* (*Zm0008a03340*) is indicated by ■. For **A-D**, barley candidates within the MbHF35 cluster are indicated by ◆.

Positive selection is acting on Mo_F35H2 and Tr_MYCa

The ratio (ω) of non-synonymous (d_N) and synonymous (d_S) substitution is a useful parameter to assess the selection pressure on evolving genes, whereby $\omega < 1$, $\omega = 1$ and $\omega > 1$ indicate purifying selection, neutral evolution and positive selection, respectively. The ω values for Eu_F35H (ω [eudi]), Mo_F35H1 (ω [mono1]) and Mo_F35H2 (ω [mono2]) were calculated under different hypotheses (**Table 2; Data S5**). Branch-specific Likelihood-Ratio Tests (LRTs)

(**Table S2**) determined that the three-ratio model fits the dataset better than the two-ratio and one-ratio models. Under the three-ratio model, the ω values for Eu_F35H, Mo_F35H1 and Mo_F35H2 were estimated to be 0.13183, 0.00144 and 0.00163, respectively (**Table 2**). This suggests that all plant F3'5'Hs are under strong purifying selection, while F3'5'Hs in monocot plants tend to be more conserved than their eudicot counterparts.

Branch-site models, which allow the detection of positive selection on specific sites in specified branches, were applied to the same dataset (**Table 2**). ω assessments showed that 14.81 % ($P_2 + P_3$) of the amino acid sites in the Mo_F35H2 lineage were under positive selection, compared to only 6.08 % in the Mo_F35H1 subclade. Thirteen amino acid sites in the Mo_F35H2 branch were identified to be under positive selection ($P \leq 0.05$). Comparison with the neutral site-specific Model M1 showed that these sites are indeed under positive selection in the Mo_F35H2 lineage. In contrast, no amino site in Mo_F35H1 could be identified as significantly under positive selection (**Table 2**).

Similar selection tests were also performed on monocot MYCs (**Table S3**) and MYBs (**Table S4**). Under the best fitting models, the Tr_MYCa branch displayed a higher ω value (0.48319) than Tr_MYCb and Tr_MYCc (0.31972), whilst no significant difference in the ω value was found for Tr_MYBa, Tr_MYBb and Tr_MYBc ($\omega = 0.20890$), except a single amino acid in Tr_MYBb was identified to be under positive selection. Taken together, these results indicate that Tr_MYCa has been under higher selection pressure than the other MYC branches. All MYB branches underwent similar selection pressure, whilst slight positive selection was detected in Tr_MYBb, acting on a single amino acid.

Table 2. Natural selection tests on plant F3'5'Hs.

Model	Number of parameters	ln(Likelihood)	Estimates of parameters ^a ($\omega = d_N : d_S$; P - percentage of site)	Positively selected sites ^b
M0: one-ratio				
$\omega[\text{eudi}] = \omega[\text{mono1}] = \omega[\text{mono2}]$	1	-22638.96	$\omega[\text{eudi}] = \omega[\text{mono1}] = \omega[\text{mono2}] = 0.12122$	Not Allowed (NA)
Branch-specific models				
$\omega[\text{eudi}] = \omega[\text{mono1}] \neq \omega[\text{mono2}]$	2	-22623.89	$\omega[\text{eudi}] = \omega[\text{mono1}] = 0.12361$, $\omega[\text{mono2}] = 0.00165$	NA
$\omega[\text{eudi}] = \omega[\text{mono2}] \neq \omega[\text{mono1}]$	2	-22608.46	$\omega[\text{eudi}] = \omega[\text{mono2}] = 0.12609$, $\omega[\text{mono1}] = 0.00132$	NA
$\omega[\text{eudi}] \neq \omega[\text{mono1}] \neq \omega[\text{mono2}]$	3	-22574.05	$\omega[\text{eudi}] = 0.13288$, $\omega[\text{mono1}] = 0.00144$, $\omega[\text{mono2}] = 0.00171$	NA
Site-specific models				
Neutral M1 (2 site classes)	2	-22484.14	$P_0 = 0.92328$ ($P_2 = 1 - P_0 = 0.07672$); $\omega_0 = 0.10742$ ($\omega_1 = 1.0$)	NA
Selection M2 (3 site classes)	3	-22484.14	$P_0 = 0.92328$, $P_1 = 0.04721$ ($P_2 = 1 - P_0 - P_1 = 0.02951$); $\omega_0 = 0.10742$ ($\omega_1 = 1.0$), $\omega_2 = 1.0$	None
Branch-site models				
Model A Null (Mo_F35H1)	3	-22482.77	$P_0 = 0.88117$, $P_1 = 0.07313$, $P_2 + P_3 = 0.04570$; $\omega_0 = 0.10695$, $\omega_1 = 1.0$, $\omega_2 = 1.0$	NA
Model A (Mo_F35H1)	4	-22479.54	$P_0 = 0.86830$, $P_1 = 0.07200$, $P_2 + P_3 = 0.05970$; $\omega_0 = 0.10743$, $\omega_1 = 1.0$, $\omega_2 = 176.67$	None
Model A Null (Mo_F35H2)	3	-22476.19	$P_0 = 0.80006$, $P_1 = 0.06702$, $P_2 + P_3 = 0.13292$; $\omega_0 = 0.10609$, $\omega_1 = \omega_2 = 1.0$	NA
Model A (Mo_F35H2)	4	-22466.66	$P_0 = 0.78808$, $P_1 = 0.06664$, $P_2 + P_3 = 0.14528$; $\omega_0 = 0.10741$, $\omega_1 = 1.0$, $\omega_2 = 359.65080$	24R, 32L, 96R, 101E, 164D, 170R, 211T, 268N, 353Q, 378P, 394R, 460K, 468R

^a In the site-specific model M1, two site classes were specified: highly conserved sites (ω_0) and neutral sites ($\omega_1 = 1$). For the site-specific model M2, there were three site classes: highly conserved sites (ω_0), neutral sites ($\omega_1 = 1$) and positively selected sites (ω_2). In Model A, four site classes were specified. The first two classes had ω ratios of ω_0 and ω_1 respectively, corresponding to highly conserved sites and neutral sites across all lineages. In the other two site classes, the background lineages had ω_0 or ω_1 while the foreground lineages had ω_2 .

^b Positively selected amino acids at P -value ≤ 0.05 are numbered according to *HvF35H* (Uniprot No: A0A287PC56) excluding the first 30 amino acids predicted as membrane targeting signal.

Mo_F35H2 proteins have undergone convergent evolution toward Eu_F35H

To analyse the amino acid site substitutions in different F3'5'H subclades, ancestral amino acid sequences for Eu_F35H, Mo_F35H1 and Mo_F35H2 were reconstructed and analysed. Using Eu_F35H as the outgroup, Tajima's Relative Rate Tests (RRT) [38] showed that Mo_F35H2 have evolved significantly faster than Mo_F35H1 ($P = 0.03815$; **Figure S11**). In contrast, when Eu_F35H and Mo_F35H1 were compared using Mo_F35H2 as the outgroup, no significant difference in evolutionary rates were detected ($P = 0.22544$; **Figure S11**). Ancestral sequence alignment revealed 173 amino acid substitutions between Eu_F35H and Mo_F35H2, compared to 185 substitutions between Mo_F35H1 and Mo_F35H2 (**Figure 6A**), supporting a closer distance between Eu_F35H and Mo_F35H2. Noteworthy, Mo_F35H2 resembled Eu_F35H instead of Mo_F35H1 at 62 sites (**Figure 6A**). This suggests that Mo_F35H2 represents an intermediate state between Mo_F35H1 and Eu_F35H and support the hypothesis of convergent evolution in Mo_F35H2.

The overall conservation of the Cytochrome P450s 3D structure has led to identification of six substrate recognition sites (SRS) responsible for catalytic activity [39]. In the reconstructed ancestral sequences SRS1, SRS2 and SRS3 exhibited a higher degree of divergence while SRS4, SRS5 and SRS6 tended to be more conserved (**Figure 6A**). Furthermore, amino acid substitutions at SRS4 and SRS5 were more frequent in Mo_F35H1. Similar observations were seen by alignment of the original F3'5'H sequences (**Data S6**). Two amino acid sites (211T, 268N) within SRS3 and SRS4, respectively, were found to be affected by positive selection (**Figure 6A**).

Selection on Mo_F35H2 appears to be driven by protein thermostability

To investigate the environmental factor driving the positive selection on Mo_F35H2, protein structure model of HvF35H was performed. As shown in **Figure 6B**, the overall structure of HvF35H is well conserved with the Cytochrome P450 protein structure (Protein Data Bank: 4R1Z). The spatial locations of the 13 amino acid residues identified under positive selection were analysed. The two residues (211T, 268N) belonging to SRS3 and SRS4 were found located at the predicted catalytic centre, with their side-chains positioned proximal to the superimposed enzyme substrate (**Figure 6C**), suggesting a potential effect on enzyme activity. Interestingly, the other 11 residues were located on the exterior surface of the HvF35H structure (**Figure 6D-6E**). Ten (24R, 32L, 96R, 101E, 164D, 170R, 378P, 394R, 460K, 468R) of those surface residues belong to non-polar amino acids. It has been reported that protein surface amino acid composition is biased between thermophilic and mesophilic proteins [40]. Protein thermostability is owing to their preference to non-polar amino acids [41]. Another well-recognised contributing factor to protein thermostability is the formation of salt bridge [41,42], which only involves the non-covalent interaction between Arg/Lys (bases) and Asp/Glu (acids). Out of the 11 surface residues affected by positive selection in HvF35H, 8 residues (24R, 32L, 96R, 101E, 164D, 170R, 394R, 468R) have the potential to form salt bridge, of which 6 residues are (24R, 101E, 164D, 170R, 394R, 468R) predicted to form salt bridges in the HvF35H model (**Data S7**). Most of other Mo_F35H2 proteins resemble HvF35H at these amino acid sites, whilst Mo_F35H1 retain amino acids that don't have the salt bridging potential (**Data S7**). These observations are consistent with previous report that

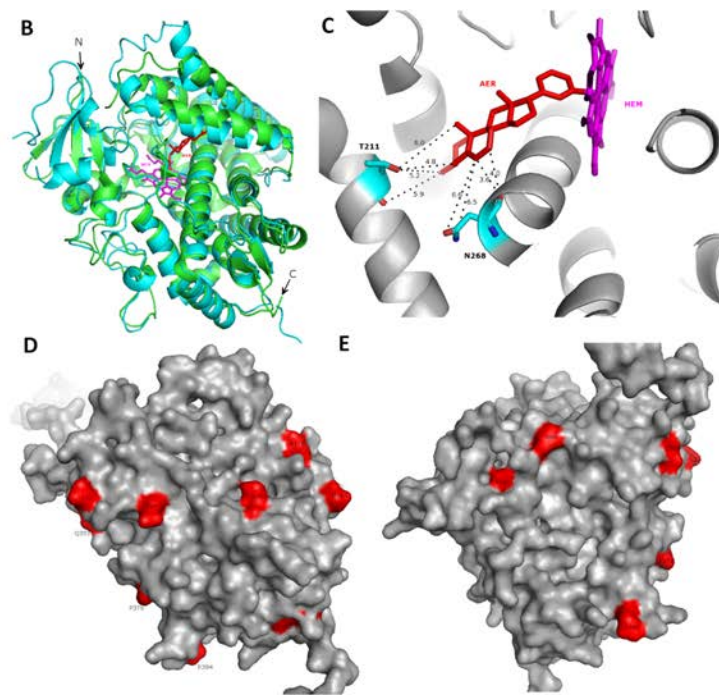


Figure 6. Ancestral sequences reconstruction and protein modelling for plant F3'5'Hs. **(A)** Alignment of the reconstructed ancestral sequences for Mo_F35H1, Mo_F35H2 and Eu_F35H. Secondary structural elements α -helices and β -strands are numbered and indicated by α and β , respectively. Strict α -turns are depicted as TTT and strict β -turns as TT. Substrate Recognition Sites (SRS) are indicated by \blacktriangle , whilst \star represent amino acids under positive selection in Mo_F35H2. Amino acid position known to reverse the enzyme activity between F3'5'H and F3'H are highlighted in yellow. **(B)** Superimposition of HvF35H model (cyan) with Zebrafish cytochrome P450 17A1 protein (4R1Z) (green). **(C)** Displays the 2 selected residues near the catalytic centre of HvF35H. **(D&E)** Display the locations of the other 11 selected residues. The superimposed substrates: AER (Abiraterone), HEM (heme). Amino acid site is numbered according to the predicted amino acid sequence of HvF35H, excluding the initial 30 amino acids.

Mo_F35H1 genes are preferentially expressed in vegetative tissues

To investigate if Mo_F35H1 genes share similar expression patterns to the Mo_F35H2 lineage, transcriptional databases were consulted (**Figure S12**). In barley, F35H1 genes were found to be almost exclusively expressed in roots and shoots, with very low expression in developing inflorescences and grains. Likewise in wheat, F35H1 genes for which data was available expression was confined predominantly in roots. According to the transcriptional

data in rice, the F35H1 expression level was also higher in roots, while remained low in seeds and barely detectable in the aleurone. A similar trend was also observed for the F35H1 gene in Sorghum (data not shown). In maize instead, F35H1 expression peaked in leaf tissue during the vegetative stage. Overall, Mo_F35H1 genes were found to be highly expressed in roots or other vegetative tissues, but only weakly expressed in grain.

Discussion

A tri-genic cluster underlies the *Blx1* locus in barley

Blue colouration in barley is associated by five complementary loci *Blx1-5* [30]. In the present study, we mapped *Blx1* to a physical interval of ~ 1.13 Mb on the long arm of chromosome 4H. This critical region is greatly reduced compared to previous mapping studies [43] and is positioned only ~ 26.13 Mb away from the acid soil tolerant gene *HvMATE* [31], corroborating the observed close linkage between the blue phenotype and *HvMATE*. A tri-genic cluster encompassing *HvMYB4H*, *HvMYC4H* and *HvF35H* (MbHF35 cluster) was identified at the mapped interval. The presence of these genes is consistent with anthocyanin accumulation: *HvMYC4H* and *HvMYB4H* belong to transcription factor classes which form the MBW complex to regulate anthocyanin biosynthesis [16], while *HvF35H* encodes a structural enzyme in anthocyanin production [17]. Of the genes identified in the MbHF35 cluster, *HvMYB4H* is proposed as a MYB TF involved in aleurone colour. In-vitro protein and DNA interaction tests showed that *HvMYB4H* is able to bind the putative promoter region of *HvF35H*, indicating *HvMYB4H* may be involved in the regulation of *HvF35H*. *HvMYC4H* is not annotated in the current genome assembly; it has been found due to its high sequence similarity to the recently published *ThMYC4E*, which has been proposed as candidate for blue aleurone phenotype in wheat [32]. An article published during the preparation of this manuscript looking at the regulatory components for blue aleurone also identified *HvMYC4H* [43]. The authors determined *HvMyc2* on chromosome 4H by sequence similarity to *HvAnt2*, as well as a F3'5'H gene in close proximity. *HvMyc2* and F3'5'H correspond to *HvMYC4H* and *HvF35H* identified in this study, thereby adding further evidence to their role in blue aleurone phenotype. Strygina, *et al.* (2017) also identified the MYB TF *HvMpc2* [43]. Results from the

present work tend to exclude *HvMpc2* as a candidate for the *Blx* loci as this gene falls outside the mapped interval. The blue aleurone mutant line (NGB20651) used in their study to characterise the blue aleurone trait is in fact an intense blue aleurone line, which contains the Intense blue aleurone locus *ibl1* (BGS 716). It's necessary for future studies to verify if *HvMpc2* corresponds *ibl1*.

Blue aleurone in wheat seems to share a similar genetic basis to barley. In wheat this trait is controlled by a single dominant locus *Ba1* that has been introgressed from wild relatives [29]. In the present study, we found that orthologues of *HvMYC4H*, *HvMYB4H* and *HvF35H* genes are also conserved in the wheat D sub-genome (designated as *TaMYC4D*, *TaMYB4D* and *TaF35H*). These genes also display higher expression levels in blue aleurone relative to white aleurone wheats. Considering the close genetic linkage of the genes within the MbHF35 cluster and their shared evolutionary pattern, we speculate that the *Ba1* locus in wheat may also involve the coordinate action of a tri-genic cluster. This study may serve as a guide for the identification of the genes responsible for blue aleurone in other species. For example, we observed the presence of the MbHF35 orthologues at a linked genetic location on chromosome 7R in rye. This is in accord with reports that blue aleurone in rye requires the dominant alleles at six loci (*vi1-vi6*), with *vi1* mapping to chromosome 7R [26,44]. Further studies are needed to verify if *vi1* corresponds to the MbHF35 homologues identified in rye. Likewise, a recent study has mapped the blue aleurone gene in *Thinopyrum ponticum*, the donor of the blue gene in blue aleurone wheat, to a small genetic segment on chromosome 4Ag [45].

Blue aleurone is a recently-evolved trait in the Triticeae

Aleurone colour is an interesting and useful agronomic trait in cereal plants. Depending on the type of anthocyanin accumulated in the aleurone cells, aleurone layer can assume different colorations. In maize, accumulation of pelargonidin and cyanidin pigments determine red or purple aleurones, respectively [22]. The production of cyanidin requires F3'H activity. Likewise, black rice accumulates abundant cyanidin and peonidin anthocyanins in the aleurone layer [20]. In contrast to other pigmentations, blue aleurone has only been observed in barley, wheat and rye. Blue coloured grains predominantly accumulate delphinidin-derived anthocyanins [21,25,26], which requires the function of F3'5'Hs. Indeed, by synteny and phylogeny we identified the MbHF35 cluster only in the Triticeae tribe. It is worth noticing that the *HvMYC4H* homolog in *A. tauschii*, is currently annotated as a low-confidence gene and is only a partial sequence. This indicates that a non-functional allele or a pseudogene is present in *A. tauschii*, thereby leaving barley, wheat and rye the only species with fully functional alleles for blue aleurone.

In Triticeae, the MbHF35 containing genes formed distinct phylogeny lineages (Tr_F35Ha, Tr_MYCa and Tr_MYBa), which diverge relatively late from Tr_F35Hb, Tr_MYCb and Tr_MYBb, respectively. The shared evolutionary pattern of the MbHF35 cluster, together with their physical linkage on chromosomes, support the hypothesis that they may have co-evolved together. Both the synteny and the colinear block analyses indicate that the MbHF35 cluster has resulted from gene insertion in strictly conserved genetic loci. Particularly, the MbHF35 clusters in barley and *A. tauschii* are positioned ~ 47.36 Mb away using barley chromosome 4H as reference, indicating the MbHF35 cluster has shifted at least one time either in barley or in *A. tauschii*. These findings are consistent with the MbHF35 cluster being

dispersed or proximal gene duplicates (**Appendix S3; Data S8**). We propose the blue aleurone trait to have emerged only recently in the common ancestor of barley and wheat D genome, which corresponds to a period about 13 million years ago (Mya) [46].

HvMYC4H and *HvMYB4H* may have undergone human selection during domestication

Transcriptional analyses show that *HvMYB4H*, *HvMYC4H* and *HvF35H* are significantly higher expressed in the grains of blue aleurone varieties, with all three genes showing a clear aleurone-specific expression by *in situ* hybridization. The aleurone-specific expression profile of *HvF35H* identified in the present study is corroborated by a recent publication [47], in which the *F3'5'H-1* (corresponding to *HvF35H*) was also found to be transcribed specifically in the aleurone cells of blue coloured barley. The expression patterns of *HvMYB4H*, *HvMYC4H* and *HvF35H* in blue and white aleurone barleys can potentially be explained by the distinct haplotypes observed. In particular, for *HvMYC4H* a 1 bp insertion was identified in the coding region of white aleurone cultivars. This frameshift mutation, also identified by [43], causes premature termination in protein translation, thus rendering the allele non-functional. Significant variations including a 4 bp indel were also identified in the 3'UTR of *HvMYB4H*, which may affect the transcriptional regulation of this gene. Speculatively, *HvMYB4H* and *HvMYC4H* may act upstream to *HvF35H* and as seen from qPCR results and EMSA. Although the 1-bp frame-shift indel in *HvMYC4H* may be the most critical mutation, we believe that this is not the sole reason for the white aleurone colour. Blue coloured grain contains a mixture of delphinidin- and cyanidin- type anthocyanins. It is known that the expression of blue colour is associated with a higher F3'5'H/F3'H expression ratio. Thus in this study, the

decreased transcriptional level of HvF35H in white barley should also play a critical role for white aleurone trait. Due to the close genetic linkage between the three genes, it is unlikely to discern by mapping the individual contribution of each gene to the blue aleurone phenotype. More studies are needed to address this hypothesis.

Interestingly, genotyping of a collection of barley landraces and wild barley originated from diverse geographic backgrounds (**Appendix S4**) showed that the identified blue alleles for *HvMYB4H* and *HvMYC4H* are strictly conserved in wild barley. The majority of these wild barley display blue aleurone. In addition to wild barley, blue alleles of *HvMYB4H* and *HvMYC4H* are conserved in all barley landraces displaying blue aleurone. In contrast, all the barley landraces with the white *HvMYC4H* and/or *HvMYB4H* alleles display the white aleurone phenotype (**Appendix S4**). These results, together with the strict conservation of *HvMYB4H* and *HvMYC4H* alleles in wild barley, indicate that white aleurone barley may have resulted from human's selection of the white *HvMYC4H* and *HvMYB4H* alleles during domestication and further breeding. This indication corroborates with previous reports that the composition of metabolites including anthocyanins in cereal grains have changed significantly during human selection [48].

Mo_F35H2 underwent convergent evolution and protein thermostability selection

The identification of a Mo_F35H2 lineage is an interesting finding in this study. In contrast to the commonly conserved Mo_F35H1 genes, Mo_F35H2 sequences were only present in Pooideae. We found that Mo_F35H2 displayed higher amino acid sequence similarity with Eu_F35H homologues. This observation indicates that convergent or parallel amino acid

sequence substitutions have occurred in the Mo_F35H2 and Eu_F35H lineages, which is confirmed by the ancestral sequence reconstruction. Although phenotypic convergence is a widespread observation in many aspects of plants and animals, examples of convergent evolution at the amino acid level are very limited, particularly in higher plants [49,50]. The data presented here demonstrates an example of adaptive convergent evolution for Mo_F35H2 in Pooideae. Despite Eu_F35H, Mo_F35H1 and Mo_F35H2 were all under strong purifying selection, the Mo_F35H2 lineage was identified to evolve significantly faster than Mo_F35H1. These observations suggest a critical role played by environmental factors in the evolution of delphinidin-derived anthocyanins in seeds. Indeed, modelling of Mo_F35H2 proteins implies a strong signal for selection of increased thermostability. These observations suggest heat or strong light may have been involved during the evolution of Mo_F35H2, which makes sense that Mo_F35H2 expressed in grains would be more exposed to light and heat effects than Mo_F35H1 in vegetative tissues (mainly root). It has been well-documented that some thermostable bacterium thrive extreme environmental conditions, for example high temperature and high pressure, through the adoption of thermostable proteins with biased amino acid composition [40-42]. Here, our study reports that a similar strategy may have been adopted by higher plants. This represents a significant advance on our understanding of environmental adaptation for plants. Climate change has been recognized as a great challenge for the sustainable production of many crops [51,52]. Recently, it was reported that the increases of heat and drought extreme events under climate change would substantially decrease world barley yield [53]. In this context, the potential association of blue coloured barley with heat stress adaptation may has an important implication for future barley breeding and production.

Gene duplication, functional divergence and environmental adaptation

Species specific expansion is a common phenomenon among various gene families, which have been suggested to play a central role for plant diversity [54]. In the present study, we showed that the regulatory network of anthocyanin biosynthesis in cereal plants, encompassing the MYB, MYC and F3'5'H genes, has undergone at least two divergence events in *Triticeae*. Recently, a similar observation has been made for the F3'H genes, another important structural gene in the anthocyanin biosynthetic pathway [55]. Together, these results may indicate a WGD event in the common ancestor of *Triticeae*. WGD can result in gene duplication and plant polyploidy, which plays an important role in the evolution of biological diversity [56]. The unique evolution of the genes within the MbHF35 cluster may have contributed to exclusive evolution of the blue aleurone trait in wheat, barley and rye. Speculatively, the absence of *HvMYC4H* and *HvMYB4H* orthologues in *T. urartu* and *T. turgidum* may have been due to gene loss after duplication, a common mechanism during gene evolution [57]. Following gene duplication, beneficial functional divergence could happen at both the gene expression level and the protein structural level [54,57]. In our study, survey of public transcriptional data suggests that Mo_F35H1 genes are mainly expressed in vegetative tissues, rather than in grains. This, together with the grain-specific expression of Mo_F35H2 genes, may explain the absence of delphinidin-derived anthocyanins in maize and rice grains [21]. Ancestral sequence and protein modelling analyses identified amino acid substitutions at the predicted substrate binding sites among different F3'5'H groups, suggesting a potential divergence in enzyme efficiency as well.

The emergence of the blue aleurone trait may be linked to adaptive advantages associated with accumulation of delphinidin-derived anthocyanin in the grains. Numerous studies report how flavonoid production is responsive to environmental stressors such as drought, frost, nutrition levels and strong UV exposure [2,6]. Indeed, the present study provides clear evidence that light exposure can significantly increase the TAC in Bowman barley grains. Flavonoid accumulation also plays a critical role in seed dormancy and viability [58,59]. This association suggests an important role for flavonoid accumulation in environmental adaptation, as plants adopt seed dormancy to avoid adverse temporary conditions. The association of blue aleurone trait with environmental adaptation also makes sense at the phytochemical level, when considering that blue anthocyanins have a broader light absorption spectrum and higher anti-oxidant capacity than red anthocyanins. This may have played a role in helping Pooideae occupy a much broader ecological niche than non-Pooideae plants such as rice and maize, for which the surrounding environmental conditions are relatively milder.

Conclusions

We have identified *HvMYB4H*, *HvMYC4H* and *HvF35H* as genes underlying the *Blx1* locus. We provide evidence that *HvMYB4H*, *HvMYC4H* and *HvF35H* have resulted from dispersed gene duplication and have co-evolved specifically in the Triticeae tribe. We demonstrate a rare example of convergent evolution for Mo_F35H2, which appears to be affected by protein thermostability selection. We speculate that blue aleurone trait has evolved quite recently as an environmental adaptation in barley, wheat and rye and this trait has been affected by human selection during barley and wheat domestication and breeding. Our study shed lights

on the evolutionary origin of blue anthocyanins in these cereal grains and highlights the need for further studies to dissect the contribution of each gene to the blue aleurone phenotype.

Experimental procedures

Plant materials and sampling

DH populations MN607/Vlamingh, Sloop/Halcyon, Clipper/Sahara and the F2 population Sahara/08S917N-260 were developed at the Department of Primary Industries and Regional Development (DPRID), Western Australia or the University of Adelaide. The NIL line was obtained from Professor Jerome Franckowiak. Other barley varieties derived from the germplasm collection at DPRID. Barley lines were grown in the glasshouse during June and October 2017. Developing seeds were sampled with three biological replicates (at least two spikes per sample). Leaf and stem samples were collected at the four-leaf stage. All samples were snap-frozen in liquid nitrogen and stored at -80 °C.

DNA extraction, genotyping, gene cloning and sequencing

Tissues were fine-ground using the TissueLyser (Qiagen, Germany) and DNA was extracted as described in Kovalchuk (2014) [60]. Genotyping PCR was performed in a Veriti™ 96-Well Thermal Cycler (Thermo Fisher Scientific, US) and were checked by 2 % agarose or 6 % polyacrylamide gel electrophoresis. KASP™ assays were designed using Primer Picker (LGC Genomics). Assays were prepared using the LGC Genomics SNPLine™ and assays were performed using KASP Master mix, following the manufacturer's instructions. For gene cloning and sequencing, target PCR fragments were purified using ISOLATE II PCR and Gel Kit (Bioline, Australia) for Sanger sequencing.

RNA extraction, cDNA synthesis and RT-PCR

Samples were ground into fine powder in a pre-cooled mortar with liquid nitrogen. For barley seeds, RNA was extracted as described by Wang *et al.* (2012) [61]. For leaf and stem, the ISOLATE II RNA Plant Kit (Bioline, Australia) was used. cDNA synthesis was carried out using SensiFAST™ (Bioline, Australia) following the manufacturer's instructions. SensiFAST™ SYBR No-ROX Kit (Bioline, Australia) was used for the RT-qPCR experiments. Each reaction contains 5 µL SensiFAST mix, 4.2 µL cDNA template, 0.8 µL forward/reverse primers (500 nM). RT-qPCR reaction was carried out using the ViiA7 Real-Time PCR System (Thermo Fisher, USA) in 384-well plates. The barley actin gene was used as reference gene. Three biological replicates were included for each time point. Each sample was run in two technical replicates. The transcription values were calculated using the comparative Ct method ($2^{-\Delta Ct}$) [62].

mRNA *in situ* hybridization

The target PCR fragment was cloned into pSPT18 and pSPT19 vectors using BamHI and SacI restriction sites. Digoxigenin (DIG)-labelled RNA probes were synthesized by *in vitro* transcription using T7 RNA polymerase from the DIG RNA Labelling Kit (Roche, US). Barley seed samples were fixed in 50 % ethanol, 5 % acetic acid and 3.7 % formaldehyde. The fixed-seeds were dehydrated in an ethanol and Histo-Clear series before embedding in Paraplast®. Microtome was used to cut longitudinal seed sections (6 µm) which were mounted on glass slides. *In situ* hybridization was carried out as described by Jackson (1991) [63] with slight modifications. Immunological detection was performed by incubation with Anti-Digoxigenin-AP (SIGMA, US). NBT-BCIP (SIGMA, US) was used for colour reaction according to the

manufacturer's instructions. All sections were visualized with the Olympus BX51 microscope system.

Synteny analysis

Orthologous genes were retrieved using BLASTP and/or TBLASTN against public sequence databases of other species: IWGSC (<https://www.wheatgenome.org>) for wheat; ATGSP (<http://aegilops.wheat.ucdavis.edu/ATGSP/>) for *A. tauschii*; WEWseq (<https://wheat.pw.usda.gov/GG3/wildemmer>) for *T. turgidum ssp. dicoccoides*; MBKBASE (<http://www.mbkbase.org/Tu/>) for *T. uratu*; IPK (<http://webblast.ipk-gatersleben.de/ryeselect/>) for rye; RGAP (<http://rice.plantbiology.msu.edu>) for rice; Gramene (http://ensembl.gramene.org/Zea_mays/Info/Index) for maize; and Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) for *B. distachyon*, *S. italica* and *S. bicolor*. Genes were confirmed by BLASTP/BLASTN back to barley, and by consulting databases of orthologous grouping: OrthoDB (<http://www.orthodb.org>) and EggNOG (<http://eggnogdb.embl.de/#/app/home>). All orthologous genes found in this way had an *E*-value < 2.3e-31.

Phylogeny inference

The amino acid sequences of HvF35H, HvMYC4H, HvMYB4H were used to query against genomic datasets of target species. The longest transcript was used when alternative transcripts were present. Sequence alignment was performed using MUSCLE [64]. The NJ phylogeny was inferred by MEGA7 v7.0 [65] using the p-distance substitution model with partial gap deletion (75 %). The branching support was assessed by the Interior-branch Test

method (1000 iterations). The Maximum Likelihood (ML) tree was inferred by MEGA7 v7.0 using JTT+G substitution model and was tested by bootstrapping (500 times). For Bayesian phylogeny, codon-based alignment of the coding sequences (CDS) were carried out using MUSCLE (8 iterations). Significant gaps and the predicted membrane-binding sites for F3'5'Hs were trimmed to reduce data noise. The phylogeny was searched by Bayesian simulations implemented in BEAST 2 [66] under strict molecular clock assumption using an unlinked substitution model Yule + G (5 categories). A single Markov Chain - Monte Carlo Chain was run for 1,500,000 generations with 1,000 pre burn-in until convergence. The final phylogenetic tree was inferred by TreeAnnotator [66] with the first 10,000 trees discarded. All phylogenetic trees in the present study were annotated using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Natural selection tests

Codon-based Maximum Likelihood estimates of ω (dN/dS) were implemented using codeml in PAML 4.7 package [67]. The CDS alignment files and Bayesian phylogenetic trees were used as input files. Branching pattern was specified using Treeview1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>). LRTs were performed to assess the significance of different hypotheses. *P*-values were calculated using the GraphPad software (<http://graphpad.com/quickcalcs/PValue1.cfm>).

Ancestral sequence reconstruction

Reconstruction of the ancestral sequences was performed using codeml in PAML 4.7 [67] based on the amino sequences back-translated from the CDS alignment files. The Empirical

Frequency model was used. The reconstructed sequences representing the ancestral status of Eu_F35H, Mo_F35H1 and Mo_F35H2 were used for Tajima's RRT [38] analyses implemented in MEGA7 v7.0 [65]. Substrate Recognition Sites (SRS) for plant F3'5'Hs were inferred based on sequence alignment as described by Gotoh (1992) [39] and Dueholm *et al.* (2015) [68].

Protein modelling

BLASTp was employed to identify homologous templates in the RCSB Protein Data Bank (<https://www.rcsb.org/>). The structural model of HvF35H was generated based on a combination of multiple structures of Cytochrome P450 proteins (chain A of PDB: 4R1Z, 4R20 and 4I8V). Protein modelling was carried out using the Modeller server, based on sequence alignment performed in Chimera (V1.12). The best models were chosen based on their lowest Discrete Optimized Protein Energy (DOPE) values and GA 341 score of 1, which indicate reliability of these models. The final model was validated by Ramachandran plot analysis using PROCHECK (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK>). The potential occurrence of salt bridge was predicted using the ESBRI tool [69]. Molecular visualizations were performed using PyMOL (Version 1.3r1. Schrodinger, LLC).

Accession numbers

The gene sequences in this study have been deposited at GenBank under accession No. MH618639-MH618662. Other data files are available in the Figshare data repository at <https://figshare.com/s/32dad83a1028b0405e38>.

Acknowledgements

We thank Professor Jerome Franckowiak (North Dakota State University) for providing the blue aleurone NIL, Gordon Thomson (Murdoch University) for his help with the mRNA *in situ* experiments, and Margaret Pallotta (The University of Adelaide) for critical revision of the manuscript.

Author contributions

CL and RW conceived and supervised the Research. YJ and CS wrote the manuscript. YJ and CS jointly performed the genetic mapping experiments. YJ and YZ undertook the developing seed sampling, RNA extraction, RT-PCR, mRNA *in situ* hybridization and anthocyanin extraction in barley. CS and RW are responsible for wheat seed sampling, RNA extraction, primer design and semi quantitative RT-PCR experiments. CS carried out the synteny analysis. YJ performed the phylogeny, gene duplication pattern, selection analyses, protein modelling, gene atlas data mining and EMSA. YJ, CS and BL performed gene sequencing. SB assisted with glasshouse trials. XZ, SW and TA performed genomic DNA extraction. PW and CT assisted with in-house genomic data retrieval.

Conflict of interests

The authors have no competing interests.

References

1. **Falcone Ferreyra ML, Rius SP, Casati P.** Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 2012;3:1–15.
2. **Mouradov A, Spangenberg G.** Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front. Plant Sci.* *Frontiers*; 2014;5:1–16.
3. **He J, Giusti MM.** Anthocyanins: natural colorants with health-promoting properties. *Annual Review of Food Science and Technology.* 2010;1:163–87.
4. **Winkel-Shirley B.** Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology.* 2001;126:485–93.
5. **Petroni K, Tonelli C.** Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science.* 2011;181:219–29.
6. **Agati G, Brunetti C, Di Ferdinando M, Ferrini F, Pollastri S, Tattini M.** Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant Physiol. Biochem.* 2013;72:35–45.
7. **Stolarzewicz IA, Ciekot J, Fabiszewska AU, Białecka-Florjanczyk E.** Plant and microbial sources of antioxidants. *Postepy higieny i medycyny doświadczalnej.* 2013;67:1359–73.
8. **Scott LC.** Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology.* 1999;70:1–9.
9. **Gould KS.** Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol.* 2004;2004:314–20.
10. **Lev-Yadun S, Gould KS.** Role of anthocyanins in plant defence. Gould K, Davies K, Winefield C, editors. *Anthocyanins: Biosynthesis, Functions, and Applications.* Springer New York; 2009. pp. 22–8.
11. **Buer CS, Imin N, Djordjevic MA.** Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* 2010;52:98–111.
12. **Hafidh RR, Abdulmir AS, Vern LS, Abu Bakar F, Abas F, Jahanshiri F, et al.** Inhibition of growth of highly resistant bacterial and fungal pathogens by a natural product. *Open Microbiol J.* 2011;5:96–106.
13. **Wang L-S, Stoner GD.** Anthocyanins and their role in cancer prevention. *Cancer Letters.* 2008;269:281–90.
14. **Lee Y-M, Yoon Y, Yoon H, Park H-M, Song S, Yeum K-J.** Dietary anthocyanins against obesity and inflammation. *Nutrients. Multidisciplinary Digital Publishing Institute*; 2017;9:1–15.
15. **Różańska D, Regulska-Ilow B.** The significance of anthocyanins in the prevention and treatment of type 2 diabetes. *Adv Clin Exp Med.* 2018;27:135–42.
16. **Xu W, Dubos C, Lepiniec L.** Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends in Plant Science.* 2015;20:176–85.
17. **He F, Mu L, Yan G-L, Liang N-N, Pan Q-H, Wang J, et al.** Biosynthesis of anthocyanins and their regulation in colored grapes. *Molecules.* 2010;15:9057–91.
18. **Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, et al.** Cytochromes P450. *The Arabidopsis Book.* The American Society of Plant Biologists; 2011;9.
19. **Bueno JM, Sáez-Plaza P, Ramos-Escudero F, Jiménez AM, Fett R, Asuero AG.** Analysis and antioxidant capacity of anthocyanin pigments. part ii: chemical structure, color, and intake of anthocyanins. *Critical Reviews in Analytical Chemistry.* 2012;42:126–51.
20. **Hu C, Zawistowski J, Ling W, Kitts DD.** Black rice (*Oryza sativa L. indica*) pigmented fraction suppresses both reactive oxygen species and nitric oxide in chemical and biological model systems. *J. Agric. Food Chem.* 2003;51:5271–7.
21. **Abdel-Aal E-SM, Young JC, Rabalski I.** Anthocyanin composition in black, blue, pink, purple, and red cereal grains. *J. Agric. Food Chem.* 2006;54:4696–704.
22. **Sharma M, Cortes-Cruz M, Ahern KR, McMullen M, Brutnell TP, Chopra S.** Identification of the *Pr1* gene product completes the anthocyanin biosynthesis pathway of maize. *Genetics.* 2011;188:69–79.
23. **Finch RA, Simpson E.** New colours and complementary colour genes in barley. *Zeitschrift fuer Pflanzenzuechtung.* 1978;81:40–53.
24. **Zeven AC.** Wheats with purple and blue grains: a review. *Euphytica.* 1991;56:243–58.
25. **Kim M-J, Hyun J-N, Kim J-A, Park J-C, Kim M-Y, Kim J-G, et al.** Relationship between phenolic compounds, anthocyanins content and antioxidant activity in colored barley germplasm. *J. Agric. Food Chem.* 2007;55:4802–9.
26. **Zykin PA, Andreeva EA, Lykholay AN, Tsvetkova NV, Voylokov AV.** anthocyanin composition and content in rye plants with different grain color. *Molecules.* 2018;23:1–12.

27. **Kushnak GD**. Utilizing linkages of genetic male sterile and aleurone color genes in hybrid barley (*Hordeum vulgare L.*) systems. Ph.D. thesis, Montana State University. 1974.
28. **Hucl P, Matus-Cadiz M**. Isolation distances for minimizing out-crossing in spring wheat. *Crop Science*. 2001;41:1348–51.
29. **Keppenne VD, Baenziger PS**. Inheritance of the blue aleurone trait in diverse wheat crosses. *Genome*. 1990;33:525–9.
30. **Shim JW, Suh SJ**. Linkage relationship of blue aleurone genes in barley. *Abst. Fifth Intl. Barley Genet. Symp., Okayama*. 1986;50.
31. **Furukawa J, Yamaji N, Wang H, Mitani N, Murata Y, Sato K, et al.** An aluminum-activated citrate transporter in barley. *Plant and Cell Physiology*. 2007;48:1081–91.
32. **Li N, Li S, Zhang K, Chen W, Zhang B, Wang D, et al.** *ThMYC4E*, candidate *Blue aleurone 1* gene controlling the associated trait in *Triticum aestivum*. *PLoS ONE*. 2017;12:1–13.
33. **Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, Menting JG, et al.** Cloning and expression of cytochrome P450 genes controlling flower colour. *Nature*. 1993;366:276–9.
34. **Hu J, Anderson B, Wessler SR**. Isolation and characterization of rice *R* genes: evidence for distinct evolutionary paths in rice and maize. *Genetics*. 1996;142:1021–31.
35. **Perrot GH, Cone KC**. Nucleotide-Sequence of the maize *R-S* Gene. *Nucleic Acids Research*. 1989;17:8003.
36. **Ludwig SR, Habera LF, Dellaporta SL, Wessler SR**. *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proc Natl Acad Sci USA*. 1989;86:7092–6.
37. **Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, et al.** Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc. Natl. Acad. Sci. U.S.A.* 2010;107:21611–6.
38. **Tajima F**. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics*. 1993;135:599–607.
39. **Gotoh O**. Substrate recognition sites in cytochrome p450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* 1992;267:83–90.
40. **Fukuchi S, Nishikawa K**. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *Journal of Molecular Biology*. 2001;309:835–43.
41. **Panja AS, Bandopadhyay B, Maiti S**. Protein thermostability is owing to their preferences to non-polar smaller volume amino acids, variations in residual physico-chemical properties and more salt-bridges. *Bhattacharjya S, editor. PLoS ONE*. 2015;10:1–21.
42. **Chakravarty S, Varadarajan R**. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry*. 2002;41:8152–61.
43. **Strygina KV, Börner A, Khlestkina EK**. Identification and characterization of regulatory network components for anthocyanin synthesis in barley aleurone. *BMC Plant Biol*. 2017;17:109–17.
44. **Voylov AV, Lykholay AN, Smirnov VG**. Genetic control of anthocyanin coloration in rye. *Russ J Genet Appl Res*. 2015;5:262–7.
45. **Liu L, Luo Q, Li H, Bin Li, Li Z, Zheng Q**. Physical mapping of the blue-grained gene from *Thinopyrum ponticum* chromosome 4Ag and development of blue-grain-related molecular markers and a FISH probe based on SLAF-seq technology. *Theor Appl Genet*. 2018;131:2359–70.
46. **Gaut BS**. Evolutionary dynamics of grass genomes. *New Phytol*. 2002;154:15–28.
47. **Vikhorev AV, Strygina KV, Khlestkina EK**. Duplicated flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes in barley genome. *PeerJ*. 2019;7:e6266.
48. **Sang T, Li J**. Molecular genetic basis of the domestication syndrome in cereals. *Cereal Genomics II*. Springer Netherlands; 2013. pp. 319–40.
49. **Zou Z, Zhang J**. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol Biol Evol*. 2015;32:2085–96.
50. **Washburn JD, Bird KA, Conant GC, Pires JC**. Convergent evolution and the origin of complex phenotypes in the age of systems biology. *International Journal of Plant Sciences*. 2016;177:305–18.
51. **Lobell DB, Schlenker W, Costa-Roberts J**. Climate trends and global crop production since 1980. *Science*. 2011;333:616–20.
52. **Wheeler T, Braun von J**. Climate change impacts on global food security. *Science*. 2013;341:508–13.

53. **Xie W, Xiong W, Pan J, Ali T, Cui Q, Guan D, et al.** Decreases in global beer supply due to extreme drought and heat. *Nature Plants*. 2018;4:964–73.
54. **Flagel LE, Wendel JF.** Gene duplication and evolutionary novelty in plants. *New Phytol*. 2009;183:557–64.
55. **Jia Y, Li B, Zhang Y, Zhang X, Xu Y, Li C.** Evolutionary dynamic analyses on monocot flavonoid 3'-hydroxylase gene family reveal evidence of plant-environment interaction. *BMC Plant Biol*. 2019;19:347–16.
56. **Van de Peer Y, Mizrachi E, Marchal K.** The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017;18:411–24.
57. **Zhang J.** Evolution by gene duplication: an update. *Trends in Ecology & Evolution*. 2003;18:292–8.
58. **Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, et al.** Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol*. 2006;57:405–30.
59. **Khlestkina EK.** The adaptive role of flavonoids: emphasis on cereals. *Cereal Research Communications*. 2013;41:185–98.
60. **Kovalchuk N.** High-throughput analysis pipeline for achieving simple low-copy wheat and barley transgenics. Fleury D, Whitford R, editors. *Crop Breeding: Methods and Protocols*. Springer New York; 2014. pp. 239–52.
61. **Wang G, Wang G, Zhang X, Wang F, Song R.** Isolation of high quality rna from cereal seeds containing high levels of starch. *Phytochem. Anal*. 2012;23:159–63.
62. **Schmittgen TD, Livak KJ.** Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*. 2008;3:1101–8.
63. **Jackson D.** *In situ* hybridization in plants. Molecular plant pathology: a practical approach. Oxford University Press. 1991.
64. **Edgar RC.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004;32:1792–7.
65. **Kumar S, Stecher G, Tamura K.** MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
66. **Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al.** BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Comput. Biol*. 2014;10:1–6.
67. **Yang Z.** PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*. 2007;24:1586–91.
68. **Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Taboureau O, et al.** Evolution of substrate recognition sites (SRSs) in cytochromes P450 from *Apiaceae* exemplified by the CYP71AJ subfamily. *BMC Evol. Biol*. 2015;15:122–14.
69. **Costantini S, Colonna G, Facchiano AM.** ESBRI: A web server for evaluating salt bridges in proteins. *Bioinformatics*. 2008;3:137–8.

Legends for Supporting Information

Appendix S1. Blue aleurone barley contains higher levels of delphinidin-derived anthocyanins.

Appendix S2. MbHF35 cluster genes are 35 significantly higher expressed in blue wheat.

Appendix S3. MbHF35 cluster genes result from dispersed gene duplication.

Appendix S4. *HvMYB4H* and *HvMYC4H* alleles are conserved in both wild barley and blue aleurone barley.

Supporting Method. Anthocyanin extraction and profiling.

Supporting Method. Wheat line sampling, RNA extraction, cDNA synthesis and semi RT-PCR.

Supporting Method. Heterologous expression and purification of recombinant HvMYB4H protein.

Supporting Method. Electrophoretic mobility shift assay.

Supporting Method. Promoter binding site prediction.

Supporting Method. Gene duplication pattern characterization.

Supporting Method. Gene expression data mining.

Figure S1. Schematic representation of the anthocyanin biosynthesis pathway.

Figure S2. Anthocyanin composition and profiling in selected barley cultivars.

Figure S3. Semi quantitative RT-PCR for wheat MbHF35 cluster genes on isolated aleurone tissue.

Figure S4. Whole seeds mRNA *in situ* hybridization of *HvF35H*.

Figure S5. Recombinant HvMYB4H purification and EMSA tests.

Figure S6. Comparison of colinear blocks containing the MbHF35 cluster in barley and *A. tauschii*.

Figure S7. Comprehensive Neighbour Joining (NJ) tree for plant F3'5'H and F3'H.

Figure S8. Maximum Likelihood (ML) phylogeny of plant F3'5'Hs based on the amino acid sequence alignment.

Figure S9. Comprehensive Neighbour Joining (NJ) tree of plant MYC family.

Figure S10. Comprehensive Neighbour Joining (NJ) tree of monocot MYB family.

Figure S11. Results from Tajima's test.

Figure S12. Transcriptional profiles of Mo_F35H1 genes.

Table S1. *In silico* promoter binding sites prediction for *HvF35H*.

Table S2. Likelihood Ratio Tests (LRTs) on the MbHF35 cluster for selection analyses.

Table S3. Natural selection tests on monocot MYC family.

Table S4. Natural selection tests on monocot MYB family.

Data S1. Genetic mapping results and primer list.

Data S2. Sequence alignment of *HvF35H*, *HvMYC4H* and *HvMYB4H* in blue and white cultivars.

Data S3. List of orthologues identified for the barley MbHF35 cluster and surrounding genes across monocot species.

Data S4. Calculation of the amino acid sequence identity across Eu_F35H, Mo_F35H1 and Mo_F35H2.

Data S5. Input and output data for selection test analyses.

Data S6. Amino acid sequence alignment of plant F3'5'Hs.

Data S7. Protein model of *HvF35H* and salt bridges prediction.

Data S8. Gene duplication pattern of monocot F3'5'H, MYB and MYC gene families.

References

Abdel-Aal ESM, Hucl P. A rapid method for quantifying total anthocyanins in blue aleurone and purple pericarp wheats. *Cereal Chem.* 1999;76:350-354.

Abdel-Aal ESM, Young JC, Rabalski I. Anthocyanin composition in black, blue, pink, purple, and red cereal grains. *J Agr Food Chem.* 2006;54:4696-4704.

Betts NS, Berkowitz O, Liu RJ, Collins HM, Skadhauge B, Dockter C, Burton RA, Whelan J, Fincher GB. Isolation of tissues and preservation of RNA from intact, germinated barley grain. *Plant Journal.* 2017;91:754-765.

Mrva K, Wallwork M, Mares DJ. alpha-Amylase and programmed cell death in aleurone of ripening wheat grains. *J Exp Bot.* 2006;57:877-885.

Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49

Zadoks JC, Chang TT, Konzak CF. Decimal code for growth stages of cereals. *Weed Res.* 1974;14:415-421.

Appendix S1. Blue aleurone barley contains higher levels of delphinidin-derived anthocyanins.

Total anthocyanin content (TAC) was extracted and compared for Halcyon (b), Sloop (w), BW063 (b), Bowman (w). As expected, TAC in blue aleurone varieties was significantly higher than that of white aleurone varieties (**Figure S2**). In particular, Halcyon (b) displayed the highest TAC concentration, which is ~ 26.5 % higher than BW063 (b). In contrast, white aleurone barley Bowman (w) had the lowest TAC, 17.7 % lower than Sloop (w). The potential effect of light exposure on anthocyanin accumulation was investigated. TAC was determined for Bowman (w) seeds exposed to light after the removal of husk following pollination. After light treatment, exposed Bowman seeds accumulated significantly higher TAC compared to intact Bowman (w) seeds.

The concentrations of six anthocyanins commonly present in cereal grains were also determined in blue and white aleurone barleys (**Figure S2**). The most abundant anthocyanins in blue varieties MN607 (b) and Halcyon (b) were malvidin-3-glucoside and delphinidin-3-glucoside, followed by a noticeable amount of malvidin-3-galactoside (malvidin also belongs to delphinidin-derived anthocyanins; **Figure S1**). Cyanidin-3-galactoside, cyanidin-3-glucoside and delphinidin-3-galactoside accumulated to a much lower level in these two varieties. Similarly, white varieties Vlamingh (w) and Sloop (w) contained high amounts of malvidin-3-glucoside and malvidin-3-galactoside, respectively; albeit to a lower degree compared to blue varieties. Vlamingh (w) and Sloop (w) also accumulated high levels of cyanidin-3-glucoside, which was found only in relatively low concentrations in MN607 (b) and Halcyon (b).

Appendix S2. MbHF35 cluster genes are significantly higher expressed in blue wheat.

The transcription of the three MbHF35 box genes in wheat (*TaF35H4D*, *TaMYC4D* and *TaMYB4D*) were tested by semi-quantitative RT-PCR on isolated aleurone tissue at 15 DPA, 19 DPA and 22 DPA. *TaF35H4D* was expressed in both white and blue wheat, with a stronger signal at 19 DPA and 22 DPA in blue aleurone wheat Sebesta Blue 3 (b) (**Figure S3**). For *TaMYC4D*, expression was only detected in Sebesta Blue (b) at 19 DPA and 22 DPA, but not at 15 DPA. No *TaMYC4D* expression was detected at any stage of seed development in white aleurone wheat Sebesta (w) (**Figure S3**). Similarly to *TaF35H4D*, *TaMYB4D* was also more strongly expressed at 19 DPA and 22 DPA in Sebesta Blue 3 (b) (**Figure S3**).

Appendix S3. MbHF35 cluster genes result from dispersed gene duplication.

The expansion patterns of the MbHF35 gene families were analysed in eight monocot species (**Data S8**). Four duplication mechanisms were specified: WGD (whole genome)/segmental duplication, tandem duplication, proximal duplication and dispersed duplication. The majority of monocot F3'5'H genes, including all F35H2 genes, were identified as either proximal or dispersed duplicates. No WGD/segmental duplication was identified for F3'5'Hs across all monocot species studied. In addition, Tr_MYCa and Tr_MYBa were all identified as dispersed duplicates (**Data S8**).

Appendix S4. *HvMYB4H* and *HvMYC4H* alleles are conserved in both wild barley and blue aleurone barley.

To explore the evolutionary origin of the blue and white alleles of *HvF35H*, *HvMYC4H* and *HvMYB4H*, indel and allele-specific Kompetitive Allele Specific PCR (KASP) markers were designed for *HvF35H*, *HvMYC4H* and *HvMYB4H*, targeting 58 the identified 6-bp and 5-bp indel for *HvF35H*, the 4-bp 3UTR indel for *HvMYB4H* and the 1-bp frame-shift indel for *HvMYC4H*, respectively. These markers were used to genotype 93 lines of wild barley (14) and landraces (79) collected from various geographic background. In addition, the genomic data of an additional 6 wild barley lines (4 from Israel and 2 from Tibet) were also searched from in-house genome resequencing database. Results showed that all of the wild barley (20) contained the blue haplotype for *HvMYB4H* and *HvMYC4H* (**Data S1G**). For *HvF35H*, interestingly, 16 wild lines have the white allele, whilst the other 4 lines displayed a different allele (termed L-allele) distinct from both the blue and white alleles identified. Of these 20 wild barleys, 15 displayed a blue aleurone phenotype, while the rest have red (2), black (2) and grey (1) seed colour. For the other barley landraces studied, 27 lines displayed blue aleurone phenotype. Genotyping results showed that all of these blue lines have the blue alleles for *HvMYB4H* and *HvMYC4H*, while the *HvF35H* alleles are diverse, including 19 blue, 3 white, 3 L-allele and 2 Morex alleles. These results indicate that *HvMYB4H* and *HvMYC4H* are strictly conserved in wild barley, while the three *HvF35H* alleles are all functional in blue aleurone barley. Notably, we also identified 10 landraces containing the blue *HvMYC4H* allele but with the white *HvMYB4H* and *HvF35H* alleles (**Data S1G**). All of these 10 landraces demonstrate white seed colour. In addition, 13 barley landraces with white aleurone were identified to contain the blue haplotype for *HvF35H*, *HvMYB4H* and *HvMYC4H*. This indicate that these lines may contain other recessive loci for the blue aleurone trait, similar as the case with BW064 (*blx3*) and BW065 (*blx4*).

Supporting Method. Anthocyanin extraction and profiling.

For Total Anthocyanin Content (TAC) measurement, mature grains were ground into fine powder using a coffee miller. Three grams of ground powder were used for anthocyanin extraction by following the extraction process described by Abdel-Aal *et al.* (2006). Three biological replicates were measured for each variety. The anthocyanin concentration was determined using a UV-vis spectrophotometer (Beckham, USA) by measuring the absorbance at 535 nm. The TAC was calculated as described by Abdel-Aal and Hucl (1999). Intact grain samples were sent for anthocyanin profiling by HPLC at Creative Proteomics, USA.

Supporting Method. Wheat line sampling, RNA extraction, cDNA synthesis and semi RT-PCR.

Wheat lines were grown under 12-hour photoperiod and 23 °C day/16 °C night temperatures. Anthesis was defined as Zadok stage Z61 (Zadoks *et al.*, 1974). Seeds were collected at 15, 19 and 22 DPA and aleurone was manually dissected using the method described by Mrva *et al.* (2006) with slight modification. All samples were snap-frozen in liquid nitrogen and stored at – 80 °C. For RNA extraction, samples were ground into fine powder in a pre-cooled mortar with liquid nitrogen. For wheat, isolated aleurone layers from 16 seeds were pooled together and total RNA extraction was performed following the method described by Betts *et al.* (2017). cDNA synthesis was carried out using ProtoScript II First Strand Synthesis (New England Biolab, USA) following the manufacturers' instructions. For semi-quantitative RT-PCR in wheat Q5 polymerase was used following manufacturer's protocol in a final reaction volume of 12.5 µL.

Supporting Method. Heterologous expression and purification of recombinant HvMYB4H protein.

The coding domain sequence of HvMYB4H gene was cloned from Halcyon using attattggatccCTTCGGCGGTGCGGCAAGAG (forward primer with BamHI site) and attattgagctcATCTACGCACGTTTGCGTCTC (reverse primer with SacI site). The amplified PCR fragment was verified by sanger sequencing and cloned into pET30a(+) vector (Novagen, Germany) using the included restriction enzyme sites. The constructed plasmid was transformed into BL21-CondonPlus-RP strain (Integrated Sciences, Australia). *In vivo* production of recombinant protein was carried out in 1 L flask. 100 mL of LB medium inoculated with *E. coli* cells harbouring the desired plasmids was incubated at 37 °C with shaking at 200 rpm until the OD₆₀₀ reached about 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were continuously grown overnight at 18 °C and harvested by centrifugation (25 minutes, 4000 rpm, 4 °C). The separated cell pellet was then suspended in 5 mL suspension buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM DTT) and

was sonicated on ice for 4 minutes with 1 minute interval every minute. Cell lysate was centrifuged at 14, 000 g, 4 °C for 25 minutes. Recombinant protein purification was performed using Profinity IMAC resin (Bio-Rad, USA) in a gravity flow column by following the manufacturer's instruction. The elution fraction containing the target recombinant protein was verified on 10 % SDS-PAGE gel using Coomassie brilliant blue R-250 staining.

Supporting Method. Electrophoretic mobility shift assay.

DNA fragments encompassing the PR1 and PR3 regions of Sloop, Halcyon were cloned by standard PCR using forward and reverse primers ACTTTGTGTACCATCGTGTGGA/GATCAGGTTTCAGGGTTTAAAG and GTGCAACACGAGATCGATTTTGC/AAGGGTCAGGTGCATGATGG. PCR products were purified using ISOLATE II PCR and Gel Kit (Bioline, Australia). EMSA tests were performed using Electrophoretic Mobility-Shift Assay (EMSA) Kit (ThermoFisher, US) by following the manufacture's instruction.

Supporting Method. Promoter binding site prediction.

The promoter binding sites for *HvF35H* were predicted using the online PlantTFDB database (http://plantregmap.cbi.pku.edu.cn/binding_site_prediction.php). The genetic region comprising 1078 bp upstream of the start codon was analysed (Morex as reference). The *P*-value threshold was set at 1e-4. Identified transcription factors and corresponding binding sites were sorted based on *P*-value. Amino acid sequences of the MYB and bHLH transcription factor hits were used as query sequences for TBLASTN against the reference barley genomic database Hv_IBSC_PGsb_v2 to identify corresponding gene IDs.

Supporting Method. Gene duplication pattern characterization.

Gene duplication patterns were characterised using the MCScanX package (Wang *et al.*, 2012). Amino acid sequence data and GTF data for *Z. mays*, *S. bicolor*, *P. hallii*, *O. sativa* and *B. distachyon* were downloaded from Phytozome databases (<https://phytozome.jgi.doe.gov/pz/portal.html#>). For *H. vulgare*, *T. turgidum ssp. dicoccoides* and *A. tauschii*, genomic data at http://webblast.ipk-gatersleben.de/barley_ibsc/, <https://www.dropbox.com/sh/3dm05grokhl0nbv/AAC3wvlYmAher8fY0srX3gX9a?dl=0> and <http://aegilops.wheat.ucdavis.edu/ATGSP/annotation/> was used. The original genomic data were further processed to generate input files for MCScanX. Intra- and inter-species genome comparisons

were performed using NCBI-BLAST-2.2.29 tool with an *E*-value threshold of 1e-05. Intra-genome all-vs-all BLAST was performed for gene duplication pattern identification.

Supporting Method. Gene expression data mining.

Transcriptional data for the target homologous genes was extracted from individual databases: for wheat (<https://wheat.pw.usda.gov/WheatExp/>), barley (<https://apex.ipkgatersleben.de/apex/f?p=284:10>), rice (<http://expression.147ic4r.org/index>), maize (<https://www.maizegdb.org/>) and sorghum (<http://sorghum.riken.jp/morokoshi/Home.html>).

Figure S1. Schematic representation of the anthocyanin biosynthesis pathway. Abbreviations of the enzymes: **CHS** - chalcone synthase; **CHI** - chalcone isomerase; **FNS** - flavone synthase; **FHT** - flavanone 3-hydroxylase; **F3'H** - flavonoid 3'-hydroxylase; **F3'5'H** - flavonoid 3',5'-hydroxylase; **FLS** – flavonol synthase; **DFR** - dihydroflavonol 4-reductase; **LAR** - leucoanthocyanidin 4-reductase; **ANS** - anthocyanidin synthase; **FGT** - flavonoid 3-O-glucosyltransferase; **AMT** - anthocyanin methyltransferase.

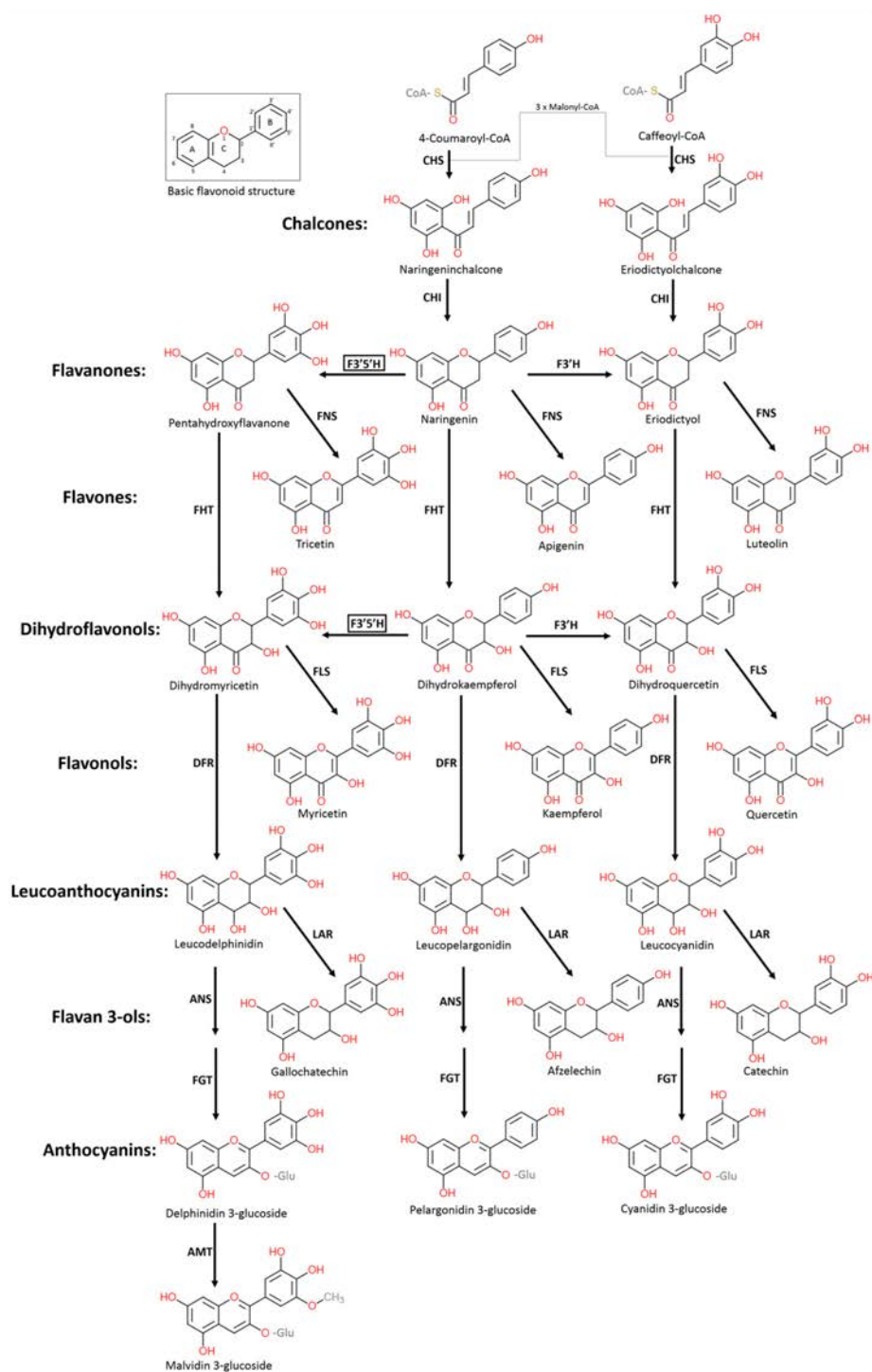


Figure S2. Anthocyanin composition and profiling in selected barley cultivars. **(a)** Mean Total Anthocyanin Content (TAC) in blue (Halcyon, BW063) and white (Sloop, Bowman) aleurone barleys. Error bars represent \pm SE, $n=3$. Student's two-tailed t -test P -values ≤ 0.01 are indicated by **. **(b)** Anthocyanin composition in blue (MN607, Halcyon) and white (Sloop, Vlamingh) aleurone barleys, $n=1$. Cyanidin-3-galactoside and cyanidin-3-glucosides are red anthocyanins, while the other anthocyanins are blue.

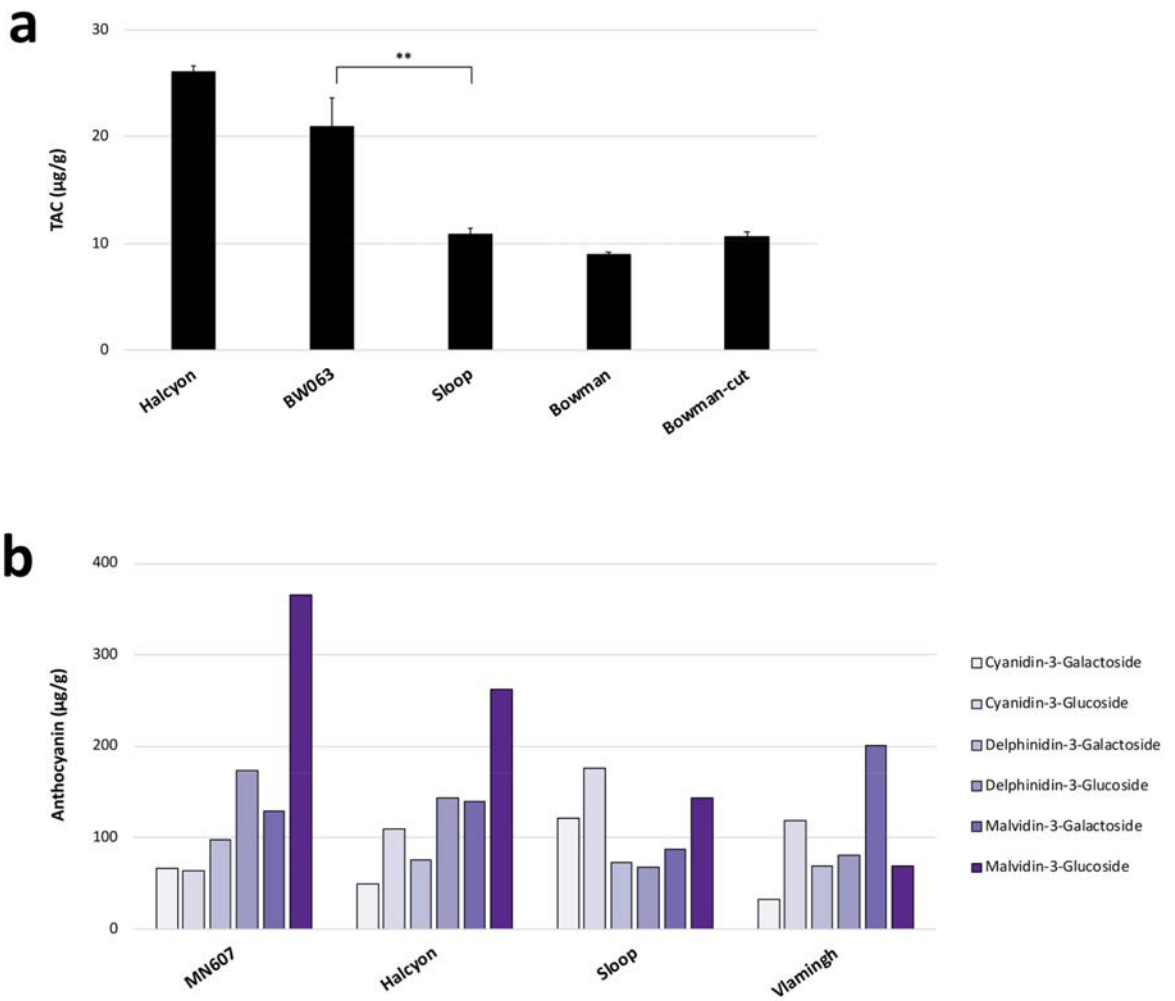


Figure S3. Semi quantitative RT-PCR for wheat MbHF35 cluster genes on isolated aleurone tissue. **(a)** *TaF35H4D*, **(b)** *TaMYC4D*, **(c)** *TaMYB4D* and **(d)** *TaGAPDH*. For **(a-d)** lanes 1-3 contain tissue from “Sebesta” at 15, 19 and 22 Days Post Anthesis (DPA), respectively. Lanes 4-6 contain tissue from “Sebesta Blue 3” at 15, 19 and 22 DPA, respectively. Lane 7 represents genomic DNA from “Sebesta” leaf tissue, whilst lane 8 represents genomic DNA from “Sebesta Blue 3” leaf tissue. Lane 9 indicates No Template Control (NTC).

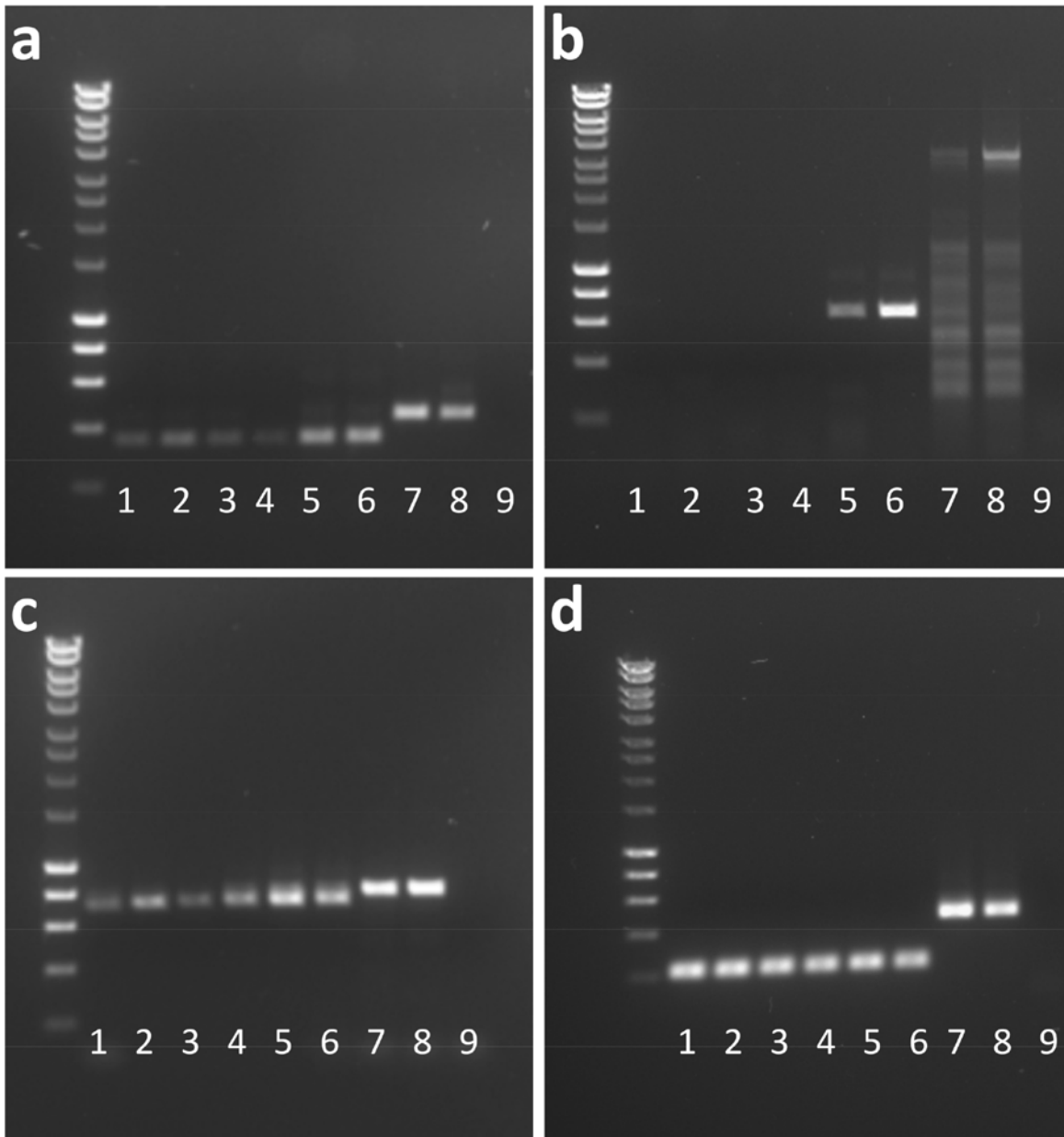


Figure S4. Whole seeds mRNA *in situ* hybridization of *HvF35H*. (a) *HvF35H*; (b) *HvMYB4H*; (c) *HvMYC4H*. Developing blue barley seeds (Halcyon, 28 DPA) were tested with antisense and sense probes.

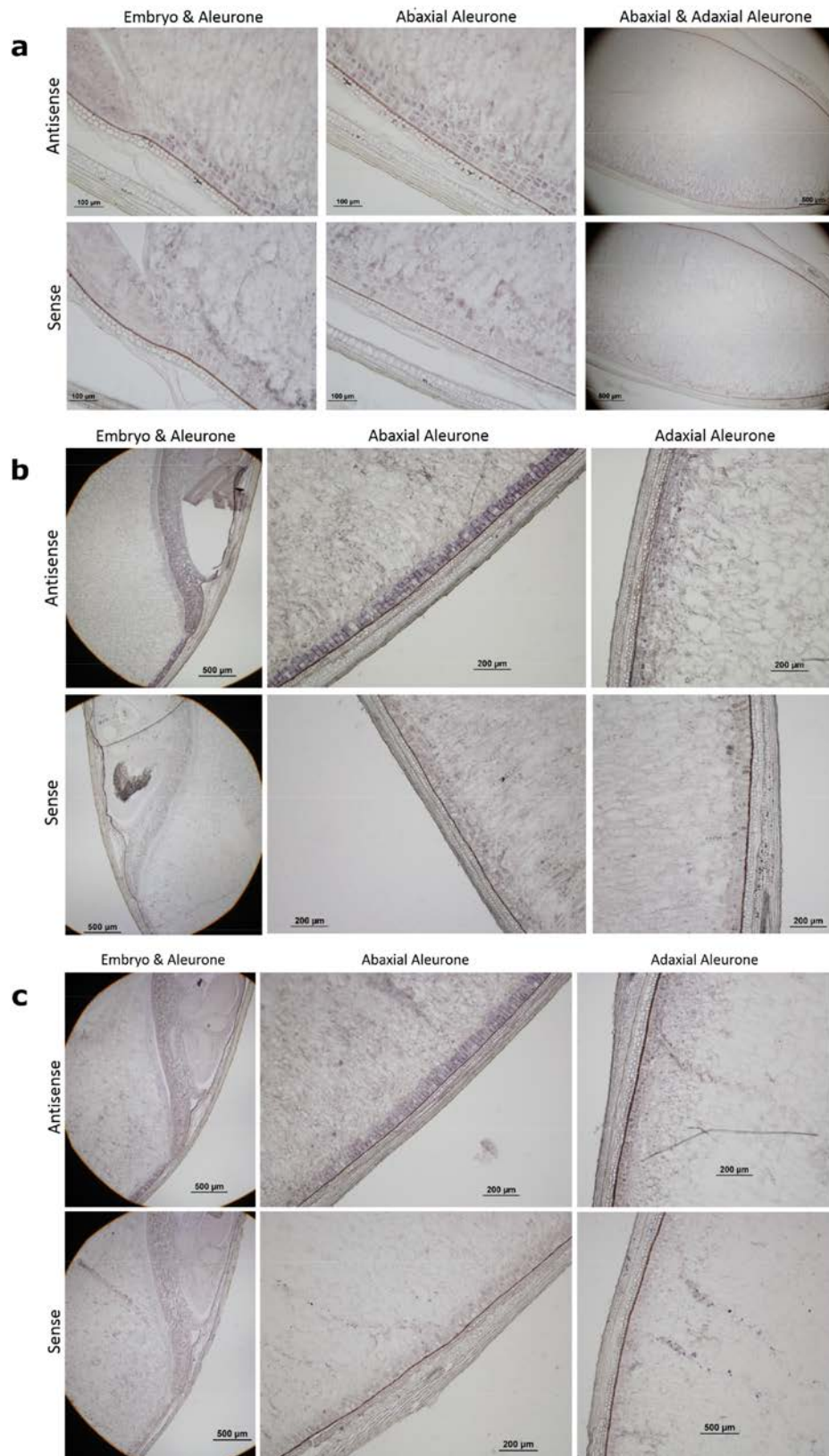


Figure S5. Recombinant HvMYB4H purification and EMSA tests. **(a)** SDS-PAGE gel check of the purified HvMYB4H protein. Lanes A: marker, B: supernatant, C: pellet, D: washing, E-H: elution fractions. Red arrow indicates the position of the expressed HvMYB4H. **(b)** EMSA tests of HvMYB4H protein with the putative promoter regions (PR1 & PR3) of *HvF35H* from Halcyon (blue allele) and Sloop (white allele). Lanes 1: no protein (control), 2-6: increasing protein from 1.0 μ L to 7.5 μ L, 7: no DNA (control).

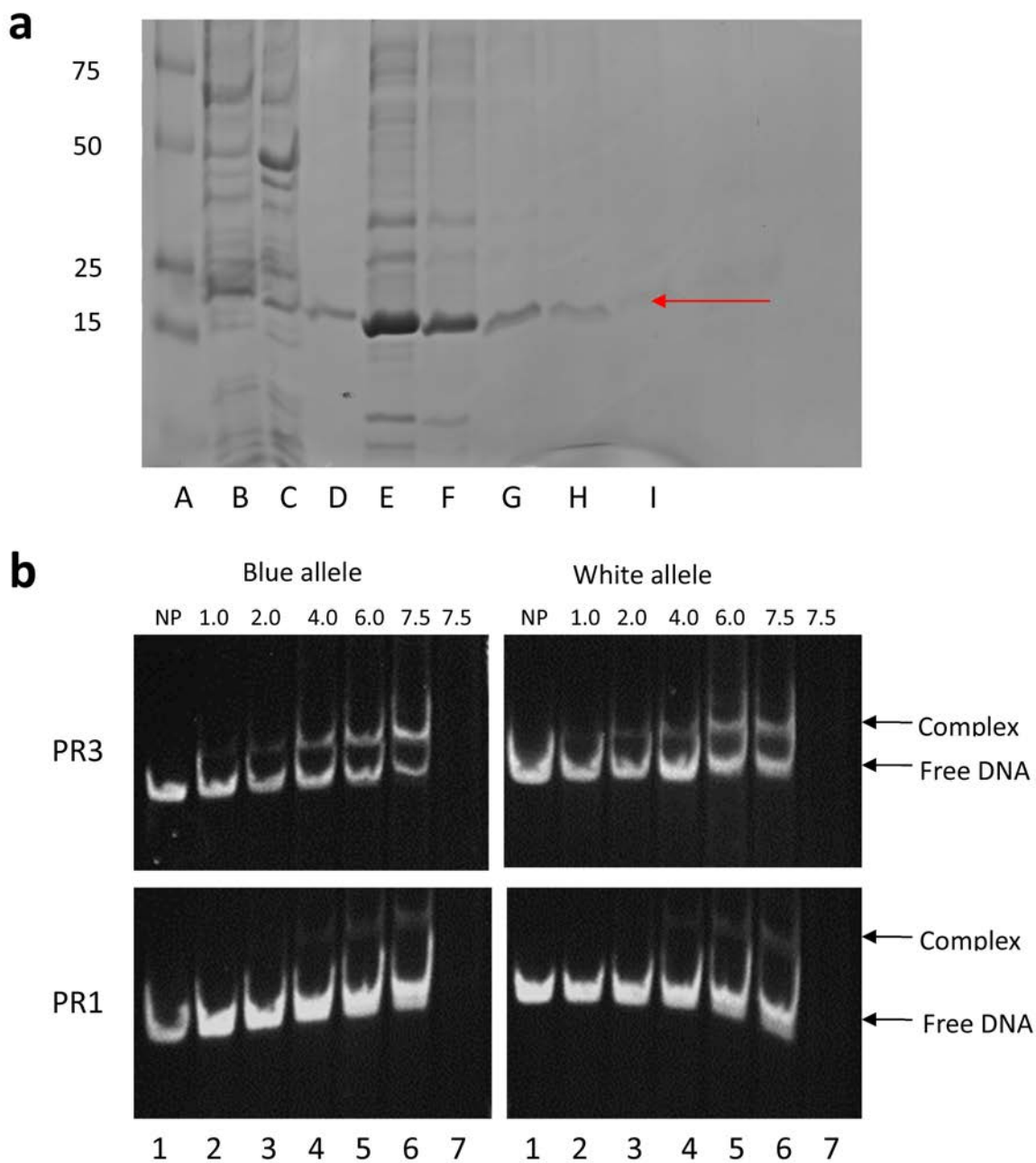


Figure S6. Comparison of colinear blocks containing the MbHF35 cluster in barley and *A. tauschii*. **(a)** Colinear block containing the *A. tauschii* MbHF35 cluster. **(b)** Position of barley MbHF35 cluster using *A. tauschii* as reference. **(c)** Colinear block containing the barley MbHF35 cluster using barley chromosome 4H as reference. The collinearity graph was generated using the MCScanX package. Species names were indicated on top. The first column refers the reference chromosome. Dark red shading indicates tandem duplicate genes. Yellow shading represents colinear genes. Light red shading indicates the location of the MbHF35 cluster.

	<i>A. tauschii</i>	<i>H. vulgare</i>	<i>B. distachyon</i>	<i>O. sativa</i>	<i>S. bicolor</i>	<i>P. hallii</i>	<i>S. italica</i>	<i>Z. mays</i>
a	AET4Gv20563600.4	HORVU4Hr1G062820.1	Bradi1g67950	LOC_Os03g14700	Sobic.001G435300	Pahal.A01365	Seita.9G467900	
	AET4Gv20563700.7	HORVU4Hr1G062880.3	Bradi1g67960	LOC_Os03g14690	Sobic.001G435400	Pahal.A01364	Seita.9G468000	Zm00008a035684
	AET4Gv20563900.1	HORVU4Hr1G062890.1	Bradi1g67980	LOC_Os03g14669	Sobic.001G435500	Pahal.A01363	Seita.9G468100	
	AET4Gv20564300.1	HORVU4Hr1G062920.2	Bradi1g67990					
	AET4Gv20564500.7	HORVU4Hr1G062930.3	Bradi1g68000	LOC_Os03g14610	Sobic.001G435700	Pahal.A01361	Seita.9G468300	Zm00008a035685
	AET4Gv20564800.1							Zm00008a035686
	AET4Gv20564700.5	HORVU4Hr1G062940.1	Bradi1g68010	LOC_Os03g14590	Sobic.001G435800	Pahal.A01360	Seita.9G468400	
	AET4Gv20564900.1							
	AET4Gv20565000.1	HORVU4Hr1G062970.3	Bradi1g68030	LOC_Os03g14570	Sobic.001G436000	Pahal.A01358	Seita.9G468600	
	AET4Gv20565200.2		Bradi1g68040	LOC_Os03g14540	Sobic.001G436100	Pahal.A01357	Seita.9G468700	Zm00008a035688
	AET4Gv20565300.3	HORVU4Hr1G063000.1	Bradi1g68050	LOC_Os03g14520	Sobic.001G436300	Pahal.A01355	Seita.9G468900	
	AET4Gv20565400.1	HORVU4Hr1G063020.9	Bradi1g68060	LOC_Os03g14510	Sobic.001G436400	Pahal.A01354	Seita.9G469000	
	AET4Gv20565600.2	HORVU4Hr1G063050.6	Bradi1g68070		Sobic.001G436500	Pahal.A01353	Seita.9G469100	Zm00008a035690
	AET4Gv20565700.7	HORVU4Hr1G063060.8						
	AET4Gv20565800.2							
	AET4Gv20566100.1							
	AET4Gv20566300.3							
	AET4Gv20566500.2		Bradi1g68110	LOC_Os03g14380	Sobic.001G436900	Pahal.A01346	Seita.9G469500	
	AET4Gv20566700.2		Bradi1g68120	LOC_Os03g14370	Sobic.001G437000	Pahal.A01345	Seita.9G469600	Zm00008a035691
	AET4Gv20566800.1	HORVU4Hr1G063140.1	Bradi1g68140	LOC_Os03g14290	Sobic.001G437100	Pahal.A01344	Seita.9G469800	Zm00008a035692
AET4Gv20567100.4	HORVU4Hr1G063150.3	Bradi1g68160	LOC_Os03g14280	Sobic.001G437200	Pahal.A01343	Seita.9G469900	Zm00008a035694	
AET4Gv20567200.4	HORVU4Hr1G063160.1	Bradi1g68170	LOC_Os03g14270	Sobic.001G437300	Pahal.A01342	Seita.9G470000	Zm00008a035695	
b	AET4Gv20572800.1	HORVU4Hr1G063640.1		LOC_Os03g14090	Sobic.001G438800	Pahal.A01327	Seita.9G471500	Zm00008a035703
	AET4Gv20572900.6	HORVU4Hr1G063650.1	Bradi1g68350	LOC_Os03g14080	Sobic.001G439000	Pahal.A01326	Seita.9G471700	
	AET4Gv20573100.1							
	AET4Gv20573500.2	HORVU4Hr1G063660.3		LOC_Os03g14030	Sobic.001G439100	Pahal.A01324	Seita.9G471800	Zm00008a035704
	AET4Gv20573600.6	HORVU4Hr1G063690.3						
	AET4Gv20573900.21	HORVU4Hr1G063710.12		LOC_Os03g14020	Sobic.001G439300	Pahal.A01322	Seita.9G472000	Zm00008a035706
	AET4Gv20574300.3	HORVU4Hr1G063790.3		LOC_Os03g14010	Sobic.001G439400	Pahal.A01320	Seita.9G472200	
	AET4Gv20574400.6	HORVU4Hr1G063810.1		LOC_Os03g14000	Sobic.001G439500	Pahal.A01319	Seita.9G472300	
	AET4Gv20574600.2	HORVU4Hr1G063820.1		LOC_Os03g13970	Sobic.001G439600	Pahal.A01318	Seita.9G472500	Zm00008a035707
	AET4Gv20574700.1							
	AET4Gv20574800.2	HORVU4Hr1G063830.8		LOC_Os03g13950	Sobic.001G439700	Pahal.A01317	Seita.9G472600	Zm00008a035708
	AET4Gv20575100.1							
	AET4Gv20575200.1	HORVU4Hr1G063870.1		LOC_Os03g13930	Sobic.001G439800	Pahal.A01316	Seita.9G472700	
	AET4Gv20575500.3	HORVU4Hr1G063880.5						
	AET4Gv20575600.1		Bradi1g68395					
	AET4Gv20575700.1				Sobic.001G439900	Pahal.A01315	Seita.9G472800	Zm00008a035709
	AET4Gv20575800.13	HORVU4Hr1G063900.2		LOC_Os03g13860	Sobic.001G440000	Pahal.A01314	Seita.9G472900	Zm00008a035710
AET4Gv20575900.2	HORVU4Hr1G063910.14		LOC_Os03g13850	Sobic.001G440100	Pahal.A01313	Seita.9G473000	Zm00008a035711	
c	HORVU4Hr1G063640.1	AET4Gv20572800.1		LOC_Os03g14090	Sobic.001G438800	Pahal.A01327	Seita.9G471500	Zm00008a035703
	HORVU4Hr1G063650.1	AET4Gv20572900.6	Bradi1g68350	LOC_Os03g14080	Sobic.001G439000	Pahal.A01326	Seita.9G471700	
	HORVU4Hr1G063660.3	AET4Gv20573500.2						
	HORVU4Hr1G063680.1							
	HORVU4Hr1G063690.3	AET4Gv20573700.3	Bradi1g68330		Sobic.001G439100	Pahal.A01324	Seita.9G471800	Zm00008a035704
	HORVU4Hr1G063710.12	AET4Gv20573900.21	Bradi1g68320		Sobic.001G439300	Pahal.A01322	Seita.9G472000	Zm00008a035706
	HORVU4Hr1G063730.2			LOC_Os03g14050				
	HORVU4Hr1G063760.1							
	HORVU4Hr1G063770.2							
	HvMYC4H							
	HORVU4Hr1G063780.1							
HORVU4Hr1G063790.3	AET4Gv20574300.3	Bradi1g68310	LOC_Os03g14010	Sobic.001G439400	Pahal.A01320	Seita.9G472200		
HORVU4Hr1G063810.1	AET4Gv20574400.6	Bradi1g68300	LOC_Os03g14000	Sobic.001G439500	Pahal.A01319	Seita.9G472300		
HORVU4Hr1G063820.1	AET4Gv20574600.2	Bradi1g68290	LOC_Os03g13970	Sobic.001G439600	Pahal.A01318	Seita.9G472500	Zm00008a035707	
HORVU4Hr1G063860.1								
HORVU4Hr1G063830.8	AET4Gv20574800.2	Bradi1g68280	LOC_Os03g13950	Sobic.001G439700	Pahal.A01317	Seita.9G472600	Zm00008a035708	
HORVU4Hr1G063870.1	AET4Gv20575200.1	Bradi1g68270	LOC_Os03g13930	Sobic.001G439800	Pahal.A01316	Seita.9G472700		

Figure S7. Comprehensive Neighbour Joining (NJ) tree for plant F3'5'H and F3'H. The phylogeny was developed using p-distance substitution model. Branch support (interior branching 1000 times) was labelled above each branch. Sequence IDs for previously characterised plant F3'5'Hs and F3'Hs are named according to the Uniprot format. HvF35H (HORVU4Hr1G063780.1) is marked with *.

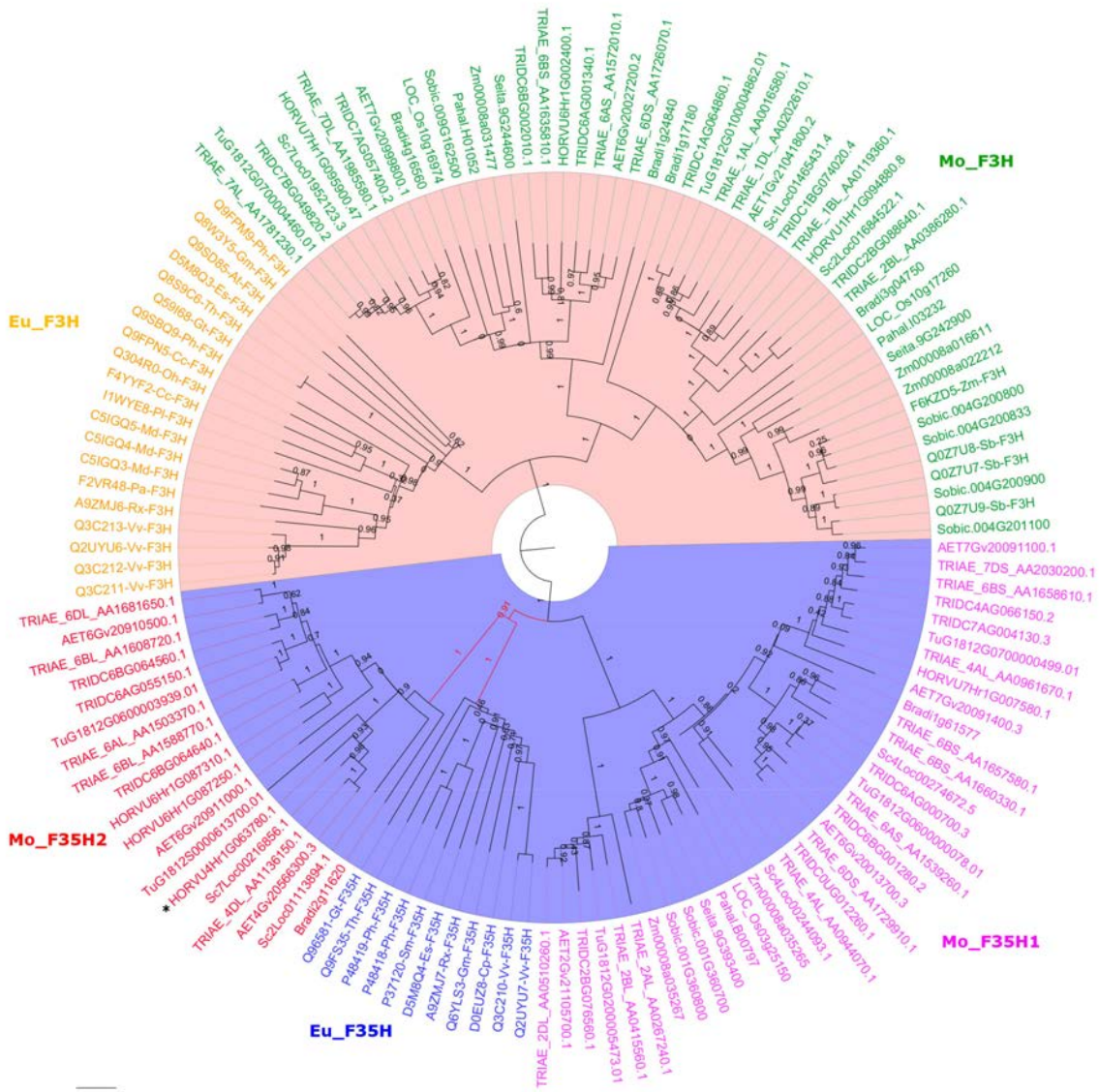


Figure S8. Maximum Likelihood (ML) phylogeny of plant F3'5'Hs based on the amino acid sequence alignment. Phylogeny was developed using JTT+G substitution model. Bootstrapping support (500 times) was indicated above each branch. Remote *Physcomitrella patens* (paten homologs) are used as an outgroup. Sequence IDs for previously characterised plant F3'5'Hs and F3'Hs are named in the Uniprot format. HvF35H (HORVU4Hr1G063780.1) is marked with *.

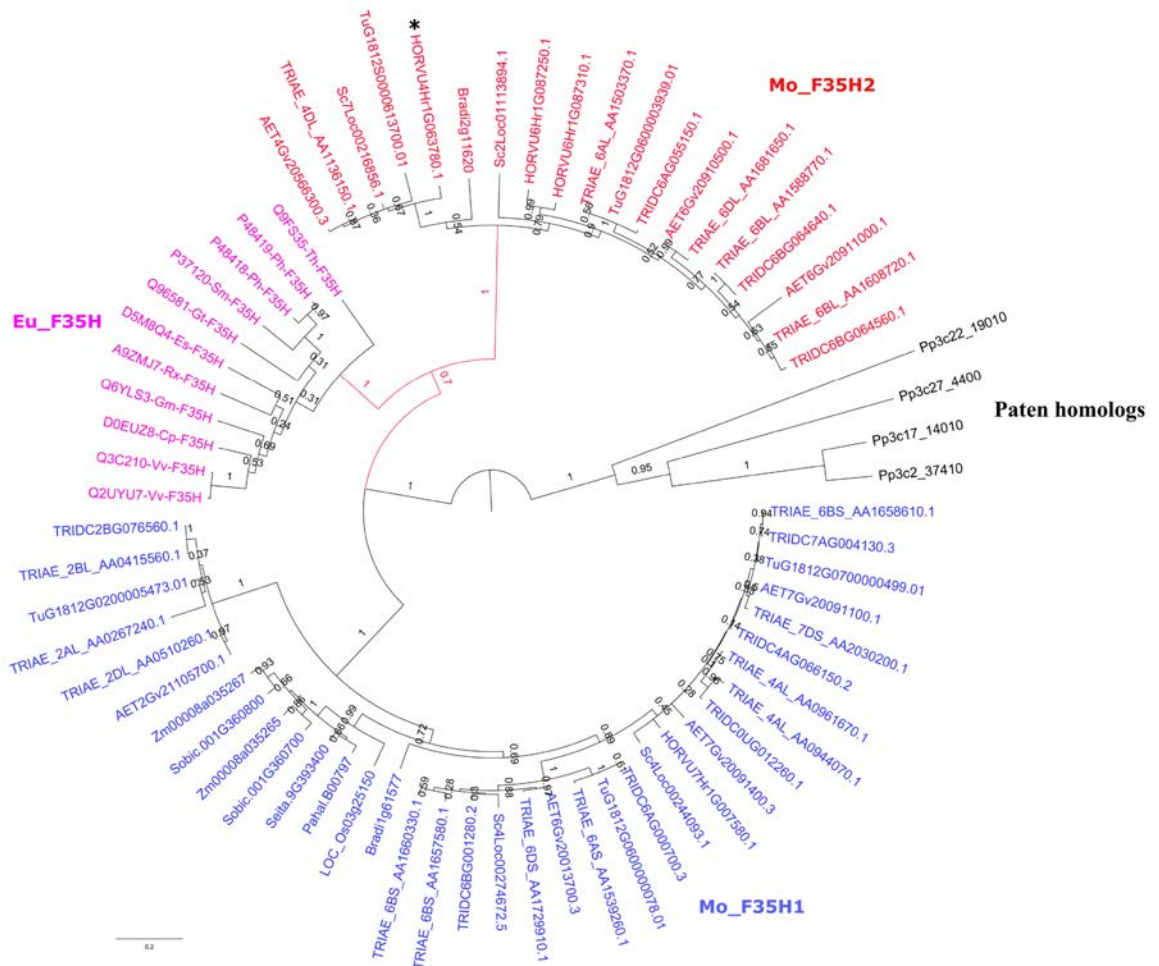


Figure S9. Comprehensive Neighbour Joining (NJ) tree of plant MYC family. Phylogeny was developed using the p-distance substitution model with 1000 times interior branching support test. Homologous MYC amino acid sequences from 12 Monocot species (*H. vulgare*, *T. aestivum*, *T. turgidum*, *T. urartu*, *A. tauschii*, *S. cereale*, *B. distachyon*, *O. sativa*, *P. hallii*, *S. italica*, *S. bicolor*, *Z. mays*) and 7 Eudicot species (*S. tuberosum*, *S. lycopersicum*, *V. Vinifera*, *P. trichocarpa*, *C. sinensis*, *T. cacao*, *G. max*) were retrieved and analysed. The branch coloured in red was selected as genuine MYC orthologues for downstream analyses. HvMYC4H and ThMYC4E are marked with *.

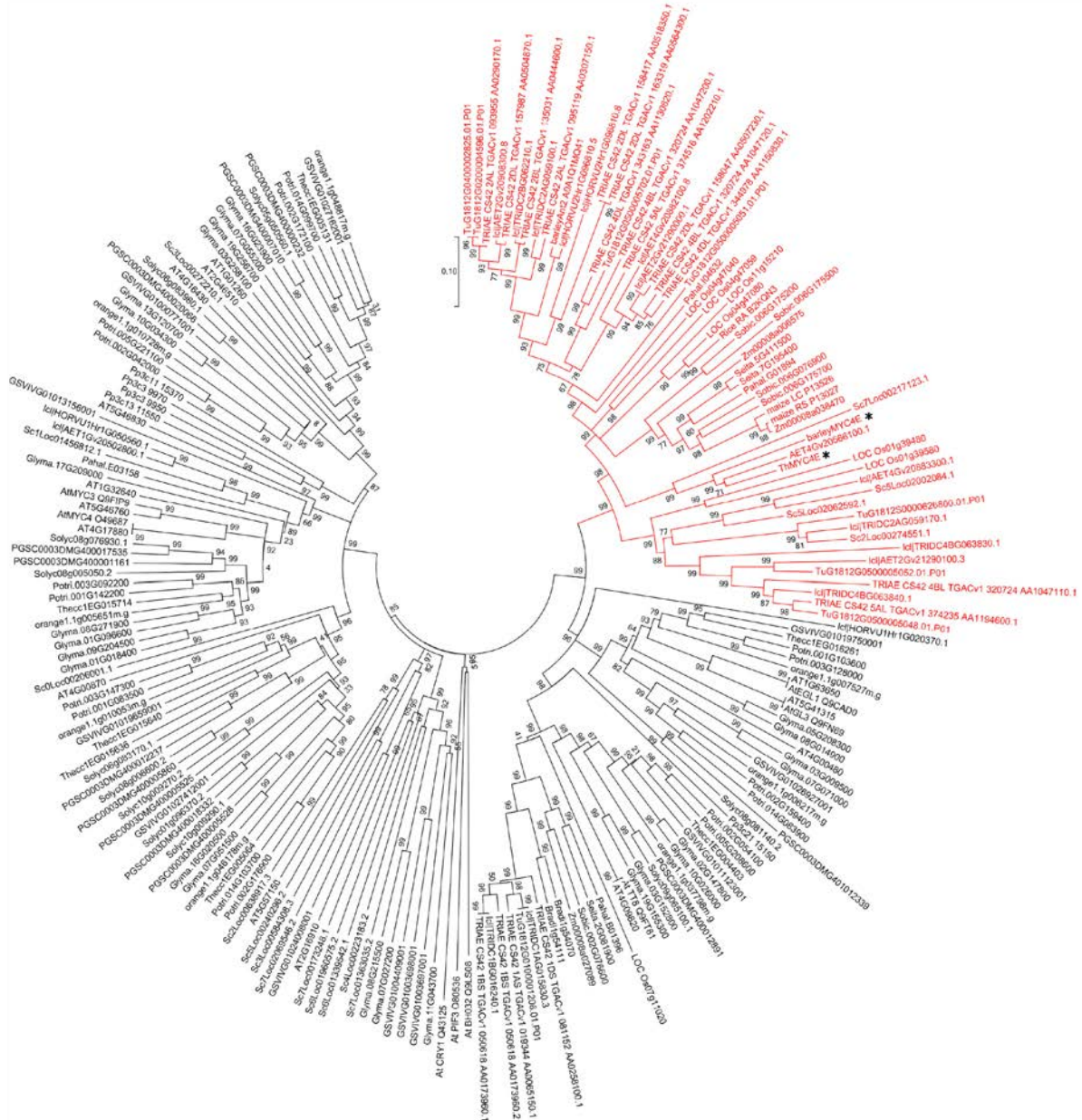


Figure S10. Comprehensive Neighbour Joining (NJ) tree of monocot MYB family. Phylogeny was developed using the p-distance substitution model with 1000 times interior branching support test. Homologous MYB amino acid sequences from 12 Monocot species (*H. vulgare*, *T. aestivum*, *T. turgidum*, *T. urartu*, *A. tauschii*, *S. cereale*, *B. distachyon*, *O. sativa*, *P. hallii*, *S. italica*, *S. bicolor*, *Z. mays*) and 7 Eudicot species (*S. tuberosum*, *S. lycopersicum*, *V. Vinifera*, *P. trichocarpa*, *C. sinensis*, *T. cacao*, *G. max*) were retrieved and analysed. The branch coloured in red was selected as genuine MYB orthologues for downstream analyses. HvMYB4H (HORVU4Hr1G063760) is marked with *. Only the subgroup containing the target HvMYB4H is displayed here. See **Data S12** for the complete NJ tree.

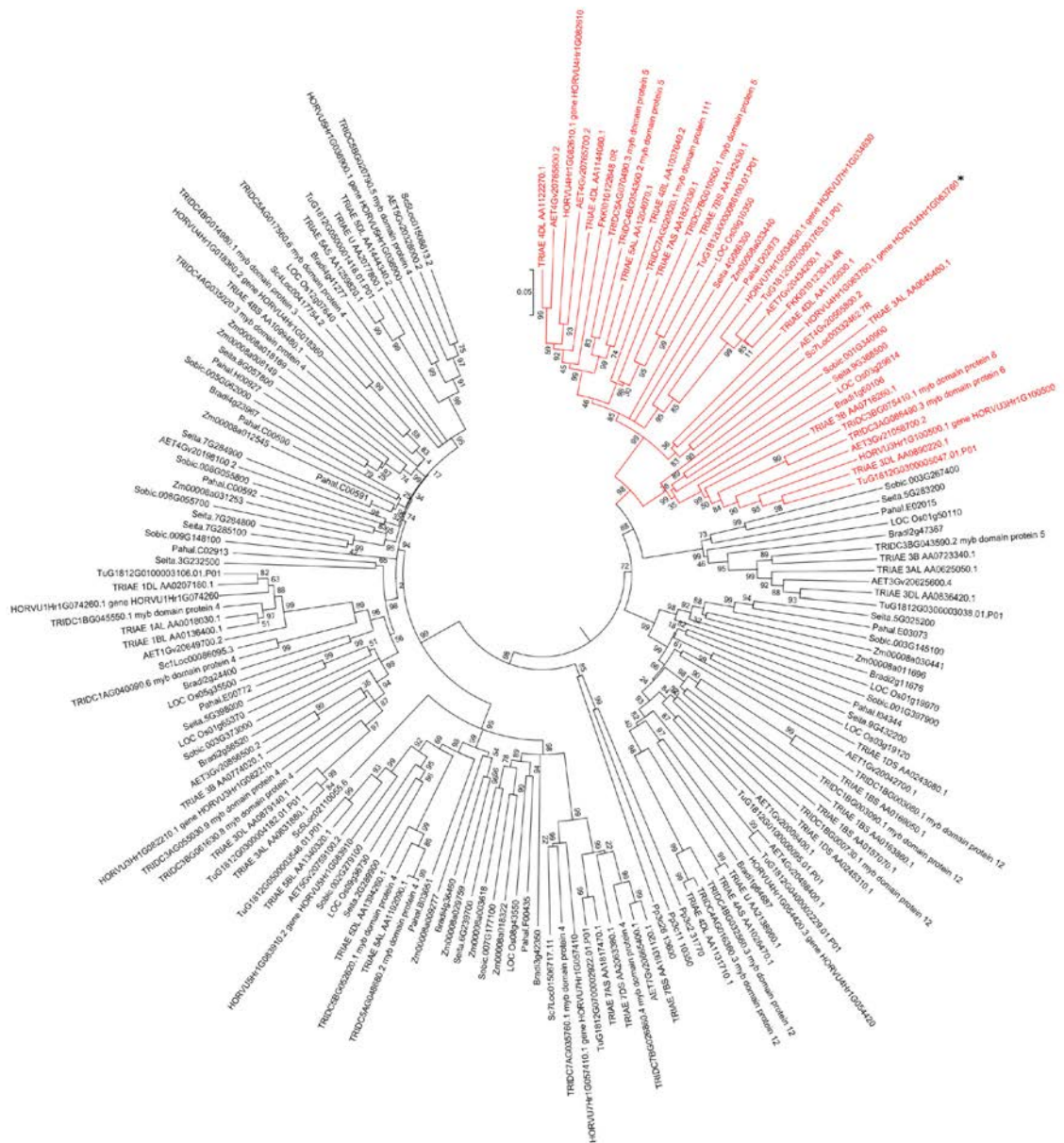


Figure S11. Results from Tajima’s test. **(a)** Comparison of Mo_F35H2 with Mo_F35H1 using Eu_F35H as an outgroup; **(b)** Comparison of Mo_F35H1 with Eu_F35H using Mo_F35H2 as an outgroup. The reconstructed ancestral sequences for each F3’5’H group were used (**Data S8**).

a

Configuration	Count
Identical sites in all three sequences	250
Divergent sites in all three sequences	51
Unique differences in Sequence A	79
Unique differences in Sequence B	55
Unique differences in Sequence C	43

NOTE.-- The equality of evolutionary rate between sequences **A** (*Mo F35H2*) and **B** (*Mo F35H1*), with sequence **C** (*Eu F35H*) used as an outgroup in Tajima's relative rate test [1]. The χ^2 test statistic was 4.30 ($P = 0.03815$ with 1 degree[s] of freedom). P -value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. The analysis involved 3 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 478 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [2].

b

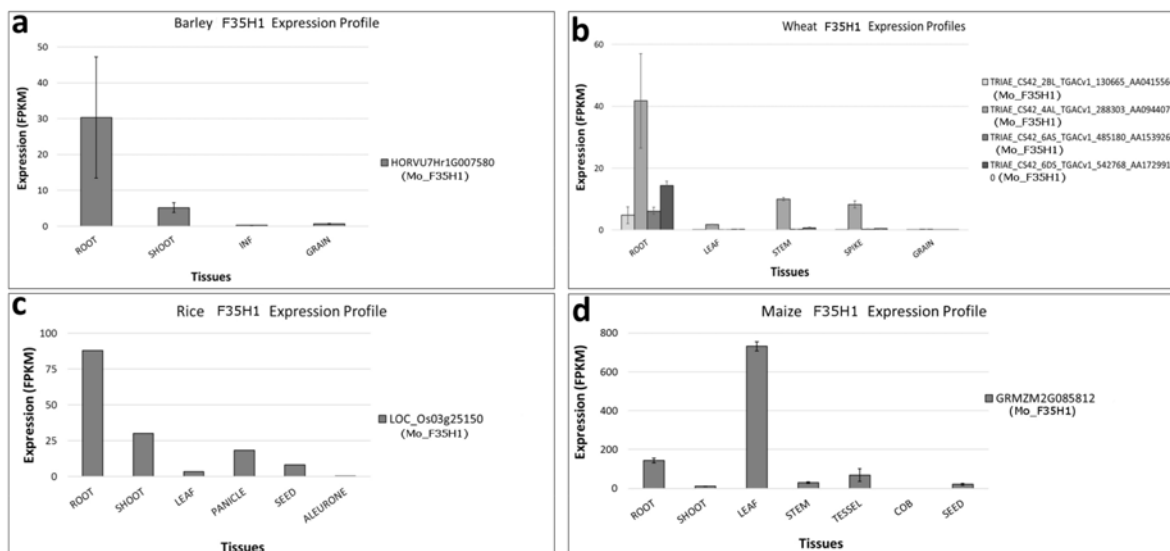
Configuration	Count
Identical sites in all three sequences	250
Divergent sites in all three sequences	51
Unique differences in Sequence A	55
Unique differences in Sequence B	43
Unique differences in Sequence C	79

NOTE.-- The equality of evolutionary rate between sequences **A** (*Mo F35H1*) and **B** (*Eu F35H*), with sequence **C** (*Mo F35H2*) used as an outgroup in Tajima's relative rate test [1]. The χ^2 test statistic was 1.47 ($P = 0.22544$ with 1 degree[s] of freedom). P -value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. The analysis involved 3 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 478 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [2].

1. Tajima F. (1993). Simple methods for testing molecular clock hypothesis. *Genetics* **135**:599-607.

2. Kumar S., Stecher G., and Tamura K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* (submitted).

Figure S12. Transcriptional profiles of Mo_F35H1 genes. Normalized transcriptional data for the Mo_F35H1 genes in barley (a), wheat (b), rice (c) and maize (d) were extracted from public databases. Descriptions of the tissues are listed below the graph.



Barley

Tissue	Description
ROOT	Roots from the seedlings (10 cm shoot stage)
SHOOT	Shoots from the seedlings (10 cm shoot stage)
INF	Developing inflorescences (1-1.5 cm)
GRAIN	Developing grain, bracts removed (15 DPA)

Wheat

Tissue	Description
ROOT	Root tissue, at 3 leaves unfolded
LEAF	Leaf tissue, at main shoot and 3 tillers
STEM	Stem tissue at 2nd detectable node
SPIKE	Spike tissue at anthesis (1/2 of flowering complete)
GRAIN	Grain tissue at 14 DAA (350°C.days). Kernel medium milk

Rice

Tissue	Description
ROOT	Root tissue at 7 days seedling
SHOOT	Shoot at 7 days seedling
LEAF	Leaf tissue at 7 days before heading to 7 days after flowering
PANICLE	Panicle at 7 days after flowering
SEED	Mature seed
ALEURONE	Mature aleurone

Maize

Tissue	Description
ROOT	Root system tissue at 7 DAS
SHOOT	Coleoptile tissue at 6 DAS
LEAF	Thirteenth leaf at vegetative stage with 9 emerged leaves
STEM	Stem and SAM at vegetative stage with 3 emerged leaf
TESSEL	Meiotic tassel at vegetative stage with 18 emerged leaf
COB	Pre-pollination cob
SEED	Whole seed at 12 DAP

Table S1. *In silico* promoter binding sites prediction for *HvF35H*. Predicted Transcription Factors (TF) located on chromosome 4H and 7H were highlighted in bold. The transcription start site for *HvF35H* is set as “0” for binding site numbering.

Predicted TFs	Gene Family	Binding site	Strand	<i>p</i> -value	<i>q</i> -value	Matched sequence
HORVU4Hr1G023510	MYB	-911 ... -893	-	3.14E-06	0.00632	GTTTTAGTAGTTGGAACAT
HORVU7Hr1G027370	MYB	-896 ... -882	+	3.21E-06	0.0069	AAACGTAACAGAAAC
HORVU5Hr1G021750	MYB_related	+5 ... +14	+	8.13E-06	0.0163	AGTTAGGTAG
HORVU6Hr1G078300	MYB	-910 ... -890	-	1.51E-05	0.0311	TACGTTTTAGTAGTTGGAACA
HORVU6Hr1G060650	MYB	-907 ... -893	-	3.33E-05	0.0668	GTTTTAGTAGTTGGA
HORVU1Hr1G073300	MYB_related	-614 ... -594	+	3.81E-05	0.0693	AAATTCGAAGATAAAATTGAA
HORVU1Hr1G017970	MYB_related	-907 ... -893	+	4.10E-05	0.0864	TCCAACACTAAAAC
HORVU4Hr1G054420	MYB	-349 ... -340	+	5.02E-05	0.103	GCACCTACCT
HORVU6Hr1G058620	MYB	-360 ... -340	+	5.71E-05	0.114	TACGAGCACATGCACCTACCT
HORVU3Hr1G004050	MYB	0 ... +14	+	7.18E-05	0.148	TTTGCAGTTAGGTAG
HORVU2Hr1G051030	MYB	-5 ... +15	+	7.37E-05	0.15	TCGATTTTGCAGTTAGGTAGT
HORVU1Hr1G050560	bHLH	-355 ... -348	+	8.82E-05	0.0948	GCACATGC
HORVU1Hr1G050560	bHLH	-355 ... -348	-	8.82E-05	0.0948	GCATGTGC
HORVU5Hr1G049880	MYB	-1 ... +9	-	9.06E-05	0.181	TAACTGCAAAA
HORVU4Hr1G018360	MYB	+1 ... +15	+	9.59E-05	0.101	TTGCAGTTAGGTAGT
HORVU4Hr1G018360	MYB	-1028 ... -1014	+	9.88E-05	0.101	CCAAAATTTGGTATG

Table S2. Likelihood Ratio Tests (LRTs) on the MbHF35 cluster for selection analyses. *df* refers the degree of freedom. $2\Delta l$ is calculated based on the $l = \ln L$ values for the various models tested in **Table 2, S2 and S3.**

F3'5'H LRTs	df	2Δl	P - value
$\omega[\text{eudi}] = \omega[\text{mono1}] \neq \omega[\text{mono2}]$ vs $\omega[\text{eudi}] = \omega[\text{mono1}] = \omega[\text{mono2}]$	1	33.82	< .0001
$\omega[\text{eudi}] = \omega[\text{mono2}] \neq \omega[\text{mono1}]$ vs $\omega[\text{eudi}] = \omega[\text{mono1}] = \omega[\text{mono2}]$	1	55.00	< .0001
$\omega[\text{eudi}] \neq \omega[\text{mono1}] \neq \omega[\text{mono2}]$ vs $\omega[\text{eudi}] = \omega[\text{mono1}] \neq \omega[\text{mono2}]$	1	94.26	< .0001
$\omega[\text{eudi}] \neq \omega[\text{mono1}] \neq \omega[\text{mono2}]$ vs $\omega[\text{eudi}] = \omega[\text{mono2}] \neq \omega[\text{mono1}]$	1	73.08	< .0001
Selection M2 vs Neutral M1	1	0.00	1
Model A (Mo_F35H1) vs Model A Null (Mo_F35H1)	1	6.38	0.0115
Model A (Mo_F35H2) vs Model A Null (Mo_F35H2)	1	20.58	< .0001
Model A (Mo_F35H2) vs Neutral M1	2	36.74	< .0001
MYC LRTs			
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$ vs $\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	1	4.34	0.0372
$\omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$ vs $\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	1	0.02	0.8875
$\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$ vs $\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	1	3.66	0.0557
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$ vs $\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	1	0.04	0.8415
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] \neq \omega[\text{Other_MYC}]$ vs $\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	2	3.82	0.1481
Selection M2 vs Neutral M1	1	0.00	1
Model A (Mo_MYCa) vs Model A Null (Mo_MYCa)	1	0.02	0.8875
Model A (Mo_MYCb) vs Model A Null (Mo_MYCb)	1	2.26	0.1328
MYB LRTs			
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$ vs $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	1	1.20	0.2733
$\omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$ vs $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	1	2.08	0.1492
$\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$ vs $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	1	0.22	0.6390
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$ vs $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	2	3.60	0.1653
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] \neq \omega[\text{Other_MYB}]$ vs $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	3	3.92	0.2702
Selection M2 vs Neutral M1	1	0.00	1
Model A (Mo_MYBa) vs Model A Null (Mo_MYBa)	1	0.00	1
Model A (Mo_MYBb) vs Model A Null (Mo_MYBb)	1	4.40	0.0359

Table S3. Natural selection tests on monocot MYC family.

Model	Number of parameters	ln(Likelihood)	Estimates of parameters ^a ($\omega = d_N : d_S$, P - percentage of site)	Positively selected sites
M0: one-ratio $\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	1	-15492.22	$\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}] = 0.32676$	Not Allowed (NA)
Branch-specific models				
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	2	-15490.05	$\omega[\text{Tr_MYCa}] = 0.48319$ $\omega[\text{Tr_MYCb}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}] = 0.31972$	NA
$\omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	2	-15492.21	$\omega[\text{Tr_MYCb}] = 0.32072$ $\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}] = 0.32699$	NA
$\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	2	-15490.39	$\omega[\text{Tr_MYCa}] = \omega[\text{Tr_MYCb}] = 0.19788$ $\omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}] = 0.33298$	NA
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}]$	3	-15490.03	$\omega[\text{Tr_MYCa}] = 0.48547$, $\omega[\text{Tr_MYCb}] = 0.30425$, $\omega[\text{Tr_MYCc}] = \omega[\text{Other_MYC}] = 0.32028$	NA
$\omega[\text{Tr_MYCa}] \neq \omega[\text{Tr_MYCb}] \neq \omega[\text{Tr_MYCc}] \neq \omega[\text{Other_MYC}]$	4	-15488.59	$\omega[\text{Tr_MYCa}] = 0.48697$, $\omega[\text{Tr_MYCb}] = 0.30445$, $\omega[\text{Tr_MYCc}] = 0.20843$, $\omega[\text{Other_MYC}] = 0.32611$	NA
Site-specific models				
Neutral M1 (2 site classes)	2	-15263.74	$P_0 = 0.72772$ ($P_1 = 1 - P_0 = 0.27228$); $\omega_0 = 0.19410$, ($\omega_1 = 1.0$)	NA
Selection M2 (3 site classes)	3	-15263.74	$P_0 = 0.72773$, $P_1 = 0.14058$ ($P_2 = 1 - P_0 - P_1 = 0.13169$); $\omega_0 = 0.19410$, ($\omega_1 = 1.0$), $\omega_2 = 1.0$	None
Branch-site models				
Model A Null (Mo_MYCa)	3	-15261.61	$P_0 = 0.61974$, $P_1 = 0.23204$, $P_2 + P_3 = 0.14882$; $\omega_0 = 0.18932$, $\omega_1 = \omega_2 = 1.0$	NA
Model A (Mo_MYCa)	4	-15261.60	$P_0 = 0.63319$, $P_1 = 0.23679$, $P_2 + P_3 = 0.13002$; $\omega_0 = 0.18948$, $\omega_1 = 1.0$, $\omega_2 = 1.1538$	None
Model A Null (Mo_MYCb)	3	-15262.38	$P_0 = 0.65131$, $P_1 = 0.24055$, $P_2 + P_3 = 0.10809$; $\omega_0 = 0.19182$, $\omega_1 = \omega_2 = 1.0$	NA
Model A (Mo_MYCb)	4	-15261.25	$P_0 = 0.71832$, $P_1 = 0.26744$, $P_2 + P_3 = 0.01425$; $\omega_0 = 0.19249$, $\omega_1 = 1.0$, $\omega_2 = 10.81706$	None

^a In the site-specific model M1, two site classes were specified: highly conserved sites (ω_0) and neutral sites ($\omega_1=1$). For the site-specific model M2, there were three site classes: highly conserved sites (ω_0), neutral sites ($\omega_1=1$) and positively selected sites (ω_2). In Model A, four site classes were specified. The first two classes had ω ratios of ω_0 and ω_1 , respectively, corresponding to highly conserved sites and neutral sites across all lineages. In the other two site classes, the background lineages had ω_0 or ω_1 while the foreground lineages had ω_2 .

Table S4. Natural selection tests on monocot MYB family.

Model	Number of parameters	ln(Likelihood)	Estimates of parameters ^a ($\omega = d_n : d_s$, P - percentage of site)	Positively selected sites ^b
M0: one-ratio				
$\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	1	-3579.85	$\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}] = 0.20890$	Not Allowed (NA)
Branch-specific models				
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	2	-3579.25	$\omega[\text{Tr_MYBa}] = 0.14185$ $\omega[\text{Tr_MYBb}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}] = 0.21446$	NA
$\omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	2	-3578.81	$\omega[\text{Tr_MYBb}] = 0.51006$ $\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}] = 0.20230$	NA
$\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	2	-3579.74	$\omega[\text{Tr_MYBa}] = \omega[\text{Tr_MYBb}] = 0.17334$ $\omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}] = 0.21101$	NA
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}]$	3	-3578.05	$\omega[\text{Tr_MYBa}] = 0.13099$, $\omega[\text{Tr_MYBb}] = 0.60138$, $\omega[\text{Tr_MYBc}] = \omega[\text{Other_MYB}] = 0.20766$	NA
$\omega[\text{Tr_MYBa}] \neq \omega[\text{Tr_MYBb}] \neq \omega[\text{Tr_MYBc}] \neq \omega[\text{Other_MYB}]$	4	-3577.89	$\omega[\text{Tr_MYBa}] = 0.13031$, $\omega[\text{Tr_MYBb}] = 0.62149$, $\omega[\text{Tr_MYBc}] = 0.16529$, $\omega[\text{Other_MYB}] = 0.21050$	NA
Site-specific models				
Neutral M1 (2 site classes)	2	-3433.06	$P_0 = 0.70626$ ($P_1 = 1 - P_0 = 0.29374$); $\omega_0 = 0.06073$, ($\omega_1 = 1.0$)	NA
Selection M2 (3 site classes)	3	-3433.06	$P_0 = 0.70626$, $P_1 = 0.13678$ ($P_2 = 1 - P_0 - P_1 = 0.15696$); $\omega_0 = 0.06073$, ($\omega_1 = 1.0$), $\omega_2 = 1.0$	None
Branch-site models				
Model A Null (Mo_MYBa)	3	-3432.73	$P_0 = 0.67520$, $P_1 = 0.27543$, $P_2 + P_3 = 0.04937$; $\omega_0 = 0.05976$, $\omega_1 = \omega_2 = 1.0$	NA
Model A (Mo_MYBa)	4	-3432.73	$P_0 = 0.67520$, $P_1 = 0.27543$, $P_2 + P_3 = 0.04937$; $\omega_0 = 0.05976$, $\omega_1 = 1.0$, $\omega_2 = 1.1538$	None
Model A Null (Mo_MYBb)	3	3432.73	$P_0 = 0.67636$, $P_1 = 0.27584$, $P_2 + P_3 = 0.04780$; $\omega_0 = 0.06028$, $\omega_1 = \omega_2 = 1.0$	NA
Model A (Mo_MYBb)	4	-3430.53	$P_0 = 0.70125$, $P_1 = 0.28102$, $P_2 + P_3 = 0.01773$; $\omega_0 = 0.06153$, $\omega_1 = 1.0$, $\omega_2 = 351.16854$	175A

^a In the site-specific model M1, two site classes were specified: highly conserved sites (ω_0) and neutral sites ($\omega_1 = 1$). For the site-specific model M2, there were three site classes: highly conserved sites (ω_0), neutral sites ($\omega_1 = 1$) and positively selected sites (ω_2). In Model A, four site classes were specified. The first two classes had ω ratios of ω_0 and ω_1 , respectively, corresponding to highly conserved sites and neutral sites across all lineages. In the other two site classes, the background lineages had ω_0 or ω_1 while the foreground lineages had ω_2 .

^b Positively selected amino acids at P -value ≤ 0.05 are numbered according to the barley MYBb gene (*HORVU4Hr1G082610.1*).

Supplementary Data S1 - S8. Data files are available in the Figshare data repository at <https://figshare.com/s/32dad83a1028b0405e38>.