Palacký University Olomouc
Faculty of Science
Department of Botany

And
Institute of Experimental Botany
Centre of Region Haná for Biotechnological and Agricultural Research
Olomouc

Zuzana Ivaničová

Mapping and identification of flowering time genes in bread wheat (*Triticum aestivum* L.)

Ph. D. Thesis

Supervisor: RNDr. Jan Šafář, Ph.D.

Olomouc 2017

Acknowledgements:
I would like to express great thanks to my supervisor, RNDr. Jan Šafář, Ph.D., for a professional guidance, inspirational ideas, patience and help. Further, I would like to thank to the head of the laboratory, prof. Ing. Jaroslav Doležel, DrSc., for the opportunity to work in the Centre of the Plant Structural and Functional Genomics and also to all the members of the team for a friendly and motivating atmosphere. My final thank goes to my family for their patience and support during the years of my studies.

Declaration:
I hereby declare that I elaborated this Ph.D. thesis independently under the supervision of RNDr. Jan Šafář, Ph.D. and using only information sources referred in the Reference chapter.

…….

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Mgr. Zuzana Ivaničová
Title: Mapping and identification of flowering time genes in bread wheat (*Triticum aestivum* L.)
Type of Thesis: Ph.D. thesis
Department: Botany
Supervisor: RNDr. Jan Šafář, Ph.D.
The Year of Presentation: 2017
Abstract: Bread wheat (*Triticum aestivum*) is one of the most important crops. It is a staple food for 30% of population. Increasing of human population leads to necessity to achieve higher yield without enlarging of arable land. One way of achievement may rest in adaptation of flowering process to microenvironmental conditions. The flowering and subsequent grain filling period is sensitive and optimal conditions during this process have profound impact on grain quality and yield.
This work focuses on identification of genes and their allelic variation which participates on adjustment of flowering.
This work can be divided into two parts. The first part of the work deals with identification and characterization of new allele of vernalization gene *VRN-A1* – *VRN-A1f-like* allele. The very original source of this allele is tetraploid wheat *Triticum millitinae* Zhuk. with AAGG genomic composition. The allele confers spring habit and is characteristic by short deletions in promoter region, insertion of miniature inverted transposable element and wide deletion in the first intron. This allele was responsible for delayed flowering in comparison to previously identified *Vrn-A1a* allele.
The second project determined that higher number of copies of photoperiodic gene *Ppd-B1* was responsible for difference in flowering time between two bread wheat varieties. Identification of flowering time variability in F_7 generation lines with three copies of *Ppd-B1* led to determination of expression and methylation pattern of chosen lines, in order to reveal nature of this intragroup variation.
The important source of new genes and alleles are domesticated tetraploid wheat. The identification of evolutionary relationship within these groups and their genetic characterization can lead to discovery of new sources of variability. The additional part of the work is dealing with evolutionary relationship within tetraploid wheats and their domestication.
Knowledge of flowering time genes, their functions and interaction in determination of this process is important prerequisite for breeding new varieties adapted to local microenvironment.
Key words: bread wheat, flowering time, genetic mapping, QTL analysis, vernalization, photoperiod

Number of pages/Appendices: 102/IX
Language: English

BIBLIOGRAFICKÁ IDENTIFIKÁCIA

Meno: Mgr. Zuzana Ivaničová
Názov: Identifikácia a mapovanie génov kvitnutia u pšenice letnej (*Triticum aestivum* L.)
Typ práce: Dizertačná práca
Odbor: Botanika
Školiteľ: RNDr. Jan Šafář, Ph.D.
Rok obhajoby: 2017
Abstrakt: Pšenica letná (*Triticum aestivum*) je základnou plodinou pre 30 % celosvetovej populácie. V súčasnej dobe je nutné jej ročný výnos zvyšovať, a to v dôsledku neustáleho rastu počtu ľudí. Vyšší výnos je však ohrozený rôznymi abiotickými a biotickými faktormi. Jedným z možných spôsobov je modifikácia času kvitnutia tak, aby bol prispôsobený lokálnym podmienkam prostredia. Optimálny čas kvitnutia je totiž dôležitý pre ochranu kvetných častí pred poškodením (v dôsledku chladu alebo sucha), a teda zabezpečenie vhodných podmienok pre vývin obiliek.
Hlavným cieľom predloženej práce je identifikácia génov/alel, ktoré sú zodpovedné za variáciu v čase kvitnutia. Vlastnú prácu je možné tematicky rozdeliť na dve časti.
Prvá časť sa zaoberá identifikáciou novej alely vernalizačného génu *Vrn-A1f-like*. Táto alela je charakteristická sekvenčnými zmenami v promótorovej oblasti (krátke delécie) a v oblasti prvého intrónu (inzercia *miniature inverted transposable element* a rozsiahla delécia). Alela *Vrn-A1f-like* bola zodpovedná za oneskorenie kvitnutia v porovnaní s už predtým identifikovanou alelou *Vrn-A1a*.
Zodpovednosť vyššieho počtu kópií fotoperiodického génu *Ppd-B1* za rozdiel v čase kvitnutia u variet pšenice letnej bola nosnou témou druhej časti práce. V štúdii bola zároveň identifikovaná variabilita v čase kvitnutia v rámci línií s troma kópiami tohto génu. V snahe zistiť príčinu týchto rozdielností došlo k analýze expresného profilu a rovnako aj metylácie sledovaného génu.
Práve identifikácia nových génov/alel zodpovedných za rozdiely v kvitnutí, ich charakterizácia a identifikácia funkcie v rámci biochemických dráh umožní úplné pochopenie tohto dôležitého procesu. Posledná časť práce sa zaoberala identifikáciou miesta domestikácie a evolučných vzťahov medzi skupinou tetraploidných pšeníc. Práve charakterizácia týchto druhov na úrovni genómu určenie vzťahov medzi nimi môže viesť k identifikácii nového, cenného zdroja genetickej variability. Implementácia týchto znalostí do procesu šľachtenia zas umožní tvorbu nových kultivarov schopných vo výraznej miere odolávať nepriaznivým podmienkam prostredia.

Kľúčové slová: pšenica letná, čas kvitnutia, genetické mapovanie, analýza QTL, vernalizácia, fotoperióda

Počet stránok/Príloh: 102/IX
Jazyk: anglický

CONTENT

1 INTRODUCTION

Initiation of flowering is one of the most important processes in plant life. Conditions during this process and following period of grain development predetermine number and grain viability or in agricultural language - yield and grain quality. It is the variation in key genes determining flowering process (vernalization requirement and photoperiod response) which enabled cultivation of wheat from equator to polar circles. However, current wheat yield is threatened by an unprecedented pressure of biotic and mainly abiotic stresses, for which currently used elite bread wheat cultivars are not suited well. This can be solved by improvement of old varieties and development of new one which will be adapted to local environmental conditions. They will be able to flower in the very best time to take advantage of local conditions on one side and to avoid unsuitable ones on the other side. The determination of genes influencing flowering and their mutual interactions should preceded breeding new suitable cultivars. For the initiation of flowering under optimal conditions, plant needs to take into consideration many external and internal cues. This is the reason why flowering is determined by many genes which very often work in comprehensive and redundant manner. Identification of these genes, their allelic variants and interaction within different biochemical pathways is therefore very complicated process. Not surprisingly, the pioneer studies which shed some lights to this process have been done using model object *Arabidopsis thaliana* L. To date, the analysis of flowering pathway in this model object made the huge progress and it is the best studied object in this area. The accumulated results confirmed complexity and complicated nature of genetic determination of flowering. The gathered information showed to be useful in the study of agriculturally important plants like rice, barley and bread wheat. However, studies identified convergent nature of some parts of the flowering pathway between monocots and dicots. Thus, the information from *Arabidopsis* cannot be used generally and analysis of wheat and rice needs to be done individually.

Current progress in molecular techniques leads to study of even such complicated plant as bread wheat. Identification of new genes and alleles is hampered not only by genome size (1C ~ 17 Gbp), presence of three homoeologous genomes and high content of repetitive elements (> 90%), but also by dominant effect of main flowering genes affecting vernalization process and perception of photoperiod. These

two groups of genes decrease manifestation of genes with just a minor effect on flowering time.

Complexity of flowering process is driving force of current research which slowly but surely heads towards final understanding of this essential process.

2 LITERATURE REVIEW

2.1 Flowering time pathway

Plant development and propagation is threatened by unsuitable and changing external conditions. Plants cannot move to overcome these threats. Thus, they need to evolve adaptations which allow instant response to external change and utilization of the most suitable conditions. Plant needs to receive and process various information from environment. It uses predominantly the length of the day and temperature to determine time and place for planning its life cycle. Various conditions throughout different altitudes and latitudes lead to development of plants which need long (LD) or short days (SD) for flowering. In addition, the photoperiod insensitive plants which are able to flower irrespective of day length have emerged. Different temperature conditions lead to development of plants with winter or spring growth habit; the first ones need cold treatment to initiate flowering.

The optimal time of flowering is ensured by network of genes which integrates both internal and external signals. When the convenient conditions are reached, these genes are responsible for initiation of flower development and flowering itself.

The research focused on understanding of flowering pathways uses mainly two species: *Arabidopsis* and rice. The model organism for dicots, *Arabidopsis thaliana*, belongs to the best studied objects. This is due to its simple manipulation, small and fully sequenced genome. Rice is a monocot plant which serves as a model organism for plants capable of flowering under SD conditions. Moreover, there are other monocot plants (barley and wheat) which belong to agronomical important crops. The flowering process is intensively studied also in those. Understanding of flowering time pathway of crops is prerequisite for successful breeding of new varieties with higher yield in the future.

2.1.1 *Arabidopsis thaliana*

A. thaliana is one of the most important model objects for plants. It is winter annual with short life cycle and small (1C ~ 153 Mbp) genome. It provides the most comprehensive knowledge of flowering time determination to date [1]. High level of conservation enables usage of gained information as the base for the research of agriculturally important plant species. However, the conservation and synteny is not limitless and thus findings cannot be used comprehensively.

LD and period of cold treatment are necessary for initiation of flowering in *A. thaliana*. The conditions for flowering are determined by network of genes which are usually divided into five categories: gibberellin, autonomous, endogenous, vernalization and photoperiod pathway. Gibberellin signaling pathway consists of genes determining requirement of gibberellic acid (GA) for normal flowering pattern. Endogenous pathway explains influence of plant age to the control of flowering time. Autonomous pathway is characterized by regulators which are independent of the photoperiod and gibberellin pathways. Vernalization pathway is responsible for perception of cold and the photoperiod pathway is responsible for perception of day length [2].

2.1.1.1 Gibberellin signaling pathway

The effect of GA on plant grow was discovered in 1926 by Japanese scientist Kurosawa. He observed tall rice plants with male sterility infected by *Gibberella fujikuroji* (Sawada) Wollenw. [2]. This, so called ´balkane disease´ was caused by GA produced by this pathogen fungus. The GA is naturally produced in plants leaves. It is considered a plant hormone and plays an important role in flowering induction during SD conditions [3,4]. GA regulates *FLOWERING LOCUS T* (*FT*) mRNA level during SD. This role is taken over by flowering activator protein CONSTANS (CO) (reviewed in: [5]).

Active forms of GAs (GA3 and GA4) are perceived by GIBBERELLIN INSENSITIVE DWARF 1 (GID) receptor. It is a soluble GA receptor localized in the nucleus [5,6]. Binding of GA on GID receptor leads to conformational change and recognition of DELLA protein RGA (repressor of ga1-3) which is followed by their ubiquitination and degradation. These proteins belong to GRAS family transcription factors (named after GIBBERELLIC-ACID INSENSITIVE -GAI, RGA and SCARECROW) and has N-terminal DELLA domain (named after five conserved aminoacids within) (Sasaki Itoh, H. et al. 2003; McGinnis 2003; Lucas et al. 2008; Feng et al. 2008). RGA proteins repress plant growth by immobilizing PHYTOCHROME INTERACTING FACTORS (PIF). Immobilizing influences DNA binding activity of PIF proteins and leads to changes in response to red and far red light and to delay of flowering time [10,11]. PIF proteins are furthermore destabilized by phytochrome B (phy B). Therefore, gibberellins permit accumulation of functional PIF proteins in nucleus by inhibiting function of RGA protein.

Mutation in DELLA domain prevents GA mediated degradation and causes dwarf phenotype [10]. Further, GA is affecting expression of important integrator of vernalization and photoperiod pathway – *SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *LEAFY* (*LFY*) [3].

Recently, physical interaction of DELLA protein with CO was identified. DELLA protein RGA has been shown to inhibit interaction of CO and NF-Y (nuclear transcription factor Y) proteins [12]. This way Xu et al. [12] identified crosstalk between photoperiod and GA in negative regulation of CO and thus flowering (Fig. 1).

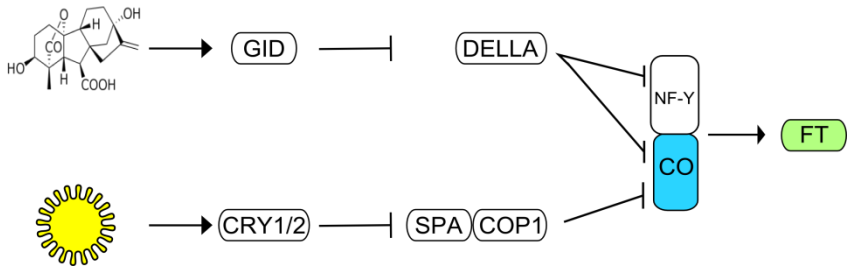


Fig. 1: The crosstalk between light and GA signalling in negative regulation linkage of flowering. CRYPTOCHROME 1 and 2 (CRY1 and CRY2) photoreceptors stabilize CO activity under LDs conditions, namely by inhibition of CONSTITUTIVELY PHOTOMORPHOGENIC 1 (COP1)/ SUPRESSOR PHYA-105 (SPA) complex activity through direct interaction CRYs with COP1 and SPA. Binding gibberellin to GID receptor enables stabilization of CO by repressing DELLA proteins activity. By this GID abolishes inhibition of CO and NF-Y by DELLA proteins. Arrows and T bars denote positive and negative regulation, respectively. Sun denote LDs. (adapted from: [12]).

2.1.1.2 Endogenous pathway

Endogenous pathway includes genes which are influencing plant senescence and carbohydrates homeostasis. Both these factors affect initiation of flowering [13,14].

Senescence has important role in flowering initiation. It is connected to decreased level of miR156. This micro RNA represses transcription of *SQAMOSA – PROMOTER BINDING PROTEIN-LIKE* (*SPL*). SPL plays a role in stimulating the production of LFY, MADS box protein (named after four founding members: MCM1, AGAMOUS, DEFICIENS and SRF), and miR172, which enhances flowering [15,16]. It also decreases *SOC1* and *FRUITFULL* (*FUL*) expression which results in flowering delay [17,18].

Carbohydrates influence flowering time in various ways and their precise role needs to be determined. Carbohydrates are interacting with hormonal (GA) signaling pathway. They have effect on level of trehalose – 6 –phosphate and level of sucrose seems to have effect on expression of *FLOWERING LOCUS T* (*FT*) and *LFY* [4,19–22].

It was also reported that flowering is affected by modification of carbohydrates balance in apical meristem [23]. The balance between hexose and sucrose is directed by invertase. Invertase influences carbon dioxide uptake, photosynthesis and plant growth with clear connection to flowering time pathway gene network (reviewed in [24]).

2.1.1.3 Autonomous pathway

Autonomous pathway consists of genes which do not form biochemical pathway in the narrow sense of this word. These genes have a few common characteristics: (i) they are working independently of photoperiod and gibberellin pathways, (ii) they inhibit expression of *FLOWERING LOCUS C* (*FLC*), (iii) they often belong to chromatin remodeling and maintenance or RNA processing factors, and last but not least (iv) the mutation of these genes delay flowering irrespective of day length [25]. Several autonomous pathway genes have been identified so far: *FLOWERING LOCUS CA* (*FCA*), *FLOWERING LOCUS PA* (*FPA*) , *FLOWERING LOCUS K* (*FLK*) which contain RNA binding domain; *FY* shows homology to RNA processing factors; *FLOWERING LOCUS D* (*FLD*) and *RELATIVE OF EARLY FLOWERING 6* (*REF6)* which have a role as a histone demethylases (reviewed in [26]). The primary effect of *FCA*, *FPA* and *FY* is associated with regulation antisense RNAs called COOLAIR (COLD INDUCED LONG ANTISENSE INTRAGENIC RNA) resulting in enhancement of flowering [27].

2.1.1.4 Vernalization pathway

Vernalization is an important adaptation of plants growing in regions where growing season is preceded by winter. The requirement for vernalization prevents flowering in the fall and ensures the occurrence of flowering in convenient period of the year, allowing sufficient time for seed development. The vernalization requirement can be facultative or obligatory.

Arabidopsis belongs to vernalization obligatory plants. The perception of cold period and following flowering is coded by vernalization pathway genes.

The necessity of cold requirement is determined mainly by expression of *FLC* gene. *FLC* is MADS box transcription factor which repress flowering before cold treatment. *FLC* levels are high prior vernalization and they are gradually decreasing during cold period. No clear orthologue of *FLC* was identified in other taxa besides *Brasiceae* [28]. This indicates convergent evolution of vernalization pathway. *FLC* expression is enhanced by binding of FRIGIDA (FRI) to *FLC* promoter region in winter annuals [29,30]. Different allelic variants of these two genes lead to emergence of rapid cycler plants (do not have vernalization requirement and thus can have more generations during the year) [29].

FLC repression is regulated in the time specific manner on various levels:
(i) in the early cold period, it is mediated by noncoding RNAs – COOLAIR and COLDAIR (COLD ASSISTED INTRONIC NONCODING RNA). COOLAIR is synthesized from promoter localized behind poly(A) signal of *FLC* locus (Fig. 2). It terminates either at a proximal poly(A) cluster in intron 6 (class I) or at the *FLC* promoter region (clas II) [27,31–33]. Significant effect of COOLAIR was confirmed by replacing its promoter with alternative promoter (terminator exchange FLC^{TEX}) which lead to reduction of *FLC* downregulation [34]. COOLAIR is regulated also by presence of R loop (RNA-DNA hybrid which covers COOLAIR promoter). Stabilization of this structure leads to inhibition of COOLAIR transcription [35]. On the contrary, COLDAIR is capped but not polyadenylated and it is synthesized in the sense manner. It arises from *FLC* intron I and it reaches the highest levels approximately 10 days after COOLAIR [33,34,36]. COLDAIR transiently associates with component of POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) (CURLY LEAF) and promotes its recruitment to *FLC* (Fig. 2) [36].

(ii) In the later cold period, repression is enhanced by epigenetic modification of histones. Trimethylation of histone 3 lysine 27 residue (H3K27me3) further decreases *FLC* expression. It is mediated by VERNALIZATION INSENSITIVE 3 (VIN3).

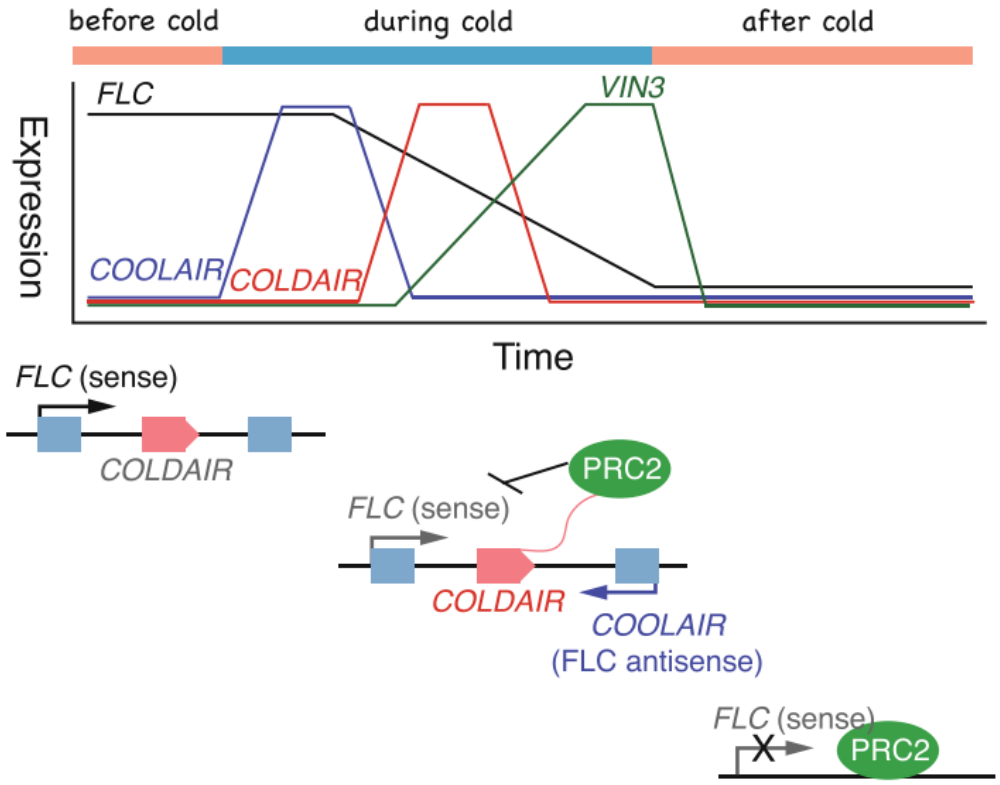


Fig. 2: Negative regulation of *FLC* during vernalization. *FLC* is actively transcribed before cold exposure. Onset of cold temperature leads to accumulation of antisense transcript COOLAIR. In the next period COLDAIR is transcribed from the first intron of FLC. It represses *FLC* sense transcripts via recruiting the PRC2 complex. After COLDAIR accumulation, the *VIN3* expression is gradually increased. It depends on the length of cold exposure. *VIN3* levels return to the same level before the cold once back to normal temperature. This transient expression of *VIN3* is also a key to the initial repression of *FLC*. The repression of *FLC* sense transcripts is maintained by activity of PRC2 after returning to warm temperature (reproduced from: [37]).

The highest level of *VIN3* expression is achieved after 4 – 6 weeks of cold (Fig. 2) [38]. VIN3 recruits VERNALIZATION 5 (VRN5 - its constitutively expressed homologue) to the first few nucleosomes of *FLC* first intron region, so called nucleation region [38–40]. VIN3 is possibly also interacting with POLYCOMB REPRESSIVE COMPLEX 2 which is responsible for trimethylation of lysine 27 at the histone 3 [36,41]. VIN3 and VRN5 are thus responsible for maintaining repressed state of FLC1, even after induction of higher temperatures. There were other genes identified playing a similar role in this process - *VERNALIZATION 1* (*VRN1*), *VERNALIZATION 2* (*VRN2*), *LIKE HETEROCHROMATIN PROTEIN 1* (*LHP1*) and autonomous pathway genes [38,40,42–45]. Local chromatin modifications of *FLC* are thus responsible for their own epigenetic heritability through multiple cell divisions (reviewed in: [46]).

Temperature influences also later stages of flowering. Recently, the role of *HEAT SHOCK PROTEIN 70* (*HSP 70*) was connected with plant response to increased temperature and to acceleration of flower development. Higher temperature affects *FLC* expression on histone level. Histone H2A is replaced with H2A.Z by ACTIN RELATED PROTEIN (ARP6) during higher ambient temperature. ARP6 is part of SWR1 chromatin remodeling complex. H2A.Z binds to *FLC* promoter and is responsible for tighter DNA wrapping and thus for repression of *FLC* gene expression [47]. Understanding of ambient temperature pathway is only at the beginning and just a few genes have been already identified and characterized (reviewed in: [48]).

2.1.1.5 Photoperiod pathway

Different latitudes and altitudes are characteristic by different day length which is changing throughout the year. Due to the correlation of suitable conditions with certain day length, the plant is able to determine place of growth and the part of the year which helps plant to set up its life cycle in order to achieve optimal conditions for flowering and seed development.

Photoperiodic pathway is responsible for perception of day length. The backbone of the pathway is represented by *GIGANTEA* (*GI*)-*CO*-*FT* genes. Activity of *CO* is responsive to light and the circadian clock. *CO* transcription is induced by the interactions of plant-specific protein GI, ubiquitin ligase FLAVIN-BINDING, KELCH-REPEAT-BOX1 (FKF1) and photoreceptors (phytochromes and cryptochromes)[49].

The *CO* expression is regulated in circadian manner and peaks in the evening, approximately 20 and 16 hours after dawn under SD and LD conditions, respectively [50]. The expression is during LD further modulated by other genes: *GI*, *FKF1* and *CYCLING DOF FACTOR* (*CDF*). Regulation of *CO* by GI and FKF1 is functional during LD only, because expression peak of these genes is overlapping just during LD conditions (13 hours after dawn) [51–53]. Interaction of proteins coded by these genes is mediated by N-terminus of GI and light, oxygen or voltage (LOV) domain of FKF1. In addition, their interaction is enabled by blue light, which is perceived by flavin-binding domain of FKF1 [53]. This complex is interacting with CDF, which then binds to *CO* promoter and represses its transcription [51,52].

Regulation of CO protein level is facilitated by two different mechanisms. COP1 and members of the SUPRESSOR PHYA-105 (SPA) protein family are taking part on the first one. COP1 is ubiquitine ligase and is responsible for degradation of CO by 26S proteasome-dependent pathway. The WD repeat domain (usually ending with tryptophan –aspartic acid (W-D)) of COP1 is responsible for COP1-CO interaction. Furthermore, the proteasome degradation of CO during SDs is enhanced by interaction SPA proteins. This interaction is again mediated by WD repeat domain (SPA) and CCT (*CO*, *CO-like*, *TOC1*) domain (CO) [54–56]. The second way of CO protein level regulation is through membrane-bound E3 ligase *DAY NEUTRAL FLOWERING* (*DNF*). Its role is targeting CO for proteasome [57].

The proteasome degradation of CO takes place during the dark. In the SD the *CO* expression overlaps with dark period and CO protein is degraded. On the contrary, during LD the expression peaks in the light and CO protein is stabilized by interaction with Ppd protein. The high CO protein level subsequently influences *FT* expression. The regulation of *FT* is complex and taking place also at the epigenetic level by PRC2 [58–60]. The *FT* expression is regulated in 24-hours period with the maximum in the evening and is transcribed in the phloem companion cells [50,61,62]. FT was identified as the florigen. Florigen is mobile factor transported from leaves through phloem to apical meristem. It transmits signal for initiation of flowering [63]. Its existence was suggested in the pioneer flowering time study of Čajlachjan in 1937. In the apical meristem, FT is interacting with FLD and activates expression of flowering transition genes as *APETALA 1* (*AP1*) and *LFY*. *AP1* and *LFY* activates A-class genes (Fig. 3) responsible for flower development [64,65]. The B- and C- classes of flower development genes (Fig. 3) are regulated by *SVP*, *SOC1* and *AGL24*.

These classes of genes are responsible for development of four different types of floral organs and their role is explained in ABC model of flower development. Enlistment of genes into one of these classes is established on phenotypic effect of their deletion mutants (A-class affecting sepals and petals development, B-class petals and stamens development and C-class affecting stamens and carpel). Later the fourth group of genes was called E-class and was added into extended ABCE model (Fig. 3). E-class genes (*SEPALLATA*) are functional in all whorls and are required for organ specification. Their ectopic expression leads to transformation of leaves into floral organs. Subsequently, integration of information from all pathways results in

activation of so called floral meristem identity genes (reviewed in: [66]).

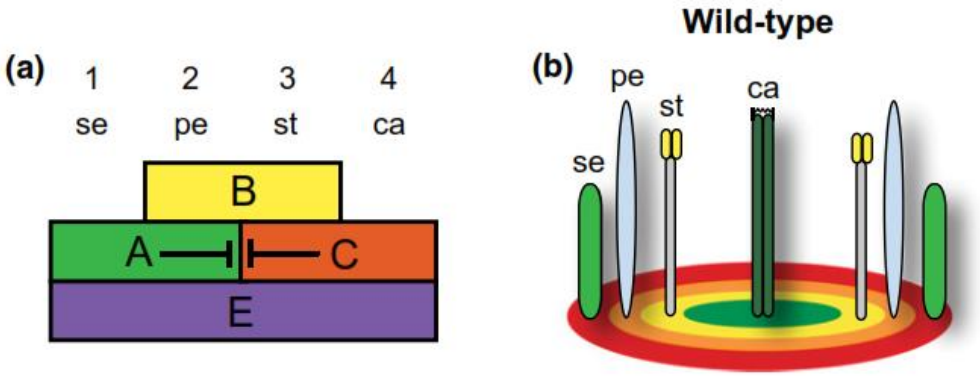


Fig. 3: ABCE model of flower development. Genes responsible for development of floral organs are divided into four classes: A-class affecting sepals and petals development, B-class petals and stamens development,C-class affecting stamens and carpel and E-class genes which are required for organ specification and are functional in all whorls (reproduced from: [66]).

Knowledge of flowering pathway gained by study of model organism *Arabidopsis thaliana* facilitates understanding of mechanism of this process. These pieces of information can be further extrapolated to agriculturally important plants.

2.1.2 *Oryza sativa*

Rice is important model organism for studying flowering time in crops. Rice, together with maize and wheat are responsible for more than 50% of world food production. Rice is a staple food for 50% of human population. It has the smallest genome within major cereal crops (1C ~ 430 Mbp) [67,68]. Rice requires SD conditions for initiation of flowering and it serves as a model organism for studies of flowering time determination in SD plants. The regulatory genes responsible for initiation of flowering in SD and LD plants are similar (reviewed in: [49]).

The florigen - mobile signal - is coded by *Heading date 3a* (*Hd3a*) gene which is orthologous to *FT* gene in *Arabidopsis* [69,70]. Also, the orthologues of other *Arabidopsis* flowering genes have been identified in rice: *GI = Oryza sativa GIGANTEA* (*OsGI*), *CO = Heading date 1* (*Hd1*) [71–73]. The pathway determined by these genes is similar to *Arabidopsis*. *OsGI* serves as an suppressor of flowering and its role is to activate *Hd1* expression [71]. *Hd1* has function as an initiator of *Hd3a* expression under SD conditions but as a repressor of *Hd3a* under LD conditions. This conversion of *Hd1* function is allowed by different timing of phytochrome signaling pathway [74]. mRNA level of *Hd3a* shows diurnal pattern

during SD but it is strongly suppressed during LD [71]. In addition to regulation by *Hd1*, *Early heading date 1* (*Ehd1*) gene also positively regulates *Hd3a* expression during SD conditions, but independently of *Hd1*. *Hd3a* is also able to initiate flowering in the absence of *Hd1* [75]. *Early heading date 1* (*Ehd1*) has no orthologous gene in *Arabidopsis*. It is a B-type response regulator protein and it binds DNA through its GARP-domain (named after: *GOLDEN2*, *ARR B*-class proteins and *Psr1*) [75,76]. *Edh1* is activated through OsGI by two pathways. The one is dependent on *OsMADS51* gene and the other is influenced by blue light signaling and lead to increase of expression at the beginning of light period [77,78]. *Edh1* is repressed during LD conditions by *Grain number, plant height and heading date 7* (*Ghd7*). *Gdh7* encodes CO-like protein which contain B-box and CCT domain and is orthologous to *VRN2* [79,80]. Rice plants planted in southern part of Asia have predominantly strong alleles of this gene and weaker and non-functional one are present in varieties planted in northern parts of Asia [80].

Protein expressed from *Hd3a* gene is then transported to apical meristem where it creates complex with 14-3-3 protein. This complex is transported to the nucleus where it interacts with OsFD1 (orthologue of *Arabidopsis* FD) and induces transcription of *OsMADS15* (orthologue of *Arabidopsis AP1*) [81–83]. Variation in promoter sequence of *Hd3a* gene, in the expression level of *Edh1* and polymorphism in Hd1 function are responsible for differences in flowering time between SD rice varieties [81].

Flowering of some rice varieties during LD conditions is enabled by second mobile signal protein coded by *RICE FLOWERING LOCUS T1* (*RFT 1*) gene. This one is regulated by *OsMADS50* (orthologue of *Arabidopsis SOC1*) through *Edh1* [81]. *RFT1* and *Hd3a* are paralogues and are located on the chromosome 6. They arose by tandem duplication and have 91% similarity in deduced amino acid sequence. These genes work redundantly and presence at least one of them is sufficient for flowering initiation. The expression of *RFT1* is very low, but it is increasing in *Hd3a* deletion lines. Flowering of plants with disrupted *Hd3a* is then delayed. Similar to Hd3a, the RFT1 is transported from leaf to apical meristem. Plants with both genes inhibited are not able to flower even after 300 days from sowing [84,85].

Rice and *Arabidopsis* flowering pathways share many similarities (Fig. 4). However, flowering is initiated by SDs in rice and by LDs in *Arabidopsis*. This is caused by

bifunctional Hd1 protein which represses flowering during LDs but enhances during SDs [74]. Although there are evolutionally well conserved genes shared by rice and *Arabidopsis*, still some unique genes and pathways have evolved to evoke flowering in rice. Especially, *Edh1* gene, which has no orthologous gene in *Arabidopsis* as well as the presence of bifunctional Hd1. Very unique adaptation is existence of second mobile signal (RTF1 protein). This adaptation is responsible for flowering during LDs and eventually for ability to grow rice under different environmental conditions [84].

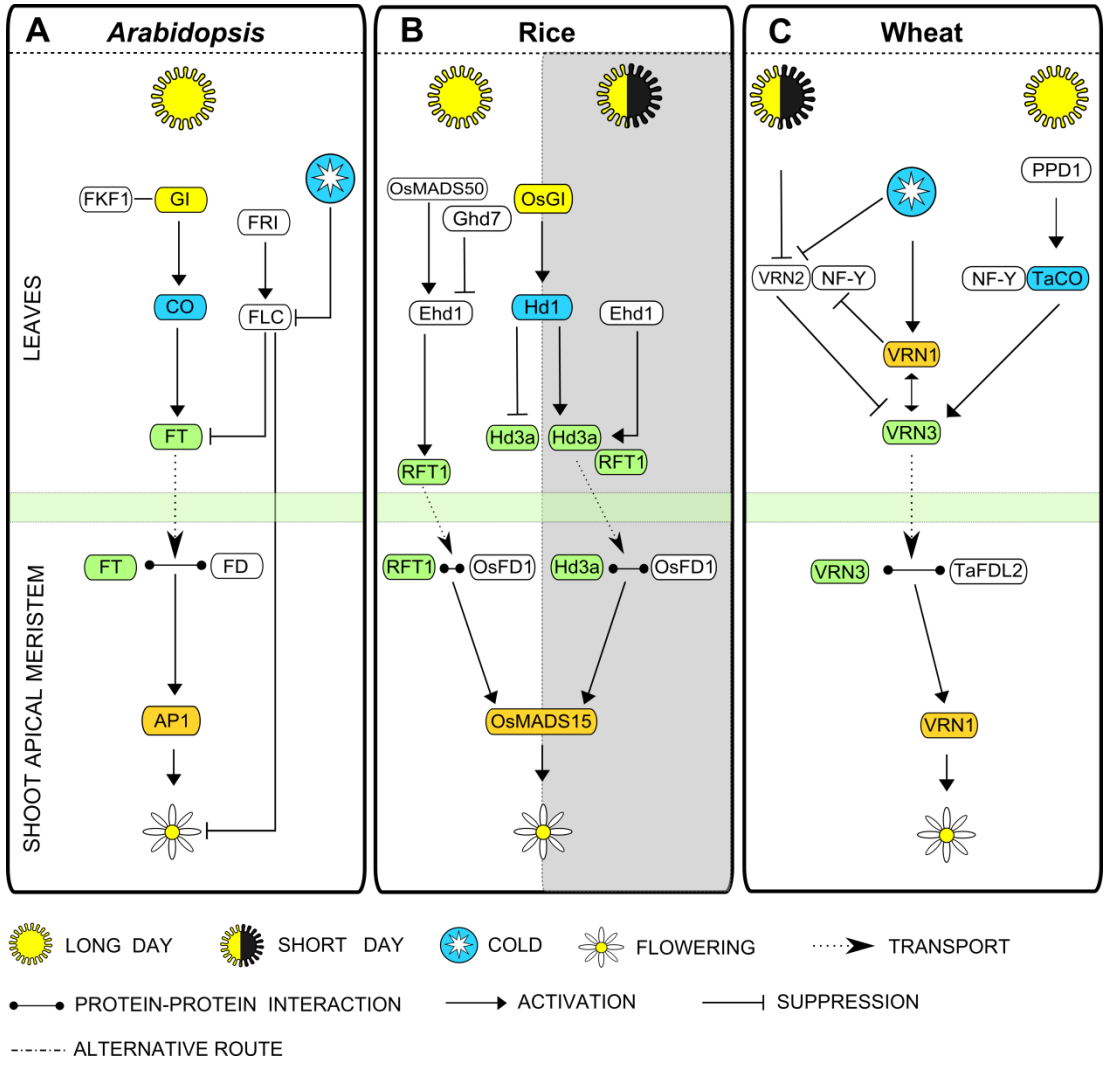


Fig. 4: Schematic comparison of the key genes involved in the flowering pathways in *Arabidopsis*, rice and wheat. In both *Arabidopsis* and wheat (A and C), the induction of flowering is promoted by LDs and vernalization, whereas in rice (B) the inflorescence is primarily induced by SDs. Despite different origins and evolution, certain genes and mechanisms are conserved among these species. The *Arabidopsis* homologue genes are highlighted in the same colour: yellow — GIGANTEA (GI), blue — CONSTANS (CO), green — FLOWERING LOCUS T (FT) and orange — APETALA 1 (AP1). The interaction of genes which lead to perception of signal from optimal external condition through transport the signal to leaves to initiation of flowering is described in detail in the chapter 2.1. The scheme is adapted from: [86].

2.1.3 *Triticum aestivum*

Bread wheat (*Triticum aestivum*) is grown on more areas than any other crop plant and is a staple food for 30% of the world's population. The wheat production provides 20% of the calories (http://faostat3.fao.org/). Bread wheat is a hexaploid species (6x = 2n=42, AABBDD) that originates from two interspecific hybridization events. Its allohexaploid nature together with prevalence of repetitive DNA (> 90%) results in huge genome size (1C - 17 Gbp). These characteristics hamper the molecular analysis of the wheat genome and thus determination of flowering gene pathways is lagging behind rice and other crops. Current progress in wheat genomics and acquiring of whole genome sequence should efficiently overcome these obstacles.

Development of new varieties with flowering tuned to local environmental conditions is one of the possible steps for increasing wheat yield. Different combinations of flowering genes/alleles are convenient in distinct environmental conditions. Varieties grown in northern latitudes or higher altitudes need to deal with long cold period. On the contrary, varieties suitable for southern latitudes are endangered by hot and dry summers. Very early flowering prevents overlapping grain filling period with these environmental conditions and eventually leads to higher yields. The holistic knowledge of flowering determination is an essential step in the process of development of new cultivars. Differences in flowering determination between monocots and dicots emphasize necessity of specific research of *T. aestivum*.

Flowering time in wheat is determined by set of genes which could be divided into three groups: (i) genes responsible for vernalization requirement - the vernalization (*VRN*) genes, (ii) genes responsible for perception of day length - the photoperiod (*PPD*) genes and (iii) genes responsible for fine tuning of flowering time - Earliness *per se* (*Eps*) genes. Allelic variation of *VRN* and *PPD* genes is responsible for 70-75% and 20-25% genetic variability in flowering time, respectively. Those variations are involved in successful spread of wheat varieties to miscellaneous parts of world. The variation in *Eps* genes is responsible for 5% of genetic variation in flowering time of wheat varieties [87]. These three groups of genes are responsible for setting up optimal time of flowering and for ensuring of survival of plant to the next generation.

2.1.3.1 Vernalization pathway

Vernalization process permits initiation of flowering after continuous chilling treatment. The wheat could be divided into two groups according to vernalization requirement. Wheat varieties with winter growth habit need cold period for flowering unlike the spring ones. However, the flowering in some spring varieties can be accelerated by previous cold exposure. Spring growth habit emerged independently many times during wheat evolution (reviewed in [88]).

The vernalization pathway consists of three major genes: *VRN1*, *VRN2* and *VRN3* (Fig. 4) [79,89,90]. Recently, the fourth *VRN4* gene was identified and characterized [91]. Variations in regulatory sequences of these genes have tremendous effect on flowering time and lead to emergence of spring varieties. The spring habit is conferred by dominant *Vrn-1*, *Vrn-3* and *Vrn-4* alleles and recessive *vrn-2* allele (Fig. 6) [79,89–94].

2.1.3.1.1 *VRN1* gene

VRN1 has an important bifunctional role in the initiation of flowering. It is closely related to *A. thaliana* paralogous genes *AP1*, *CAULIFLOWER (CAL)* and *FUL* [89,95]. The *VRN1* expression increases with the longitude of cold period and is responsible for increased expression of *VRN3* gene (Fig. 4) [96,97].

VRN1 is member of MADS box protein family [89,98,99]. Einkorn wheat *VRN1* deletion lines were not able to flower, which proved essentiality of this gene for flowering [100]. Later, Distelfeld et al. [101] showed the deletion in *T. monococcum* L. lines was not limited just to *VRN1* but includes also other loci (*Phytochrome C*) supposedly responsible for this nonflowering phenotype [102]. In addition, Chen and Dubcovsky [98] reported that *Δvrn1* bread wheat lines were able to flower and showed that paralogous MADS box proteins could functionally replace *VRN1* gene, e.g. *FUL2* and *FUL3* [98].

The *VRN1* genes are located on the long arm of homoeologous group 5 chromosomes (5AL - *VRN-A1*, 5BL - *VRN-B1* and 5DL - *VRN-D1*) [103–105]. Different *vrn1* winter alleles influence flowering time in quantitatively different ways and are responsible for difference in flowering time and length of vernalization requirements [105,106].

The *VRN1* regulatory elements are located in promoter region and in the first intron [89,105]. Variation in the promoter and in the first intron region was shown to be responsible for loss of vernalization requirement and emergence of spring wheat varieties (Fig. 5) [79,105].

The variation connected to spring wheat was identified in all three subgenomes and many spring alleles were identified so far (reviewed in: [88]).

Vrn-A1a spring allele is characterized by insertion of Miniature Inverted-repeat Transposable Element (MITE) named Spring foldback element (SFE) in promoter region. SFE is present in two copies with size of 222 bp and 131 bp, respectively. The second copy contains 91 bp deletion [94]. So far, two theories explain effect of *VRN-A1* disruption by SFE insertion. Yan et al. [94] suggested disruption of repressor binding site and Yu et al. [107] suggested positive effect of SFE on expression because of MUTATOR – like effect of this transposome [94,107]. In addition, another variant carrying 16 bp deletion upstream of 91 bp deletion were identified [108]. Yu et al. [107] found out that insertion of this element is responsible for development of very stable hairpin loops on mRNA. The insertion of this element also leads to production of miR1123 which is greatly upregulated by aging [107].

The 20 bp deletion in 5´UTR (untranslated region) and two single nucleotide polymorphisms (SNPs) in host directed duplication (HDD) define *VRN-A1b* allele [105].

The *Vrn-A1c* allele posses large deletion in the first intron [105]. There were different deletions identified in distinct wheat varieties so far. Further analysis identified 4 kbp conserved region which presumably contains regulatory elements essential for *VRN1* repression (Fig. 5) [105].

Vrn-A1d and *Vrn-A1e* alleles have been identified in tetraploid wheat and contain 32 bp and 54 bp deletion in the HDD and CArG box, respectively [105]. Pidal et al. [109] doubted the essentiality of CArG box for spring habit. Instead of this, they identified VRN box as the possible cause of spring phenotype [109]. Muterko et al. [108] underlined importance of VRN box, when they described responsibility of even one nucleotide substitution in A and C rich regions of VRN box for determination of growth habit [108].

Vrn-A1f allele has specific deletion (8 bp and 50 bp) in the promoter region; MITE insertion (424 bp) and 2,753 deletion in first intron [110,111]. Recently, the MITE insertion was shown to be solely responsible for spring habit emergence [112].

Vrn-B1a allele has 440 bp deletion in first intron. *Vrn-B1b* allele has one more 36 bp deletion and the SNP [113]. *Vrn-B1c* contains a 5,463 bp insertion of retrotransposon in 5´UTR [114]. The *Vrn-B1d* allele has 817 bp deletion accompanied by duplication/translocation of 450 bp in the first intron [115,116].

Vrn-D1a is characteristic by deletion in intron 1. Recently, the SNP was further connected to emergence of spring *Vrn-D1b* and facultative *Vrn-D1c* alleles. The SNP in *Vrn-D1c* allele disrupted MADS box protein binding site [117].

The *Vrn-A1* is epistatic to other *Vrn-1* alleles (*Vrn-B1* and *Vrn-D1*) and has the most profound effect on flowering time. The strength of effect on phenotype is decreasing as follows: *Vrn-A1*, *Vrn-D1*, *Vrn-4* and *Vrn-B1* [118]. Unlike *Vrn-A1*, the *Vrn-B1* and *Vrn-D1* spring alleles are responsible for decreasing but not diminishing of vernalization requirement [119]. Moreover, the *Vrn-A1* is expressed already in the first leaf stage. The other two (*Vrn-B1* and *Vrn-D1*) are detected in the second and third leaf stage [99].

Various spring alleles have dissimilar effect on flowering time and confer optimal response to different environments. Thus, their frequencies vary in different latitudes and altitudes. The frequency of the dominant *Vrn-D1* allele is increasing towards to equator, while the frequencies of *Vrn-A1* and *Vrn-B1* are decreasing [87]. High frequencies of *Vrn-D1* were identified also in central Asia and in the landraces from Pakistan, Afghanistan and China [120–122]. The *Vrn-D1* allele is responsible for adaptation to different environmental conditions by accelerating or delaying flowering time [123] and is present in three the most agriculturally used cultivars worldwide (Pastor, Attila and Kauz) [124]. The European varieties possess *Vrn-B1a* gene and *Vrn-A1a* prevails in high latitudes [87,125].

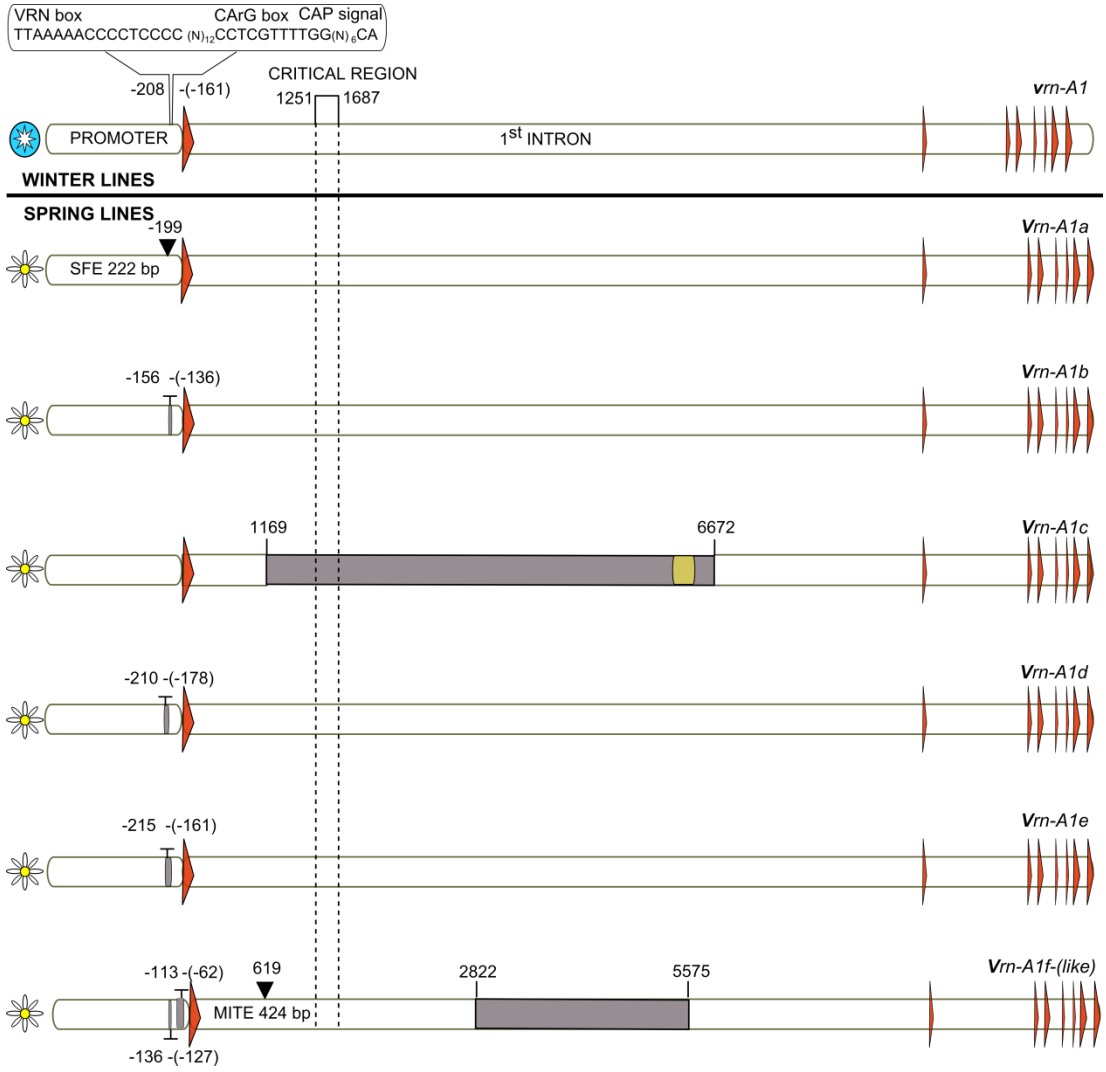


Fig. 5: The *VRN-A1* alelles. The schematic representation of sequence changes in ***dominant*** *Vrn-A1* alleles present in ***spring wheat*** compared with **recessive** *vrn-A1* (**winter wheat**). The exons are marked by orange arrows and introns by white rectangles. Promoter is illustrated by white rectangle in front of the first exon. The positions of important regulatory elements in the promoter (CArG box, VRN box and CAP signal [89,109]) and the 1st intron (critical region – delimited by dotted lines [105]) are outlined in the *vrn-A1*. Each dominant allele is characteristic by insertion or/and deletion in promoter or/and the first intron which are designated by black triangle or grey rectangle, respectively. Their position is indicated by number of base pairs from the START codon in comparison to *vrn-A1* allele. The yellow rectangle symbolizes duplication in the *Vrn-A1c* sequence. The NCBI database sequences used for schematic comparison: *vrn-A1* (AY616455.1), *Vrn-A1a* (AY616458.1), *Vrn-A1b* (AY616461.1), *Vrn-A1c* (AY747599.1), *Vrn-A1d* (AY616462), *Vrn-A1e* (KT361213), *Vrn-A1f* (GQ451750.1) and *Vrn-A1f-like* (KT696537.1). It is important to note that the database sequences of *Vrn-A1a*, *Vrn-A1b*, *Vrn-A1c*, *Vrn-A1d*, *Vrn-A1e* and *Vrn-A1f* do not contain whole gene sequence but just promoter region and partial CDS. The sequence of *Vrn-A1f* and *Vrn-A1f-like* were merged together in scheme. *Vrn-A1c* allele contains sequence of the 1st intron only.
SFE – Spring Foldback Element, MITE – Miniature Inverted-repeat Transposable Element.

2.1.3.1.2 *VRN4* gene

Recently, the new gene responsible for loss of vernalization requirement was identified on the short arm of chromosome 5D. This gene evolved by extensive duplication of region from 5AL chromosome and it is paralogous to *VRN1* gene. Kippes et al. [93] identified three SNPs in the first intron which were segregating with spring habit and were considered responsible for emergence spring habit in some Chinese wheat cultivars [93].

2.1.3.1.3 *VRN2* gene

VRN2 gene is considered to be a major repressor of flowering in wheat. Well described is its role in *VRN3* repression before arrival of cold conditions during winter. Its expression is decreased by SD and cold treatment [79,98]. *VRN2* was localized on 5A and 5B chromosomes in tetraploid wheat. *VRN2* contains Zinc-finger protein domain and CCT domain and are very often called ZCCT genes. The CCT domain is responsible for interaction with NF-Y proteins. ZCCT protein works antagonistically to CO protein and blocks binding of CO to *VRN3* promoter and inhibits its expression [79,98,102].

Interrelationship between *VRN1* and *VRN2* has been discussed in many studies. Yan et al. [89] supposed that *VRN2* represses *VRN1*. The low temperature decreases *VRN2* expression and releases the *VRN1* from its repression. Loukoianov et al. [99] suggested that cold induces *VRN1* expression and *VRN1* further represses *VRN2*. Decreased expression of *VRN2* removes repression of *VRN3* (Fig. 6) [101]. Recently, Dubcovsky et al. [126] observed so called “short day vernalization”. The sufficiency of SD for repression of *VRN2* and subsequently for flowering questioned role of *VRN2* in repression during winter and introduced putative presence of another repressive protein [126].

The spring habit is also connected with recessive *vrn-2* allele (Fig. 6). This spring allele has mutation in conserved CCT domain [79]. For emergence of barley spring habit the whole gene deletion of *VRN2* was also responsible [79,127].

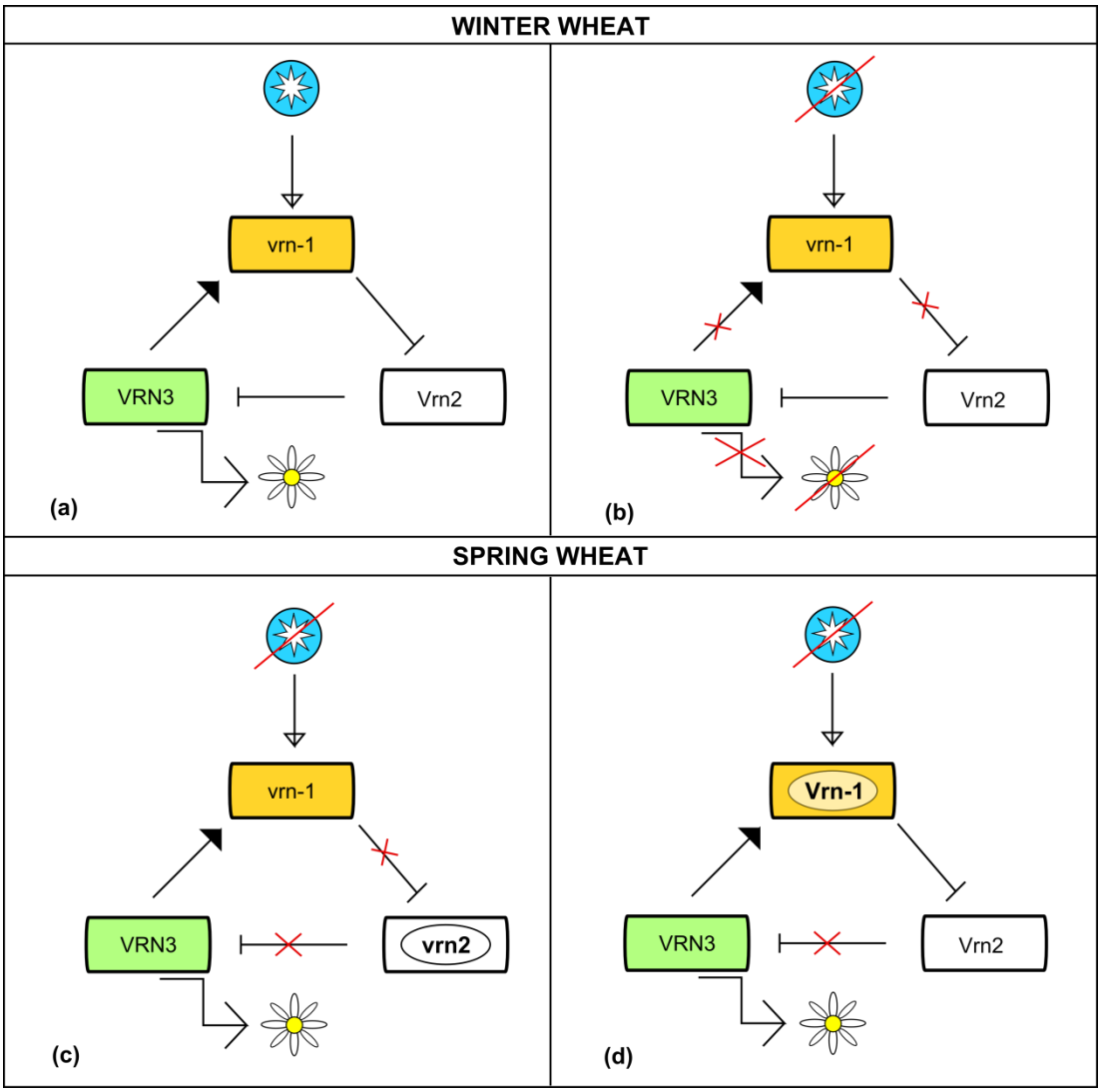


Fig. 6: The schematic role of *VRN* genes for determination of winter and spring habit in wheat. The interaction between three main *VRN* genes and their allelic variants is responsible for necessity of cold requirement for flowering. (a) Winter wheat when cold requirement is fulfilled. (b) Winter wheat when cold requirement is not fulfilled. (c) Spring wheat with recessive *vrn2* allele which initiates flowering without previous cold period (removes downregulating activity of vrn2 protein on *VRN3* gene). (d) Spring wheat with dominant *Vrn1* allele which initiates flowering without previous cold period (enables repressing inhibitory effect of VRN2 on *VRN3* gene). Arrows and T bars *represent* induction and repression activity of proteins, respectively. The crosses represent absence of this activity. Stars represent cold period and flower represents induction of flowering. Crossed pictograms of cold period or flower means absence of cold period or flowering, respectively. Circles mark genes which were structurally changed in wheat lines and are responsible for spring habit.

2.1.3.1.4 *VRN3* gene

VRN3 gene is coding mobile part of flowering pathway which transfers flowering signal from leaves to apical meristem. The gene was originally considered vernalization gene, but it is orthologous to *FT* gene of *A. thaliana* [90].

VRN3 homoeologous genes were identified on 7A, 7B and 7D chromosomes [90,92,128]. They are classified as RAF kinase inhibitors. They work through

repression of MAP kinase pathway. *VRN3* gene is responsible for activation of *VRN1* expression. As mention above, it is orthologous to *A. thaliana FT* gene. Unlike *FT* in *A. thaliana*, the higher level of *VRN1* expression in wheat is detected not only in the apical meristem but also in the leaves [129]. However, *FT* ectopic expression in *A. thaliana* leaves leads similarly to increase of *VRN1* expression [64]. Recently, the duplication of this *VRN1* gene in all members of grasses family was identified. This duplication leads to emergence of *Triticum aestivum FLOWERING LOCUS T 2* (*TaFT2*) gene which could play a role in flowering pathway (Li and Dubcovsky 2008). Duplication which leads to *TaFT2* gene preceded divergence of grasses but not divergence of dicots and monocots [129].

Two VRN3 monomers are interacting with two *Triticum aestivum* Flowering Locus D – like (TaFDL) proteins (DNA-binding bZIP transcription factors) and then this complex binds to promoter region of many genes and influence their expression, among others *VRN1* and *TaFT2* [90,129]. This interaction is mediated by 14-3-3 protein bridge. The final hexameric complex is transported from cytoplasm to nucleus [130]. 14-3-3 proteins are essential for this transportation while the *TaFDL* genes are responsible for interaction of complex to ACTG domains in the *VRN1* promoter region [129]. As mentioned above, it regulates also expression of *TaFT2*. TaFT2 protein can potentially interact with other TaFDL factors and influences expression of different flower meristem identity genes. Li et al. [130] discovered ability of VRN3 protein to interact with different 14-3-3 and TaFDL (TaFDL2 and TaFDL6) proteins. Different permutation can affect different genes and this can serve in decoding of information carried by *VRN3* proteins. It is important to mention that just TaFDL2 was essential for interactions with *VRN1* promoter region [130]. The *TaFDL2* expression is influenced by neither the growth stage of apical meristem nor the day length and cold treatment [129].

Similarly to *VRN1,* the dominant *VRN3* allele confers spring phenotype. Allelic variation of *VRN3* gene was identified at the 7B loci. Three spring alleles (*Vrn-B3a*, *Vrn-B3b* and *Vrn-B3c*) were detected so far. The *Vrn-3Bb* allele has 890 bp insertion in 5′UTR region while *Vrn-B3c* allele has 5,300 bp insertion and 20 bp and 4 bp deletion in 5′UTR [90,131].

2.1.3.2 Photoperiod pathway

The qualitative and quantitative properties of light are important factors for flower development. The variation in day length during the year correlates with the distance from equator. The advent of winter is connected with shortening day length and on the contrary, the summer is associated with day length elongation, in the direction to poles. Plants are using this correlation of day length and suitable environmental conditions for determination of optimal flowering time. In other words, plants growing in distinct latitudes are adapted to different day length conditions for optimizing of flowering.

Photoperiod has more important role in timing of flowering in spring cultivars compared to the winter ones. The reason is that in winter cultivars, the photoperiod influences flowering time after satisfaction of vernalization requirement [132]. However, the spring cultivars grown in higher latitudes are not influenced by photoperiod because day length is long enough to satisfy photoperiod requirements [133,134].

2.1.3.2.1 The role of photoreceptors in flowering

The important role in flowering has also different light quality. Flowering is accelerated when plants experience shift in the light quality from red (620-750 nm) to far red (710-850 nm) light spectrum. This shift is connected with suboptimal light conditions caused by rate of growth of other plants and leads to accelerated stem elongation and precocious flowering, so called “shade avoidance” mechanism. The day neutral response can be induced by high intensities of blue (450-495 nm) light which indicate importance of different light lengths in flowering acceleration (reviewed in: [135]). Plants measure the length of the day with specialized proteins called photoreceptors. These are able to percept light with different wave length and thus plant is able to distinguish the (sub)optimal external conditions and to adjust life cycle. There are few main classes of photoreceptors: light-oxygen-voltage (LOV) sensors, xanthopsins, phytochromes, blue-light sensors using flavin adenine dinucleotide (BLUF), cryptochromes, and rhodopsins [136]. Two major plant photoreceptors are phytochromes (red/far red light) and cryptochromes (blue/UV-A). The function of these receptors in light perception and flowering pathway will be discussed in the following subchapter.

Phytochromes control plant developmental processes starting with seed germination and seedling de-etiolation. They are also responsible for perception of neighboring plants and shade and influence transition from vegetative to generative stage [137]. The phytochromes are particularly responsible for perception of red and far-red light. There are three classes of phytochromes in monocot plants: *phyA, phyB* and *phyC*. The fourth and fifth class in dicots raised by duplication of *phyB* [138,139]. Phytochromes contain covalently bound linear tetrapyrrole chromophores (bilin) and create dimmers [140]. PhyA is synthesized in inactive red-light absorbing form - Pr (660 nm) and after exposure to red light is photoconverted into biologically active far-red absorbing form - Pfr (730 nm). In Pfr form, the phytochromes are translocated into nucleus where they regulate transcriptional signaling network [140–143]. Reversely, the exposure to far-red light leads to conversion to red-absorbing form [136].

Cryptochromes are responsible for photomorphogenesis, flowering and clock resetting in plants [135]. They are flavoproteins and are able to perceive blue and UV-A radiation (320-520 nm) [144]. Cryptochromes have two important domains. The light is perceived by photolyase homology region domain (PHR) at the N-terminal end and the interaction with other proteins and thus signal transfer is mediated by CCT domain located at the C-terminus [145,146]. Three cryptochromes have been identified in *Arabidopsis*. Proteins CRY1 and CRY2 are localized predominantly in the nucleus [147]. CRY1 is interacting with COP1 and SPA proteins in the light dependent manner and lead to transcription of HY5 gene [148,149]. This regulates transcription of genes important for de-etiolation. On the other hand, the CRY2 was shown to interact with cryptochrome-interacting basic-helix-loop-helix (CIB1) protein. This interaction accelerates *FT* transcription and subsequently flowering [150]. CRY3 is localized in mitochondria and chloroplasts [151] and was found to be responsible for repairing damages of DNA [152,153].

2.1.3.2.2 The genetic determination of photoperiod (in)sensitivity

Originally, the wheat needs LD for flowering. Photoperiod sensitivity is predominantly determined by series of photoperiodic (*Ppd*) genes (*Ppd-A1*, *-B1*, *-D1*) located on short arm of chromosomes from the second homoeologic group [154,155]. Kheleskina et al. [156] identified new *Ppd-B2* gene located on the *7BS*

chromosome. In contrast to *Ppd-1* genes which accelerate flowering irrespective of day length, the *Ppd-B2* accelerates flowering during LD only [156].

The *Ppd* genes belong to pseudo-response regulators (*PRR*) gene family [157] and contains CCT domain responsible for protein-protein interaction [158]. This domain is related to components of circadian clock genes. *Ppd* genes do not have a role in circadian regulation, but they are regulated by circadian clock mechanism [157,159].

Photoperiod insensitivity is caused by misexpression of *Ppd* genes. The insensitivity is dominant over sensitivity; dominant alleles have been given suffix *a*, while recessive alleles have been designated with suffix *b* for better determination. [155]. Different *Ppd-A1a*, *Ppd-B1a* and *Ppd-D1a* alleles lead to independent emergence of photoperiod insensitivity in many different wheat varieties [106,155,160–162].

To date, two different origins of insensitive alleles were examined: (i) structural changes in promoter region and (ii) copy number variation (Fig. 7).

(i) In this type of photoperiod insensitivity, the indel in the promoter region disrupts putative transcription factors binding sites [155,162,163]. This leads to loss of *Ppd* repression followed by its expression resulting in acceleration of flowering time during SD. The *Ppd-D1a* allele has 2,089 bp deletion in the promoter region [155]. The *Ppd-A1a* insensitive alleles with 1,027, 1,117 bp and 1,085 deletions have been recognized in tetraploid and hexaploid wheat, respectively [160,162]. Recently, the *Ppd-B1a.1* allele characterised by 308 bp insertion in the promoter was identified [162]. These structural changes cover joint region of 886 bp which is presumed to be important for *Ppd1* repression. This is supported by presence of *CCA1-HIKING EXPEDITION* (*CHE*) recognition elements in this area, which was disrupted by insertion in *Ppd-B1a.1* allele [162]. *CHE* are known to be responsible for repressoric effect [164].

(ii) Díaz et al. [106] identified presence of higher number of copies of *Ppd-B1a* gene as another cause of photoperiod insensitivity in bread wheat. In addition, Sun et al. [165] identified hypermethylation of promoter connected to duplication of this region and presumably co-responsible for loss of *Ppd* repressing.

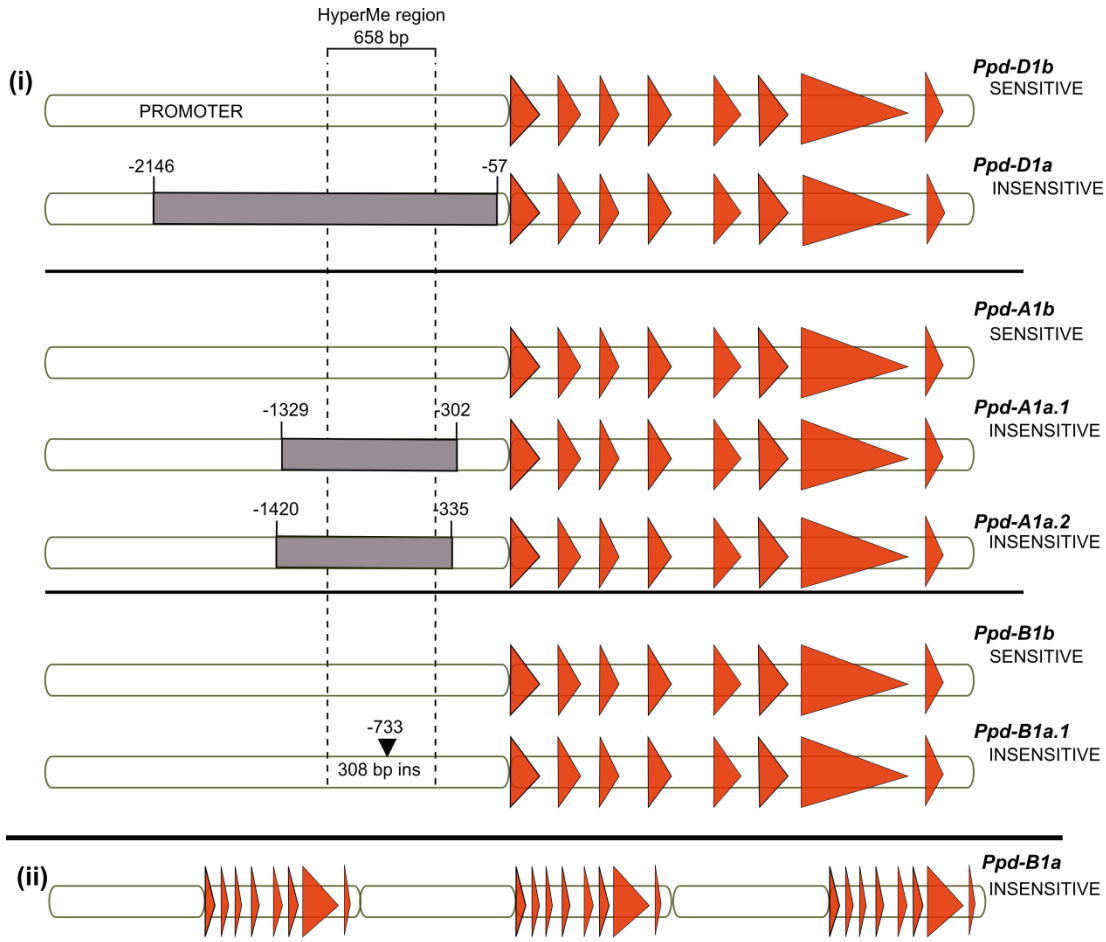


Fig. 7: The *Ppd1* alleles. The schematic representation of sequence changes in dominant *Ppd1a* alleles present in photoperiod insensitive wheat varieties compared with recessive *Ppd1b* alleles of all three homoeologous subgenomes (photoperiod sensitive wheat). The exons are marked by orange arrows and introns by white rectangles. Promoter is illustrated by white rectangle in front of the first exon. The position of important regulatory region which is common in so far all insensitive varieties and was identified as hypermethylated (HyperMe) [165] is delimited by dotted lines. Each dominant allele is characteristic by insertion or deletion in promoter which are designated by black triangle or grey rectangle, respectively. Their position is indicated by number of base pairs from the START codon in comparison to *Ppd1b* allele. The NCBI database sequences used for schematic comparison: *Ppd-D1b* (AB646977), *Ppd-D1a* (AB646976), *Ppd-A1b* (AB646972), *Ppd-A1a.1* (EU117146), *Ppd-A1a.2* (AB646973), *Ppd-B1b* (AB646975), Ppd-*B1a.1* (AB646974) and *Ppd-B1a* (JF946486). (i) *Ppd1a* alleles with structural changes in promoter region conferring photoperiod insensitivity in all three homoeologous 2 chromosomes (ins – insertion). (ii) *Ppd1a* allele with higher number of copies conferring photoperiod insensitivity identified solely on 2B chromosome.

The comparison of flowering acceleration alleles conferring photoperiod insensitivity showed the *Ppd-D1a* has the most profound effect followed by *Ppd-A1a* and *Ppd-B1a* [106]. The strength of new *Ppd-B1a.1* allele is questionable. Tanio and Kato [163] considered it the most effective flowering promoter, but Nishida et al. [162] did not confirm this dominant effect of *Ppd-B1a.1*. One of the possible explanations

of differences in the effect of insensitive alleles determined by CNV and promoter disruption is their distinct expression pattern. The *Ppd-D1a* showed misexpression (peaks three hours before onset) [155], although expression pattern of *Ppd-B1a* was similar to wild type (peaks three hours after onset), but upregulated [106].

The Ppd protein is affecting *VRN3* expression trough stabilization of CO protein. It has antagonistic relationship with VRN2. CCT domains of Ppd and also VRN2 proteins serve for interaction with NF-Y transcription factors. The VRN2 has a role especially in early fall when avoids CO stabilization, *VRN3* expression and thus flowering before winter. During winter, *VRN1* expression increases which downregulates *VRN2* and releases NF-Y for interaction with Ppd1 when sufficient day length is achieved. Ppd-CO interaction stabilizes CO protein and positively influence expression of *VRN3* (Fig. 8) [97,98].

Introduction of photoperiod insensitive alleles to wheat cultivars lead to 35% yield increase in Southern Europe and 15% in Central Europe. The most photoperiod insensitive European varieties carry the *Ppd-D1a* allele coming from Japanese cultivars Akakomugi and Saitama 27 [166]. Higher yield is mainly achieved due to balance shift in incorporation of assimilates between stem elongation process and spikelet initiation to spikelet formation exclusively [167]. The effect of insensitive alleles on yield is highly dependent on environmental conditions. Different studies showed yield decrease (-1.8%) in cool humid summer, 5% yield increase in hot dry summer or no yield changes [133,166,168]. Frequency of different sensitive and insensitive alleles is changing through different latitudes. The breeders have selected the most suitable *Ppd* allele combination just by indirect selection for agronomical important traits [169]. Generally, the winter wheat varieties grown in northern latitudes are highly sensitive to photoperiod and varieties from southern latitudes are highly photoperiod insensitive [168,170].

Recently, there were identified also different pathways repressing photoperiod sensitivity, e.g. the decrease of function of *EARLY FLOWERING 3* (*EFL3*) or *PHYTOCLOCK/LUX1* reduced photoperiod sensitivity in wheat [171–175].

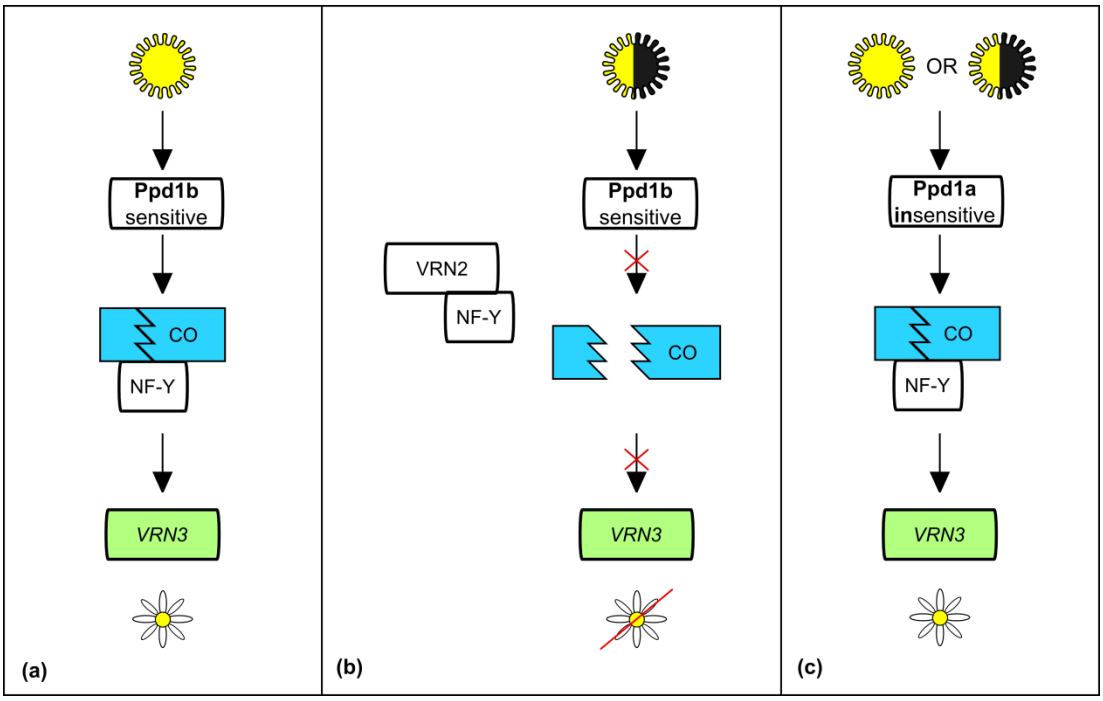


Fig. 8: The schematic model of determination of photoperiod (in) sensitivity in wheat. (a) Photoperiod sensitive wheat with recessive *Ppd1b* allele during LD conditions (>12 hours). (a) Photoperiod sensitive wheat with recessive *Ppd1b* allele during SD conditions (<12 hours). (c) Photoperiod insensitive wheat with dominant *Ppd1a* allele irrespective of day length (it is able to flower during both, SDs and LDs. Expression of *Ppd* gene leads to stabilization of CO protein by creating CO – NF-Y complex. The other way (during SDs in varieties sensitive to photoperiod) the *CO* expression is taking place during night. The NF-Y proteins are captured by VRN2 and the CO is degraded. The misexpression of *Ppd1a* leads to CO stabilization in both day length conditions. CO positively influences VRN3 expression and participates in induction of flowering. Arrows represent induction activity of proteins. The crosses represent diminishing of this activity. The full and half shadowed suns represent LD or SD conditions and flower represents induction of flowering. Crossed pictogram of flower means absence of flowering.

2.1.3.3 Earliness per se genes (*Eps*)

Eps genes belong to the third group of genes which influence flowering time. This group is created artificially and consists of genes which play role in different parts of flowering pathway. They have two main common features: they influence flowering time when the day length and vernalization requirement had been fulfilled [176] and have just minor effect on flowering time. These genes are generally mapped as quantitative trait loci (QTLs) [177]. Their effect can be determined only when overlapping effect of *VRN* and *PPD* genes is eliminated [178].

Lewis et al. [179] identified effect of *Eps-A^{m}1* from double ridge stage to terminal spikelet initiation but not from terminal spikelet initiation to heading. However, different study showed that *Eps* genes can accelerate flowering time in any stage of

development [180]. They can be used for shortening life cycle of wheat [181]. Many studies revealed their pleiotropic effect not only on flowering but also on spikelet number and number of grains per spike [173,179,182].

The first studies of these genes suggested their independence on environmental factors. However, later studies revealed modulation of effect of these genes by temperature [180,182].

The *Eps* loci are located on all wheat chromosomes of all homoeologous groups (reviewed in: [178]). Just a few *Eps* loci were positional cloned so far [173,183,184]. The *Eps-3A*^{m} gene was located on chromosome 3A of *T. monococcum* [173]. *Eps2* gene was identified on 2H chromosome of *Hordeum vulgare* L. and was determined as *CENTRORADIALIS* [183]. It codes for phosphatidyl-ethanolamine binding protein which is distantly related to FT protein family. The aminoacid substitution in this gene leads to flowering delay in spring barley varieties [183].

Furthermore, the regions of other *Eps* genes were delimitated. Three genes were identified as candidate for *Eps- A*^{m}*1* loci (*MOLYBDATE TRANSPORTER 1*, *FILAMENTOUS TEMPERATURE SENSITIVE H 4*) [182,185]. Recently, Alvarez et al. [186] identified *ELF3* gene as the candidate gene for *Eps- A*^{m}*1*. Three aminoacid changes in the protein sequence were connected to differential expression of downstream targets of this gene (*PPD1*, *FT1* and *PIFs*).

The fine mapping of *Eps* gene located on 1DL chromosome of bread wheat lead to identification *Triticum aestivum EARLY FLOWERING 3-D1* (*TaELF3-D1*) gene as a possible candidate for earliness *per se* locus [187]. They further correlated reduced expression of *TaEps3-D1* gene to altered expression of *TaGI*. However, the lack of recombination in analyzed region impeded the precise determination of the candidate gene [187].

The determination of *Eps* genes is challenging due to their subtle effect on flowering time. Detailed study of *Eps* genes needs precise genetic stocks and controlled growth conditions. However, their usage in breeding process can lead to development of varieties which will be able to meet local environmental conditions. The usage of *elf3* gene in commercial cultivars of barley was reported from Sweden, where early flowering conferred by this gene is advantageous [188].

2.2 Wheat evolution and domestication

2.2.1 *Polyploidization*

Bread wheat is allohexaploid plant and its genome consists of three subgenomes and have evolved by two hybridization events. The first hybridization, 500,000 years ago, occurred between *T. urartu* (Tumanian ex Gandilian) as a A genome donor [189,190] and the relative of *Aegilops speltoides* Tausch. as a B genome donor [191]. This hybridization was followed by polyploidization. The whole genome rearrangements lead to stabilization of tetraploid wheat subgenome, to restoring diploid-like behavior and to creation of wild emmer wheat *T. dicoccoides* (Koern. ex Asch. & Graebn.) [191]. Later, domesticated *T. dicoccum* (Schrank) Schuebl. has evolved [192] (Fig. 9). The second hybridization event occurred 10,000 years ago between yet unknow tetraploid wheat and *Ae. tauschii* Coss. and led to emergence of hexaploid bread wheat [193,194]. Since *Ae. tauschii* and *T. dicoccoides* do not have common habitat, domesticated tetraploid wheat is the most probable donor of AB genomes of hexaploid wheat. The bread wheat origin involved multiple hybridization events and subsequent intercrossing followed by formation of single gene pool [193]. The *Tg* locus of *Ae. tauschii* has led to emergence of hulled hexaploid wheat. Anyway, there was no hulled ancestor of bread wheat identified. Nesbitt et al. (2002) suggested the high selection pressure against hulled wheat led to fast emergence of free-threshing hexaploid wheat [195].

In addition, hybridization between *T. urartu* (A genome) and relative of *Ae. speltoides* (G genome) led to emergence of tetraploid wheat *T. araraticum* Jakubz. [191,196]. Wild tetraploid wheat was also subject of cultivation and later gave rise to domesticated wheat *T. timopheevii* Zhuk. (Zhuk). The hexaploid wheat *T. zhukovskii* Menabde & Ericzjan arises from hybridization between *T. timopheevii* and *T. monococcum* [189] (Fig. 9). The wheat from these groups are not used in modern agriculture but they are currently found in limited areas in Georgia [192].

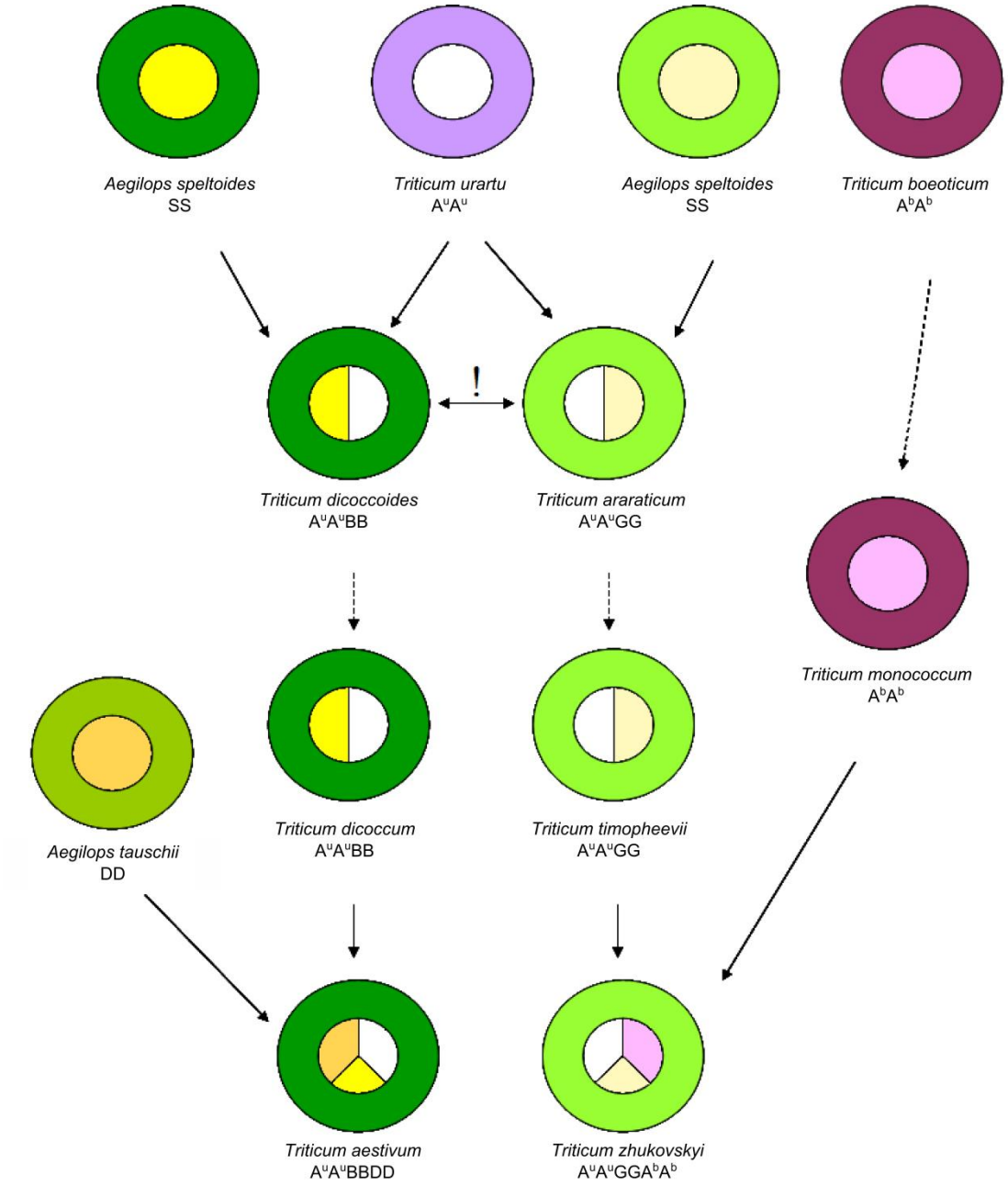


Fig. 9: The scheme of polyploid wheat evolution. The circles represent nuclear (inner) and extranuclear (outer) genetic information. Genetic information coming from different diploid wheat is indicated with different color and further the genome composition is written under the name of given species. The exclamation mark indicates uncrossability of species *T. dicoccoides* and *T. araraticum*. These species comes from cross of different individuals of the same species which was divided by thousand years and happened at different places of Fertile Crescent.

2.2.2 *Domestication syndrome*

Domestication leads to significant genetic changes which affected phenotypic traits of agronomically important wheat species. The human-induced selection is often going against natural selection [197]. The traits which distinguish domesticated wheat from wild ancestor are called domestication syndrome traits. They are usually

influencing yield, harvesting and processing of spikes (non-brittle rachis and free-threshing). However, the emergence of varieties with these important characteristics is associated with lower genetic variability [198].

Domestication process genetically influenced rachis brittleness, glum tenacity, heading date, plant height, grain size and yield components [199]. Approximately 70 domestication QTLs were located on all chromosomes but mainly clustered to 4A, 5A and 5B of *T. dicoccum* [200]. Seven domestication syndrome factors were identified and each of them influences 5 - 11 traits [199].

The first sign of domestication is loss of brittle rachis which is easy to distinguish in the excavations. Appearance of non-brittle wheat spike started approximately 10,000 years ago and allows dating of domestication event [192]. This first and very important character is conferred by genes located on chromosomes of homoeologous group 3. This trait has positively influenced yield by making harvesting process easier [201,202].

Another important trait which has effect on spike processing is emergence of free-threshing form/character. Originally, seeds were closed in chaffs which were hard to separate. Free-threshing varieties facilitate releasing of grains and also increase kernel proportion of spikelets [203]. Threshability is under the control of *Q* locus which is located on 5A chromosome. *Q* gene is a member of *AP2* family which influences flowering time in *A. thaliana*. *Q* locus influence also speltoid character of spike and rachis fragility [204,205]. Free-threshing is influenced by additional genes located on chromosomes of homoeologous group 2: *tenacity glumes* (*Tg*) gene controlling glumes toughness and non-orthologous *soft glumes* (*sog*) gene which controls glumes softness [206,207].

The short and wide grains with increased size were also selected during domestication process. These traits are controlled by different genes [208]. Loss of asymmetric dormancy is another domestication syndrome trait. Gene responsible for uniform time of germination is located on chromosome 4B. In wild emmer wheat, one kernel in the spikelet needs prolonged period of time to overcome dormancy. This trait serves to ensure probability of survival in the case of unsuitable conditions during year [209].

2.2.3 *Domestication process*

Domestication process (together with polyploidization) is another important factor influencing wheat genome plasticity. Domestication plays important role in transition from hunter-gatherer style to agrarian life style. This Neolithic revolution took place in Levant 10,000 years ago and was connected to domestication of a few staple crops in this area: barley, pea, lentil, chickpea, einkorn wheat (*T. monococcum*) and emmer wheat (*T. dicoccum*) (reviewed in: [192]). Domesticated emmer wheat appeared at the same time as domesticated einkorn wheat. The experimental data suggests that fixation of the first domestication trait – brittle rachis - could have happened in the short period after the beginning of cultivation [210]. However, cultivation of wild wheat started 19,000 years ago. This leads to idea of protracted domestication in which domestication was long and unconscious process [211]. The archaeobotanical and genetic data was used in many studies in order to reveal where, when and how emmer wheat was domesticated. However, the domestication site of wheat has not been undoubtedly determined yet. In addition, the results gained by genetic analysis do not correlate with archaeobotanical findings. These incongruences within genetic data and between genetic and archaeological data lead to suggestion of new hypothesis of reticulated origin of domesticated wheat. Civáň et al. [212] suggested spreading of pre-domesticated tetraploid wheat throughout Levant corridor and hybridization of these plants with the local populations. This enables spread of southern Levant diversity. The hybridization is mediated by humans and is beyond our ability to precisely determine place of domestication and thus emergence of domesticated tetraploid emmer wheat. This spread also facilitates fixation of domesticated traits. This hypothesis also explains unexpected degree of genetic variability of domesticated wheat [212]. The gene flow between domesticated and wild emmer gene pool could have significantly decreases founder effect [194]. The founder effect is much larger in hexaploid wheat. This is probably caused by small number of hybridization events between tetraploid wheat and diploid *Ae. tauschii*. It is also connected to decreased possibility of hybridization between tetraploid and hexaploid wheat [193,213]. Indeed, reduction of diversity is more pronounced in the D genome of hexaploid wheat. This indicates at least restricted genetic flow between hexaploid and tetraploid wheat [214].

2.3 The genetic mapping

Current wheat supplies are threatened by extreme weather changes, decreasing area of arable land and various biological threats (bacterial, fungal, viral diseases etc.). Improving currently used wheat varieties is one of the main aims for wheat research and breeding programs. Discovery of new genes and alleles which confer desired phenotype and their integration into elite varieties needs to be preceded by few steps. The genetic mapping is very often the prerequisite which, followed by QTL analysis, leads to identification of locus responsible for desired phenotype.

Genetic mapping determines relative position of anonymous specific sequences of DNA (markers) along chromosomes and assigns phenotypic trait to chromosome locus. It anonymously distributes markers along genome/chromosome. Markers are subsequently ranked according to number of recombination events. The probability of recombination events between markers is reducing with decreasing their mutual distance. Genetic map distances are determined as the mean number of recombination events within chromatid per meiosis. It is reported in centimorgans (cM). However, the number of recombination events is unevenly distributed with increasing rate towards to telomeres. Thus, the relative marker distance in cM does not reflect real physical distance in bp (reviewed in: [215]).

The amount of recombination events is limited in one individual. Thus, the mapping population ranges from several hundred to several thousand of individuals which are used for creation of genetic map. The appropriate mapping population is created from parental lines which differs in the studied phenotype and are sufficiently polymorphic at the genotypic level [216].

There are several types of mapping populations which are used for genetic mapping. An F_2 mapping population consists of members of second filial generation which was developed from self-pollination of F_1 individuals. Backcross mapping population is developed by backcrossing F_1 lines to recurrent parental line. Repetitive backcross for six generations followed by additional two steps of self-pollination leads to development of nearly isogenic lines (NILs). NILs are characterized by significantly decreased level of heterozygosity. In the ideal case, these lines will contain only 1% of genetic background from donor parental line [217]. Recombinant inbred lines (RILs) are developed by multiple self-pollination of

individuals until F_6-F_8 generation is achieved. This method is called single seed descent method and is mainly used for RILs development. Its main advantage is elimination of number of intermediate steps. Another type of mapping population is doubled haploid lines mapping population which is created by doubling gametes of F_1 or F_2 generation. The chromosome doubling of pollen grains is followed by regeneration using tissue culture techniques. This approach enables to achieve null level of heterozygosity rapidly (reviewed in: [216]).

F_2 and BC mapping populations are considered temporary mapping populations (mortal); however, their development is easy and not time-demanding. DH, RILs and NILs are considered immortal (true-breeding) mapping populations because of high level homozygosis which enables to propagate mapping population lines without genetic change (reviewed in: [216]).

The suitable set of markers is necessary for development of precise genetic map. High density of DNA markers is important for quantitative trait loci localisation in the genome. The fine genetic maps further lead to map-based positional cloning which is even further narrowing down the region of interest. The final stage of positional cloning includes delimitation of region below 1 cM and is followed by the physical mapping stage. This consists of chromosome walking process and subsequent identification of sequence between flanking markers [218]. The availability of physical map of the whole genome simplifies the process. Then, the laborious chromosome walking is replaced by finding of BAC (bacterial artificial chromosome) clones with flanking markers and ideally, identification of one physical map region where the region spanning flanking markers is located (Fig. 10) [219]. The precise identification and following validation of gene responsible for studied phenotype is described in subchapter 2.4 Functional genomics.

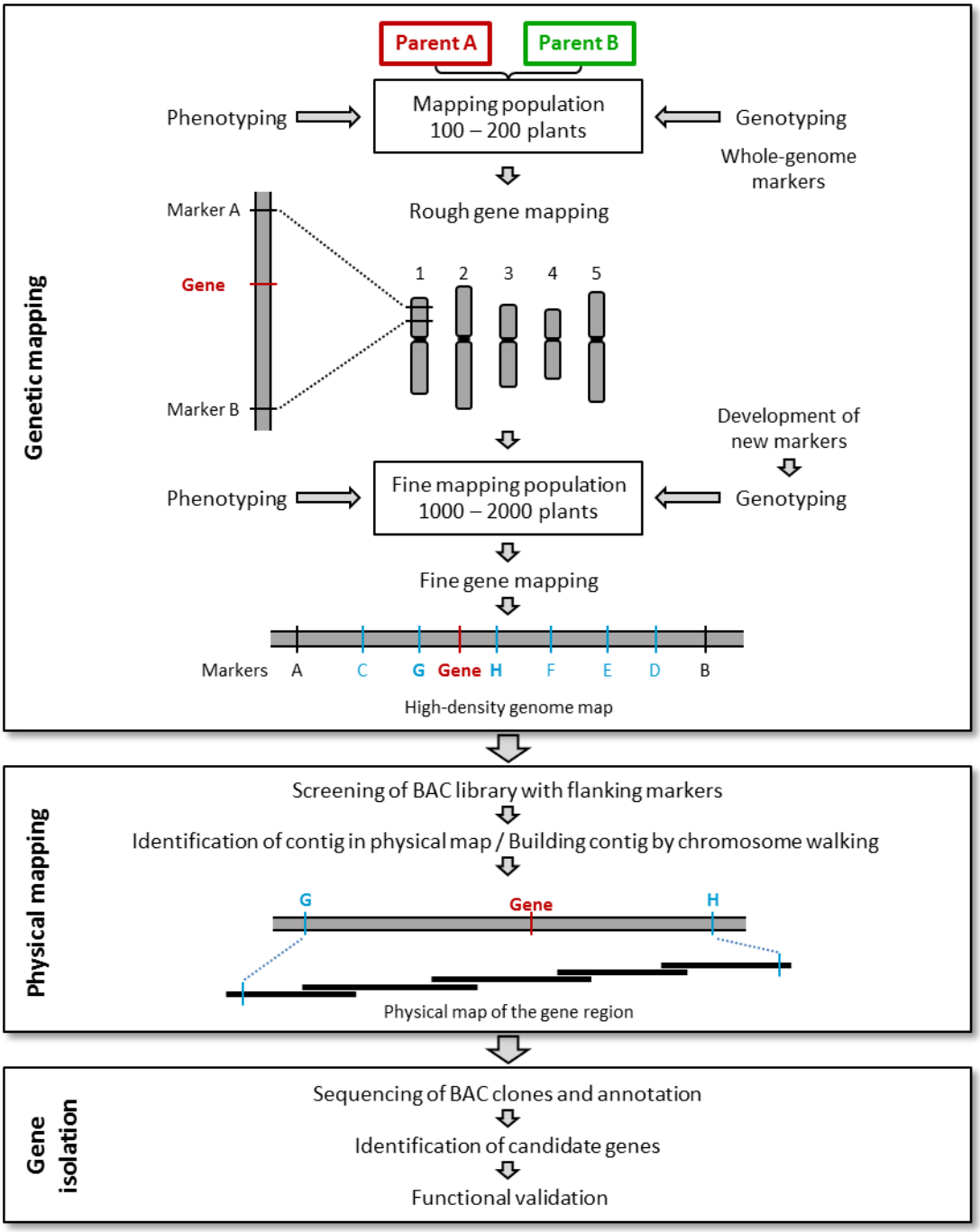


Fig. 10: The scheme of candidate gene identification. Identification and functional validation of gene responsible for studied phenotype must be preceded by genetic mapping and further by positional cloning which leads to developing of genome map with high density of markers within region of interest. The localisation of desired gene within the narrowest region is followed by physical mapping. The gene is localized into contig of BAC clones which is delimited by flanking markers. The sequencing of delimited region leads to identification of candidate genes (reproduced from: [220]).

The markers can be divided into (i) phenotypic, (ii) biochemic and (iii) DNA markers. DNA markers are based on the variation in DNA sequence. Increasing level of genomics leads to evolution of DNA markers from low throughput hybridization

based (restriction fragment length polymorphism - RFLP), trough medium-throughput PCR markers (sequence-characterised amplified region - SCAR, cleaved amplified polymorphic sequence - CAPS, amplified fragment length polymorphism - AFLP and repetitive sequence based markers – simple sequence repeats - SSR, inter-simple sequence repeat - ISSR, inter-retrotransposon amplified polymorphism - IRAP, retrotrasposon-microsatellite amplified polymorphism - REMAP, insertion site binding polymorphism - ISBP) and microarray based markers (diversity arrays technology - DArT) to high-throughput sequencing based markers (reviewed in: [221]). Nowadays, the new high-throughput genotyping platforms are developed and have been progressively replacing older and more laborious forms of genotyping. The whole genome sequencing is used for detection of thousands of SNPs in many accessions. Application of SNP was originally based on high-throughput analysis of Illumina (Inc., San Diego, CA, USA) developed bead-array technology based platforms (GoldenGate and Infinium). The number of possible SNP markers ranges from 90K as recorded by new Infinium to 500K and 4M by Illumina shortgun WGS sequencing. Later, microarray-based platform – Affymetrix (Inc., Santa Clara, CA, USA) offered another way of SNP genotyping. The alternative approach to these platforms are KBiosciences´ competitive allele specific PCR (KASP) and TaqMan technology coupled with Life Technology (reviewed in:[222]).

2.3.1 *Statistical processing of genotyping and phenotyping data*

The genotypes of individuals of mapping population are used for ordering markers and for development of genetic map. The data is processed by linkage mapping software (e.g. JoinMap, MapMaker, Map Manager) [223–226]. The statistical software computing genetic maps needs to deal with high amount of input data provided by high-throughput genotyping platforms. These challenges could be solved by dividing the process into three steps: (i) “grouping”, (ii) “ordering” and (iii) “spacing” [227].

(i) Grouping is responsible for dividing markers into linkage groups. Many algorithms were developed for marker grouping. One of the examples is *nearest neighbor locus*. The markers are step by step added into group according to lowest recombination frequency with the members of the linkage group. This strategy is employed in MAPMAKER. The MAPMAKER calculates two-point maximum likelihood distances and logarithm of odds (LOD) score for every pair of marker.

Further it connects markers with LOD score higher than 3 into the same group. Another type is *graph partitioning grouping method*. It creates graph of all markers connected to all other markers. The connecting graph edges are weighted by two point functions of data. All edges with LOD score above threshold value are chopped and the resultant partitioning lead to creating of linkage groups [227].

(ii) Ordering determines order of markers within each linkage group. The marker ordering is based on minimizing/maximizing value of objective function. Objective function evaluates the quality of a given marker order and describes how one marker order is better than another. The *sum of adjacent recombination fractions* is one of the examples of objective function. The loci are considered nearest when the smallest value of recombination fraction is achieved. Then, the sum of adjacent recombination fraction value is minimized in case of suitable marker order. Another objective functions are *maximum sum of adjacent LOD scores*, *the minimum number of crossovers*, *the product of adjacent recombination fractions*, *the minimum entropy*, *the minimum weighted least squares marker order*, *the maximum likelihood* and *the maximum number of fully informative meiosis* (reviewed in: [227]).
However, ordering of huge number of markFers has high computational demands. This task is considered non-deterministic polynomial (NP) - hard combinatorial problem. Such problem is not possible to solve in the real time. Therefore, the approaches (a *seriation* and a *branch and bound*) which reduce amount of scanned marker orders were developed. The *seriation* approach determines the closely linked marker pair, further it adds random marker in position which leads to optimization of objective function value. The *branch and bound* strategy starts with determination of initial order of markers. In the next step the partial marker order is constructed. Those which are worse than the initial order are eliminated. Further the new order is established and the last step is repeated. This is repeating until objective function is not decreasing anymore [223,228].

(iii) Spacing means determination of marker distances which is based on recombination fraction between markers [227].

2.3.2 *QTL analysis*

QTL analysis is linking phenotyping data and genotyping data in order to identify locus responsible for studied phenotype. It is used for polygenic traits. Analysis can

be performed by three methods: *single factor analysis of variance* (or *single marker analysis*), *simple interval mapping* (SIM) and *composite interval mapping* (CIM).

The *single factor analysis of variance* uses *t*-test, analysis of variance (ANOVA) and linear regression. These methods are calculating probability of linkage phenotype to every single marker. The linkage map is not necessary and this method can be used in cases when the dense genetic map is missing. Linear regression is using the coefficient of determination (R^2) or *P*-value. The marker connected to phenotype has the lowest *P*-value and the highest R^2. R^2 defines the importance of marker in conferring phenotype, *P*-value determines probability of linkage phenotype to marker.

SIM method analyses intervals between all pairs of adjacent markers located on the genetic map for detecting position of QTL. SIM performs likelihood ratio statistic at every position of the interval. The QTL is thus located in the interval defined by flanking markers [229].

The CIM combines interval mapping with multiple regression analysis. Similar to SIM, it analyses position of QTL in markers interval. In addition, it considers effect of background markers – cofactors [230–233]. It reduces bias of QTLs located in other regions, so the their effect can be separated [230,231].

Results of SIM and CIM are shown in LOD or likelihood ratio statistic (LRS) statistic. LOD score can be calculated from LRS by dividing LRS value by 4.61. LOD score is counted as probability of linkage of two loci with given recombination value over probability they are not [225,234]. If LOD value of marker recombination exceeds the critical LOD score, the markers are considered linked. The LOD score 3 is generally considered indicator of marker linkage. It means that there is 1,000 higher probability the markers are linked than that they are not [225]. However, later studies showed that even higher LOD value could lead to false-positive linkage. To identify position of QTL within linkage map, the LOD (LRS) values are compared. These values must exceed the threshold values above which the results are statistically significant. The thresholds are determined by permutation tests. In permutations test the phenotypic values of individual plant are randomly distributed while the genotype is held the same. The QTL analysis is then performed in order to determine amount of false-positive marker QTL associations. This process is then

repeated for 500 – 1,000 times. The highest LOD (LRS) which exceeds threshold value is further linked to analyzed phenotype. The output of these mapping methods is graph with marker position on *x* axis and with LOD (LRS) score on the *y* axis [235].

The identified QTLs could be divided according to their effect on phenotype as suggestive (P=0,05), significant (P=0,01) and highly significant (P=0,001) [236].

When map distances are below 10 cM, the recombination frequency reflects marker distance. However, this relation is not observed when the distance is above 10 cM. Two algorithms are used for recalculation of this distance. The Haldane´s mapping function does not assume interference between crossovers; and the Kosambi´s mapping function suggests the certain level of interference [216].

2.4 Functional genomics

Functional genomics is a section of genomics which attempts to describe gene functions and their mutual interactions. It focuses on explanation of DNA function at the levels of genes, RNA transcripts and protein products. The main goal is to understand the relationship between an organism's genome and its phenotype. It is believed the functional genomics will provide more comprehensive picture of biological processes taking place in the cell, the tissue and whole organisms and how they are arising from the information encoded by DNA in an organism's genome.

The mutagenesis underpin molecular basis of functional genomics. The mutagenesis is induced by chemical, physical and biological compounds (reviewed in: [49]).

The functional analysis use both forward and reverse genetic approaches. Forward genetics carry out gene identification through the characterization of a mutant phenotype. Transposon tagging mutagenesis has been preferentially used in past. Extensive transposon derived resources have been developed for *Arabidopsis*, rice, maize, tomato and barley [237]. Nowadays, huge increase of sequence data produced by genomic projects leads to prevalence of reverse genetic approach in gene function analysis. Reverse genetics screens generally does not require large scale phenotypic screening, but does require knowledge of a candidate gene sequence for recovering and assessment of mutation in gene of interest.

Function of genes can be determined on different levels of their effects like transcription, translation and regulation. There are numerous techniques and approaches which can be used for revealing gene functions. At the beginning, the methods were based on chemical or radiation mutagenesis [238]. Later, the methods of genetic engineering and biotechnology were put in the forefront including microarray, serial analysis of gene expression (SAGE), gene knockout, RNA interference (RNAi), virus gene silencing (VIGS), targeting induced local lesions in genomes (TILLING), targeted mutagenesis (zinc-fingers nucleases, meganucleases, transcription activator-like effector nucleases (TALENs), Clustered Regularly Interspaced Short Palindromic Repeats/Cas 9 (CRISPR/Cas 9) (reviewed in: [239,240].

Because the functional genomics integrates broad portfolio of methods, only the selected techniques will be discussed in detail.

2.4.1 *RNA interference*

RNA interference is method which disrupts gene function and respective phenotype. The RNA interference is based on development of double stranded RNA which is included in cleavage of mRNA or in blocking of mRNA at the transcriptional level. The stages of RNA interference includes cleavage of double stranded RNA into 21-26 bp small RNA (smRNA) molecules by DICER-like enzyme. Further, the smRNA molecules are recruited into RNA-induced silencing complex (RISC). The RISC complex mediates cleavage of target complementary mRNA. Since RNA interference is able to block expression of all homoeologous copies of genes, it is very useful in silencing of genes in wheat. The usage of RNA interference in functional genomics studies is conditioned by ability to transform plant with construct used for dsRNA transcription (reviewed in: [241]).

Posttranscriptional regulation in wheat was used in identification of effect of many genes. Yan et al. [79] used 347 bp targeted sequence from *T. monococcum* for silencing of *VRN2* gene. They were able to decrease mRNA level down to 40% and to correlate it to flowering time acceleration. Loukoianov et al. [99] used similar approach (294 bp dsRNA designed from *T. monococcum*) for identification of *VRN1* effect on flowering time [99].

2.4.2 *Targeted genome editing*

The targeted genome editing approach is based on modification of specific DNA sequence without damaging of other parts of DNA. Advent of these method eliminated problems with high copy number of inserted sequences and the regulation of expression. These techniques are using ability of enzymes to perform double strand breaks in specific DNA sequence - nucleases and ability of genome to repair double strand breaks by non-homologous ends joining (NHEJ) or by homologous recombination (HR). The HR can introduce desired mutations by using specific DNA (repair DNA) as a template. However plants prefer NHEJ as a dominant repairing

mechanism. NHEJ can introduce unpredicted errors – especially small deletions and substitution which can lead to knockout of the desired gene (reviewed in: [242]). Increased double strand breaks showed to increase frequency of HR [243–245]. This can be used in order to achieve specific mutations in required DNA sequence positions.

The targeted genome editing is connected with zinc-fingers nucleases, meganucleases and TALENs [238]. Recently, these tools are overshadowed by CRISPR/Cas9 method.

2.4.2.1 Zinc-fingers nucleases

Zinc fingers nucleases are consisting of three or four zinc fingers and Fok I endonuclease. Every single finger recognizes three nucleotides on target DNA. They works in dimers and thus they recognize 18-24 bp. Fok I endonuclease is a type II restriction enzyme. It works as a dimmer and provides DNA sequence cleavage. The binding of dimeric Zinc-fingers nucleases to specific sequence brings Fok I monomers to close proximity and activates its restrictive activity (Fig. 11) (reviewed in: [238]).

2.4.2.2 Meganucleases

Meganucleases (or homing meganucleases) are derived from proteins coded by mobile introns [246,247]. The meganucleases have DNA binding domain which is able to recognize DNA sequence larger than 12 bp and cleavage domain with endonuclease activity. Coding sequence of both domains is overlapping which complicates modification of its specificity to different DNA targets (Fig. 11) (reviewed in: [238]). Dimeric formation of meganucleases is needed for their appropriate function, however monomeric protein can be linked and can work as a single peptide [248].

2.4.2.3 TALENs

The principle of targeted genome editing using TALENs is similar. TALENs recognize specific DNA sequence and create double strand break (Fig. 11). TALENs provide many advantages in comparison to previous methods. They are easy to assembly, have reliable function and broad targeting range.

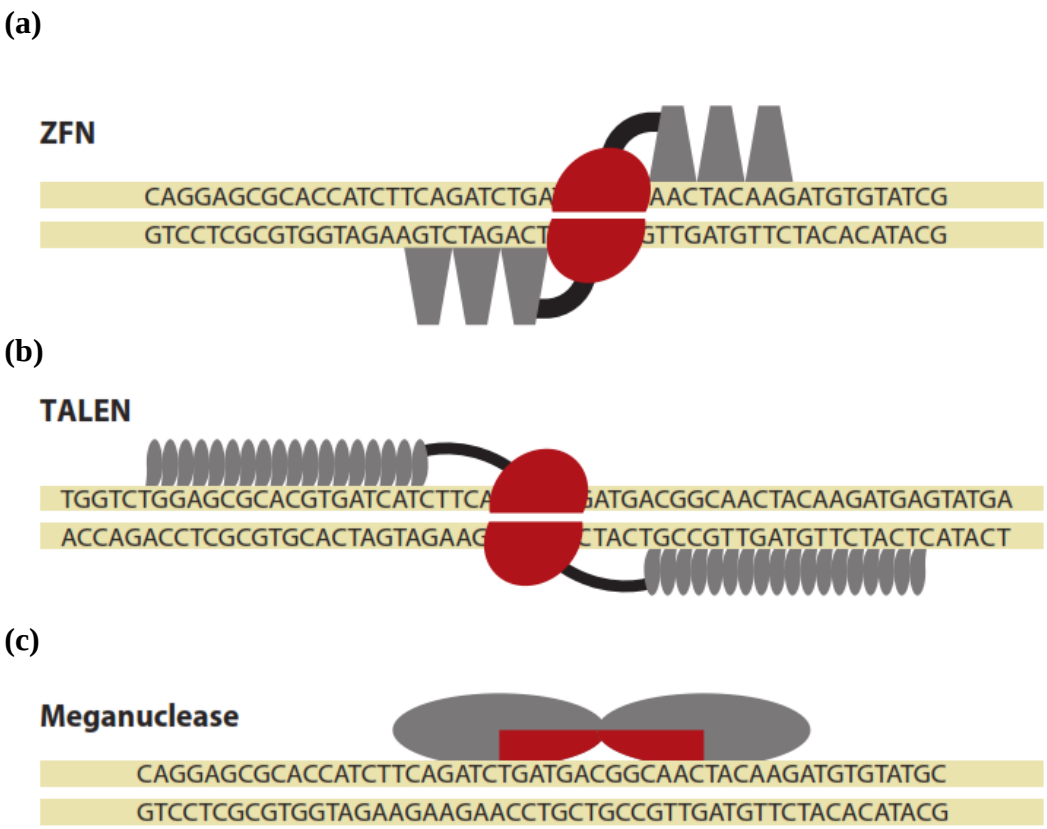


Fig. 11: The three classes of sequence-specific nucleases. (a) Two zinc-finger nucleases – ZFNs - bind to their target site. Each monomer is made of of three zinc fingers (gray) that make base-specific contacts. The zinc finger arrays are fused to the catalytic domain of the FokI endonuclease (red), which upon dimerization creates a double-strand break. (b) Two transcription activator-like effector nucleases (TALENs) bound to their target site. Variable di-residues in each repeat in the transcription activator-like effector array (gray) make a base-specific contact with the target DNA. As with zinc-finger nucleases, the transcription activator-like effector repeat arrays are fused to FokI. (c) A meganuclease bound to a DNA target. The catalytic domain, which also determines base specificity, is shown in red (reproduced from: [238]).

The DNA binding domain of these transcription-activator-like proteins comes from bacterial genus *Xantomonas*. Originally, they are used for binding promoter sequence of specific genes after plant cell infection. This way, they modulates its expression and increases virulence [249]. The DNA binding domain contains 13 - 28 copies of 34 amino acid repeats. Each repeat recognizes single base. Every single repeat creates hairpin structure with 12^{th} and 13^{th} amino acid residues on the top. Besides 12^{th} and 13^{th} residues, the other amino acids are highly conserved [250,251]. The 12^{th} amino acid is responsible for hairpin stabilization and 13^{th} makes a base specific contact with major groove of DNA. Repeats recognize different bases and can be

engineered in order to recognize target sequence. The DNA binding domain is bound to Fok I restriction endonuclease which provides sequence cleavage [252].

2.4.2.4 CRISPR/Cas9

Recently, the CRISPR/Cas9 method found to be useful in targeted method editing. CRISPR system provides bacteria and archaea with adaptive immunity against viruses and plasmids. There are three types of CRISPR (reviewed in: [253]). CRISPR system type II is used in gene editing and was adopted from *Streptococcus pyogenes*. In this system, Cas9 protein is responsible for endonuclease activity. The Cas9 structure is made from two lobes (globular recognition and small nuclease lobe). The lobes are composed of two domains which creates grooves for binding of nucleic acid. 20 - nucleotide-long small guide RNA (sgRNA) is an essential part of this system and it is complementary to targeted sequence. Binding of sgRNA to Cas9 leads to conformational change and thus to positioning of RNA-DNA duplex (Fig. 12). The 3 bp conserved motive (NGG), so called photospacer adjacent motif, at the 3′end is crucial for recognizing by Cas9 protein. Simple modification of sgRNA sequence permits recognition of specific target DNA and thus enables targeted editing [253].

The CRISPR gene editing was used in plants for the first time in 2013 [254–256]. In wheat, the modification of *Mlo* genes (conferring powdery mildew resistance) at the single homoeologous loci was achieved using this method [257].

Editing of genome must be preceded by introducing of Cas9 system coding construct to the plant. This can be achieved by *Agrobacterium tumefaciens* transformation. For developing of modified plants without any integrated vector DNA, several generations of crossing are needed. This time-consuming crossing can be avoided by transfection by non-integrating plasmids. Nevertheless, nuclease cleavage of these plasmids can lead to integration of short sequences into genome. Recently, the direct incorporation of Cas9-guide RNA construct into cell was achieved. Woo et al. [258] used polyethylene glycol for induction of endocythosis directed transport of Cas9-guide RNA complex and attained 45% efficiency of transformation. However, some studies detected off target activity of Cas9-guide RNA complex [259–262].

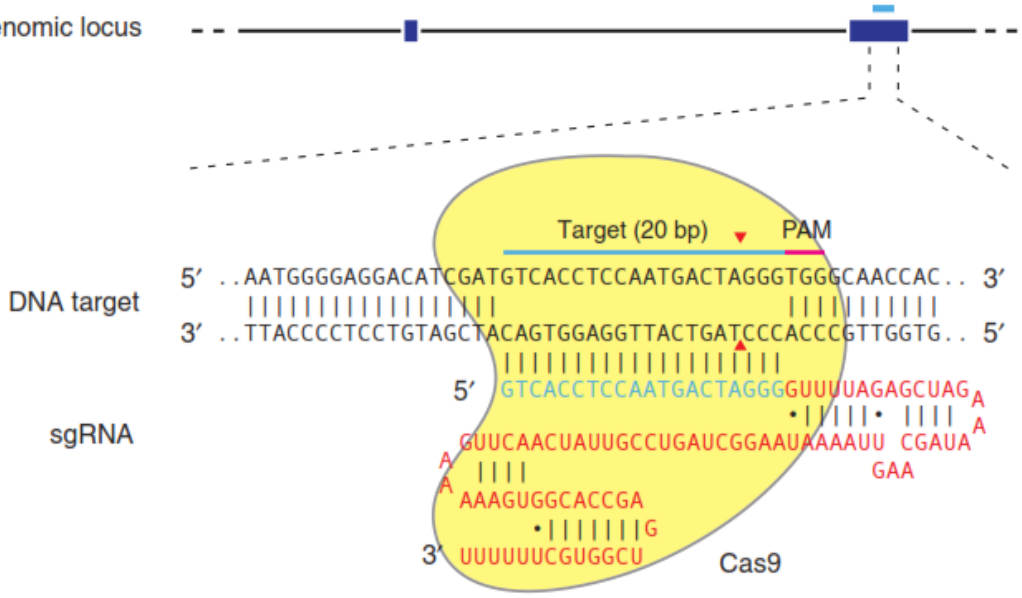


Fig. 12: Scheme of the RNA-guided Cas9 nuclease. The Cas9 nuclease from *Streptococcus pyogenes* (in yellow) is targeted to genomic DNA by sgRNA consisting of a 20 nucleotides guide sequence (blue) and a scaffold (red). The guide sequence pairs with the DNA target (blue bar on top strand), directly upstream of a requisite 5'-NGG adjacent motif (PAM; pink). Cas9 mediates a double strand break ~3 bp upstream of the PAM (red triangle) (reproduced from: [263]).

Despite the progress in transformation and genome editing techniques [264], the approach which facilitates simultaneous and specific modification of large number of genes is still missing. The genome editing is further influenced by unpredictable methylation statuses. The solution of this task can be development of engineered mini-chromosomes [265].

2.4.3 *Transformation of cereals*

Validation of gene function in specific stage/tissue is performed by introduction of exogenous DNA transfer to organism by process called transformation. There are usually used two main approaches: transformation by *A. tumefaciens* or by biolistic methods.

The transfection by *Agrobacterium* is using ability of Ti plasmid to incorporate into host DNA. Ti plasmid is used as a vector for incorporation of desired DNA fragment into the host plant. The prime experiments with wheat transformation had a very low efficiency (< 5%) and were cultivar-dependent [266]. This increased the cost of this method and made it unavailable for extensive usage. However, Ishida et al. [267] achieved efficiency up to 50% which can lead to increasing usage of this method in practice.

The biolistic approach is using a gene gun for incorporating DNA molecules bound to wolfram or gold nanoparticles to plant tissue [268].

When comparing both methods, the advantage of biolistic is based on the lack of optimizing needs to different wheat cultivars. The *Agrobacterium* transformation protocol needs to be optimized for every single cultivar. However, *Agrobacterium* mediated transgenic plants have a low number of inserted copies. In addition, the efficiency and stability of insert expression was higher than in the plants developed by biolistic [269].

High-quality transformations events are considered those with single copy insertion in the recipient genome and with stable expression. The expression of gene is influenced by many circumstances. Important is the choice of suitable promoter, e.g. the promoter from maize ubiquitin gene was successfully used in barley and wheat [270]. Transgenic expression can also be influenced by inclusion of untranslated regulatory domains and signal peptides. The addition of intron regulatory sequences can also increase transgenic expression [271]. Main disadvantage of transformation is inability to determine precise site of transformation and to regulate number of insertion events. However, the insertion sites are often located non-randomly in gene rich sites and their physical location can be determined by FISH [272].

There are different stages of transformation which can be influenced for acquiring higher transformation efficiency:
(i) higher number of successful transformations can be achieved by increasing susceptibility of target tissue. Vacuum infiltration and sonication were used in barley [273]. The immature embryos centrifugation and heat treatment increased efficiency in maize and rice [266]. High efficiency of transformed wheat was gained by using immature embryos as explants for transformation. (ii) Modification of gene expression can also influence transformation efficiency, e.g. overexpression of histone genes in *Arabidopsis* leads to higher efficiency [274]. (iii) Higher efficiency can be achieved also by influencing regeneration rate of transformed plant by modification of medium composition. The number of regenerating plants is increasing with addition of thidiazuron, picloram or lipoitic acid [275–277]. Higher amount of copper in the medium can double regeneration rate [278]. However, the

effect of these changes is cultivar dependent, it can lead to higher regeneration in one but not in another wheat variety [279,280].

Until now, the transformation of wheat for flowering time modification was not achieved. However, two barley cultivars were transformed by *Arabidopsis CRY2* gene. These transgenic plants flowered 25 days earlier than original cultivar [281].

2.5 References

[1] Andrés F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 2012;13:627–39. doi:10.1038/nrg3291.

[2] Srikanth A, Schmid M. Regulation of flowering time: All roads lead to Rome. Cell Mol Life Sci 2011;68:2013–37. doi:10.1007/s00018-011-0673-y.

[3] Moon J, Suh SS, Lee H, Choi KR, Hong CB, Paek NC, et al. The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. Plant J 2003;35:613–23. doi:10.1046/j.1365-313X.2003.01833.x.

[4] King RW, Evans LT. GIBBERELLINS AND FLOWERING OF GRASSES AND CEREALS : Prizing Open the Lid of the "Florigen" Black Box. Annu Rev Plant Biol 2003;54:307–28. doi:10.1146/annurev.arplant.54.031902.135029.

[5] Eriksson S, Böhlenius H, Moritz T, Nilsson O. GA4 Is the Active Gibberellin in the Regulation of LEAFY Transcription and Arabidopsis Floral Initiation. Plant Cell 2006;18:2172–81. doi:10.1105/tpc.106.042317.

[6] Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, et al. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 2005;437:693–8. doi:10.1038/nature04028.

[7] Sasaki Itoh, H. A, Gomi K, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Jeong D-H, et al. Accumulation of Phosphorylated Respressor for Gibberellin Signaling in an F- box Mutant. Science 2003;299:1896–9. doi:10.1126/science.1081077.

[8] McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun T, et al. The Arabidopsis SLEEPY1 Gene Encodes a Putative F-Box Subunit of an SCF E3 Ubiquitin Ligase. Plant Cell 2003;15:1120–30. doi:10.1105/tpc.010827.

[9] Dill A, Thomas SG, Hu J, Steber CM, Sun T-P. The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. Plant Cell 2004;16:1392–405. doi:10.1105/tpc.020958.

[10] Lucas M, Davière J-M, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, et al. A molecular framework for light and gibberellin control of

cell elongation. Nature 2008;451:480–4. doi:10.1038/nature06520.

[11] Feng S, Martinez C, Gusmaroli G, Wang Y, Zhou J, Wang F, et al. Coordinated regulation of Arabidopsis thaliana development by light and gibberellins. Nature 2008;451:475–9. doi:10.1038/nature06448.

[12] Xu F, Li T, Xu PB, Li L, Du SS, Lian HL, et al. DELLA proteins physically interact with CONSTANS to regulate flowering under long days in Arabidopsis. FEBS Lett 2016;590:541–9. doi:10.1002/1873-3468.12076.

[13] Corbesier L, Lejeune P, Bernier G. The role of carbohydrates in the induction of flowering in Arabidopsis thaliana: Comparison between the wild type and a starchless mutant. Planta 1998;206:131–7. doi:10.1007/s004250050383.

[14] Ohto M, Onai K, Furukawa Y, Aoki E, Araki T, Nakamura K. Effects of sugar on vegetative development and floral transition in Arabidopsis. Plant Physiol 2001;127:252–61. doi:10.1104/pp.127.1.252.

[15] Borner R, Kampmann G, Chandler J, Gleißner R, Wisman E, Apel K, et al. A MADS domain gene involved in the transition to flowering in Arabidopsis. Plant J 2000;24:591–9. doi:10.1046/j.1365-313X.2000.00906.x.

[16] Vekemans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruelens P, et al. Gamma paleohexaploidy in the stem lineage of core eudicots: Significance for MADS-BOX gene and species diversification. Mol Biol Evol 2012;29:3793–806. doi:10.1093/molbev/mss183.

[17] Wu G, Poething RS. Temporal Regulation of Shoot Development in Arabidopsis thaliana by Mir156 and its target SPL3. Development 2006;133:3539–47. doi:10.1097/MPG.0b013e3181a15ae8.Screening.

[18] Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. Specific effects of microRNAs on the plant transcriptome. Dev Cell 2005;8:517–27. doi:10.1016/j.devcel.2005.01.018.

[19] Wingler A, Delatte TL, O'Hara LE, Primavesi LF, Jhurreea D, Paul MJ, et al. Trehalose 6-phosphate is required for the onset of leaf senescence associated with high carbon availability. Plant Physiol 2012;158:1241–51. doi:10.1104/pp.111.191908.

[20] Matsoukas IG, Massiah AJ, Thomas B. Florigenic and antflorigenic signaling

in plants. Plant Cell Physiol 2012;53:1827–42. doi:10.1093/pcp/pcs130.

[21] Rolland F, Baena-Gonzalez E, Sheen J. SUGAR SENSING AND SIGNALING IN PLANTS: Conserved and Novel Mechanisms. Annu Rev Plant Biol 2006;57:675–709. doi:10.1146/annurev.arplant.57.032905.105441.

[22] Paparelli E, Parlanti S, Gonzali S, Novi G, Mariotti L, Ceccarelli N, et al. Nighttime Sugar Starvation Orchestrates Gibberellin Biosynthesis and Plant Growth in Arabidopsis. Plant Cell 2013;25:1–11. doi:10.1105/tpc.113.115519.

[23] Heyer AG, Raap M, Schroeer B, Marty B, Willmitzer L. Cell wall invertase expression at the apical meristem alters floral, architectural, and reproductive traits in Arabidopsis thaliana. Plant J 2004;39:161–9. doi:10.1111/j.1365-313X.2004.02124.x.

[24] Bolouri Moghaddam MR, den Ende W Van. Sugars, the clock and transition to flowering. Front Plant Sci 2013;4:22. doi:10.3389/fpls.2013.00022.

[25] Michaels SD. Flowering time regulation produces much fruit. Curr Opin Plant Biol 2009;12:75–80. doi:10.1016/j.pbi.2008.09.005.Flowering.

[26] Hepworth J, Dean C. Flowering Locus C's Lessons: Conserved Chromatin Switches Underpinning Developmental Timing and Adaptation. Plant Physiol 2015;168:1237–45. doi:10.1104/pp.15.00496.

[27] Hornyik C, Terzi LC, Simpson GG. The Spen Family Protein FPA Controls Alternative Cleavage and Polyadenylation of RNA. Dev Cell 2010;18:203–13. doi:10.1016/j.devcel.2009.12.009.

[28] Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ. Control of fowering time by FLC orthologues in Brassica napus. Plant J 2001;28:545–53.

[29] Johanson U, West J, Lister C, Michaels S, Amasino R. Molecular Analysis of FRIGIDA , a Major Determinant of Natural Variation in Arabidopsis Flowering Time. Science 2000;290:344–8. doi:10.1126/science.290.5490.344.

[30] Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 1999;11:949–56. doi:10.1105/tpc.11.5.949.

[31] Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. Targeted 3' Processing

of Antisense Transcripts Triggers Arabidopsis FLC Chromatin Silencing. Science 2010;327:94–7. doi:10.1126/science.1180278.

[32] Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 2009;462:799–802. doi:10.1038/nature08618.

[33] Heo JB, Sung S. Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science 2011;331:76–9. doi:10.1126/science.1197349.

[34] Csorba T, Questa JI, Sun Q, Dean C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. Proc Natl Acad Sci U S A 2014;111:16160–5. doi:10.1073/pnas.1419030111.

[35] Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science 2013;340:619–21. doi:10.1126/science.1234848.

[36] Heo JB, Sung S. Encoding memory of winter by noncoding RNAs. Epigenetics 2011;6:544–7. doi:10.4161/epi.6.5.15235.

[37] Yamaguchi A, Abe M. Regulation of reproductive development by non-coding RNA in Arabidopsis: To flower or not to flower. J Plant Res 2012;125:693–704. doi:10.1007/s10265-012-0513-7.

[38] Sung S, Amasino RM. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. Nature 2004;427:159–64. doi:10.1038/nature02238.1.

[39] De Lucia F, Crevillen P, Jones AME, Greb T, Dean C. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc Natl Acad Sci U S A 2008;105:16831–6. doi:10.1073/pnas.0808687105.

[40] Greb T, Mylne JS, Crevillen P, Geraldo N, An H, Gendall AR, et al. The PHD Finger Protein VRN5 Functions in the Epigenetic Silencing of Arabidopsis FLC. Curr Biol 2007;17:73–8. doi:10.1016/j.cub.2006.11.052.

[41] Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G. A Polycomb-group gene regulates homeotic gene expression in

Arabidopsis. Nature 1997;386:44–51. doi:10.1038/386044a0.

[42] Gendall AR, Levy YY, Wilson A, Dean C. The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. Cell 2001;107:525–35. doi:10.1016/S0092-8674(01)00573-6.

[43] Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. Multiple Roles of Arabidopsis VRN1 in Vernalization and Flowering Time Control. Science 2002;297:243–6. doi:10.1126/science.1072147.

[44] Sung S, He Y, Eshoo TW, Tamada Y, Johnson L, Nakahigashi K, et al. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. Nat Genet 2006;38:706–10. doi:10.1038/ng1795.

[45] Mylne J, Greb T, Lister C, Dean C. Epigenetic regulation in the control of flowering. Cold Spring Harb Symp Quant Biol 2004;69:457–64. doi:10.1101/sqb.2004.69.457.

[46] Berry S, Dean C. Environmental perception and epigenetic memory: mechanistic insight through FLC. Plant J 2015;83:133–48. doi:10.1111/tpj.12869.

[47] Kumar SV, Wigge PA. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. Cell 2010;140:136–47. doi:10.1016/j.cell.2009.11.006.

[48] Lutz U, Posé D, Pfeifer M, Gundlach H, Hagmann J, Wang C, et al. Modulation of Ambient Temperature-Dependent Flowering in Arabidopsis thaliana by Natural Variation of FLOWERING LOCUS M. PLoS Genet 2015;11. doi:10.1371/journal.pgen.1005588.

[49] Milec Z, Valárik M, Bartoš J, Šafář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv 2014;32:200–14. doi:10.1016/j.biotechadv.2013.09.008.

[50] Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature 2001;410:1116–20. doi:10.1038/35074138.

[51] Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Rühl M, Jarillo J a., et al.

Arabidopsis DOF Transcription Factors Act Redundantly to Reduce CONSTANS Expression and Are Essential for a Photoperiodic Flowering Response. Dev Cell 2009;17:75–86. doi:10.1016/j.devcel.2009.06.015.

[52] Imaizumi T, Schultz F, Harmon FG, A HL, Kay SA. FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in Arabidopsis. Science 2005;309:293–7. doi:10.1126/science.1110586.

[53] Sawa M, Nusinow D a, Kay S a, Imaizumi T. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science 2007;318:261–5. doi:10.1126/science.1146994.

[54] Liu L-J, Zhang Y-C, Li Q-H, Sang Y, Mao J, Lian H-L, et al. COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. Plant Cell 2008;20:292–306. doi:10.1105/tpc.107.057281.

[55] Laubinger S, Marchal V, Le Gourrierec J, Wenkel S, Adrian J, Jang S, et al. Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Development 2006;133:3213–22. doi:10.1242/dev.02481.

[56] Hoecker U, Quail PH. The Phytochrome A-specific Signaling Intermediate SPA1 Interacts Directly with COP1, a Constitutive Repressor of Light Signaling in Arabidopsis. J Biol Chem 2001;276:38173–8. doi:10.1074/jbc.M103140200.

[57] Morris K, Thornber S, Codrai L, Richardson C, Craig A, Sadanandom A, et al. DAY NEUTRAL FLOWERING represses CONSTANS to prevent Arabidopsis flowering early in short days. Plant Cell 2010;22:1118–28. doi:10.1105/tpc.109.066605.

[58] Yoshida N, Yanai Y, Chen L, Kato Y, Hiratsuka J, Miwa T, et al. EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. Plant Cell 2001;13:2471–81. doi:10.1105/tpc.010227.

[59] Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, et al. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J 2006;25:4638–49.

doi:10.1038/sj.emboj.7601311.

[60] Ko C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W. Arabidopsis MSI1 is a component of the MEA / FIE Polycomb group complex and required for seed development. EMBO J 2003;22.

[61] Takada S, Goto K. TERMINAL FLOWER2 , an Arabidopsis Homolog of HETEROCHROMATIN PROTEIN1 , Counteracts the Activation of FLOWERING LOCUS T by CONSTANS in the Vascular. Plant Cell 2003;15:2856–65. doi:10.1105/tpc.016345.flowering.

[62] Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, et al. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science 2000;290:2110–3. doi:10.1126/science.290.5499.2110.

[63] Wickland DP, Hanzawa Y. The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. Mol Plant 2015;8:983–97. doi:10.1016/j.molp.2015.01.007.

[64] Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, et al. Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. Science 2005;309:1056–9. doi:10.1126/science.1114358.

[65] Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, et al. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. Science 2005;309:1052–6. doi:10.1126/science.1115983.

[66] Ó'Maoiléidigh DS, Graciet E, Wellmer F. Tansley review: Gene networks controlling Arabidopsis thaliana flower development. New Phytol 2014;201:16–30. doi:10.1111/nph.12444.

[67] Yu J, Hu S, Wang J, Dai L, Zhou Y, Zhang X, et al. A Draft Sequence of the Rice Genome (Oryza sativa L . ssp .). Science 2002;296:79–82. doi:10.1126/science.1068037.

[68] Earle ED. Nuclear DNA Content of Some Important Plant Species 2016. doi:10.1007/BF02672069.

[69] Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, et al. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol

2002;43:1096–105. doi:10.1093/pcp/pcf156.

[70] Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a protein is a mobile flowering signal in rice. Science 2007;316:1033–6. doi:10.1126/science.1141753.

[71] Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 2003;422:719–22. doi:10.1038/nature01549.

[72] Plant S, Yano M, Kojima S, Takahashi Y, Lin H, Sasaki T. Genetic Control of Flowering Time in Rice , a NATURAL VARIATIONS : A NEW RESOURCE FOR. Society 2001;127:1425–9. doi:10.1104/pp.010710.1.

[73] Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 2000;12:2473–84. doi:10.1105/tpc.12.12.2473.

[74] Izawa T, Oikawa T, Sugiyama N, Tanisaka T, Yano M, Shimamoto K. Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. Genes Dev 2002;16:2006–20. doi:10.1101/gad.999202.

[75] Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, et al. Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes Dev 2004;18:926–36. doi:10.1101/gad.1189604.

[76] Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, et al. Arabidopsis transcription factor genome wide comparative analysis among eukaryotes. Science 2000;290:2105–9. doi:10.1126/science.290.5499.2105.

[77] Kim SL, Lee S, Kim HJ, Nam HG, An G. OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1, OsMADS14, and Hd3a. Plant Physiol 2007;145:1484–94. doi:10.1104/pp.107.103291.

[78] Itoh H, Nonoue Y, Yano M, Izawa T. A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nat Genet 2010;42:635–8. doi:10.1038/ng.606.

[79] Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 2004;303:1640–4. doi:10.1126/science.1094305.

[80] Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet 2008;40:761–7. doi:10.1038/ng.143.

[81] Tsuji H, Taoka KI, Shimamoto K. Regulation of flowering in rice: Two florigen genes, a complex gene network, and natural variation. Curr Opin Plant Biol 2011;14:45–52. doi:10.1016/j.pbi.2010.08.016.

[82] Taoka K, Ohki I, Tsuji H, Furuita K, Hayashi K, Yanase T, et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature 2011;476:332–5. doi:10.1038/nature10272.

[83] Purwestri YA, Ogaki Y, Tamaki S, Tsuji H, Shimamoto K. The 14-3-3 protein GF14c acts as a negative regulator of flowering in rice by interacting with the florigen Hd3a. Plant Cell Physiol 2009;50:429–38. doi:10.1093/pcp/pcp012.

[84] Komiya R, Yokoi S, Shimamoto K. A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. Development 2009;136:3443–50. doi:10.1242/dev.040170.

[85] Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K. Hd3a and RFT1 are essential for flowering in rice. Development 2008;135:767–74. doi:10.1242/dev.008631.

[86] Milec Z, Valárik M, Bartoš J, Šafář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv 2014;32:200–14. doi:10.1016/j.biotechadv.2013.09.008.

[87] Stelmakh AF. Genetic systems regulating flowering response in wheat. Euphytica 1998;100:359–69. doi:10.1023/a:1018374116006.

[88] Kamran A, Iqbal M, Spaner D. Flowering time in wheat (Triticum aestivum L.): A key factor for global adaptability. Euphytica 2014;197:1–26. doi:10.1007/s10681-014-1075-7.

[89] Yan L, Loukoianov a, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci

U S A 2003;100:6263–8. doi:10.1073/pnas.0937399100.

[90] Yan L, Fu D, Li C, Blechl a, Tranquilli G, Bonafede M, et al. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci U S A 2006;103:19581–6. doi:10.1073/pnas.0607142103.

[91] Kippes N, Zhu J, Chen A, Vanzetti L, Lukaszewski A, Nishida H, et al. Fine mapping and epistatic interactions of the vernalization gene VRN-D4 in hexaploid wheat. Mol Genet Genomics 2014;289:47–62. doi:10.1007/s00438-013-0788-y.

[92] Yoshida T, Nishida H, Zhu J, Nitcher R, Distelfeld A, Akashi Y, et al. Vrn-D4 is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. Theor Appl Genet 2010;120:543–52. doi:10.1007/s00122-009-1174-3.

[93] Kippes N, Debernardi JM, Vasquez-Gross HA, Akpinar BA, Budak H, Kato K, et al. Identification of the VERNALIZATION 4 gene reveals the origin of spring growth habit in ancient wheats from South Asia. Proc Natl Acad Sci U S A 2015;112:E5401–10. doi:10.1073/pnas.1514883112.

[94] Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J. Allelic variation at the VRN-1 promoter region in polyploid wheat. Theor Appl Genet 2004;109:1677–86. doi:10.1007/s00122-004-1796-4.

[95] Danyluk J, Kane N a, Breton G, Limin AE, Fowler DB, Sarhan F. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol 2003;132:1849–60. doi:10.1104/pp.103.023523.

[96] Murai K, Miyamae M, Kato H, Takumi S, Ogihara Y. WAP1, a Wheat APETALA1 Homolog, Plays a Central Role in the Phase Transition from Vegetative to Reproductive Growth. Plant Cell Physiol 2003;44:1255–65. doi:10.1093/pcp/pcg171.

[97] Li C, Distelfeld A, Comis A, Dubcovsky J. Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. Plant J 2011;67:763–73. doi:10.1111/j.1365-313X.2011.04630.x.

[98] Chen A, Dubcovsky J. Wheat TILLING Mutants Show That the Vernalization

Gene VRN1 Down-Regulates the Flowering Repressor VRN2 in Leaves but Is Not Essential for Flowering. PLoS Genet 2012;8. doi:10.1371/journal.pgen.1003134.

[99] Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J. Regulation of VRN-1 vernalization genes in normal and transgenic polyploid wheat. Plant Physiol 2005;138:2364–73. doi:10.1104/pp.105.064287.

[100] Shitsukawa N, Ikari C, Shimada S, Kitagawa S, Sakamoto K, Saito H, et al. The einkorn wheat (Triticum monococcum) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene. Genes Genet Syst 2007;82:167–70. doi:10.1266/ggs.82.167.

[101] Distelfeld A, Li C, Dubcovsky J. Regulation of flowering in temperate cereals. Curr Opin Plant Biol 2009;12:178–84. doi:10.1016/j.pbi.2008.12.010.

[102] Distelfeld A, Tranquilli G, Li C, Yan L, Dubcovsky J. Genetic and molecular characterization of the VRN2 loci in tetraploid wheat. Plant Physiol 2009;149:245–57. doi:10.1104/pp.108.129353.

[103] Pugsley AT. Additional genes inhibiting winter habit in wheat. Euphytica 1972;21:547–52. doi:10.1007/BF00039355.

[104] Dubcovsky J, Lijavetzky D, Appendino L, Tranquilli G. Comparative RFLP mapping of Triticum monococcum genes controlling vernalization requirement. Theor Appl Genet 1998;97:968–75. doi:10.1007/s001220050978.

[105] Fu D, Szucs P, Yan L, Helguera M, Skinner JS, Von Zitzewitz J, et al. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Mol Genet Genomics 2005;273:54–65. doi:10.1007/s00438-004-1095-4.

[106] Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 2012;7. doi:10.1371/journal.pone.0033234.

[107] Yu M, Carver BF, Yan L. TamiR1123 originated from a family of miniature inverted-repeat transposable elements (MITE) including one inserted in the

Vrn-A1a promoter in wheat. Plant Sci 2014;215-216:117–23. doi:10.1016/j.plantsci.2013.11.007.

[108] Muterko A, Kalendar R, Salina E. Novel alleles of the VERNALIZATION1 genes in wheat are associated with modulation of DNA curvature and flexibility in the promoter region. BMC Plant Biol 2016;16:65–81. doi:10.1186/s12870-015-0691-2.

[109] Pidal B, Yan L, Fu D, Zhang F, Tranquilli G, Dubcovsky J. The CArG-box located upstream from the transcriptional start of wheat vernalization gene VRN1 is not necessary for the vernalization response. J Hered 2009;100:355–64. doi:10.1093/jhered/esp002.

[110] Golovnina K a, Kondratenko EY, Blinov AG, Goncharov NP. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. BMC Plant Biol 2010;10:168. doi:10.1186/1471-2229-10-168.

[111] Ivaničová Z, Jakobson I, Reis D, Šafář J, Milec Z, Abrouk M, et al. Characterization of new allele influencing flowering time in bread wheat introgressed from Triticum militinae. N Biotechnol 2016;33. doi:10.1016/j.nbt.2016.01.008.

[112] Shcherban AB, Schichkina AA, Salina EA. The occurrence of spring forms in tetraploid Timopheevi wheat is associated with variation in the first intron of the. BMC Plant Biol 2016;16. doi:10.1186/s12870-016-0925-y.

[113] Santra DK, Santra M, Allan RE, Campbell KG, Kidwell KK. Genetic and molecular characterization of vernalization genes Vrn-A1 , Vrn-B1, and Vrn-D1 in spring wheat germplasm from the pacific northwest region of the U.S.A. Plant Breed 2009;128:576–84. doi:10.1111/j.1439-0523.2009.01681.x.

[114] Chu C-G, Tan CT, Yu G-T, Zhong S, Xu SS, Yan L, et al. A Novel Retrotransposon Inserted in the Dominant Vrn-B1 Allele Confers Spring Growth Habit in Tetraploid Wheat (Triticum turgidum L.). Genes, Genomes, Genet 2011;1:637–45. doi:10.1534/g3.111.001131.

[115] Shcherban a. B, Efremova TT, Salina E a. Identification of a new Vrn-B1 allele using two near-isogenic wheat lines with difference in heading time. Mol Breed 2012;29:675–85. doi:10.1007/s11032-011-9581-y.

[116] Milec Z, Tomková L, Sumíková T, Pánková K. A new multiplex PCR test for the determination of Vrn-B1 alleles in bread wheat (Triticum aestivum L.). Mol Breed 2012;30:317–23. doi:10.1007/s11032-011-9621-7.

[117] Zhang J, Wang Y, Wu S, Yang J, Liu H, Zhou Y. A single nucleotide polymorphism at the Vrn-D1 promoter region in common wheat is associated with vernalization response. Theor Appl Genet 2012;125:1697–704. doi:10.1007/s00122-012-1946-z.

[118] Goncharov NP. Response to vernalization in wheat: its quantitative or qualitative nature. Cereal Res Commun 2004;32:323–30.

[119] Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. MADS box genes control vernalization-induced flowering in cereals. Proc Natl Acad Sci U S A 2003;100:13099–104. doi:10.1073/pnas.1635053100.

[120] Manickavelu A, Niwa S, Ayumi K, Komatsu K, Naruoka Y, Ban T. Molecular evaluation of Afghan wheat landraces. Plant Genet Resour 2014;12:S31–5. doi:10.1017/S1479262114000203.

[121] Iqbal M, Shahzad A, Ahmed I. Allelic variation at the Vrn-A1, Vrn-B1, Vrn-D1, Vrn-B3 and Ppd-D1a loci of Pakistani spring wheat cultivars. Electron J Biotechnol 2011;14:1–8. doi:10.2225/vol14-issue1-fulltext-6.

[122] Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH. Allelic variation at the vernalization genes Vrn-A1, Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. Crop Sci 2008;48:458–70. doi:10.2135/cropsci2007.06.0355.

[123] Stelmakh AF. Genetic effects of Vrn genes on heading date and agronomic traits in bread wheat. Euphytica 1993;65:53–60. doi:10.1007/BF00022199.

[124] Van Beem J, Mohler V, Lukman R, Van Ginkel M, William M, Crossa J, et al. Analysis of genetic factors influencing the developmental rate of globally important CIMMYT wheat cultivars. Crop Sci 2005;45:2113–9. doi:10.2135/cropsci2004.0665.

[125] Iqbal M, Navabi A, Yang RC, Salmon DF, Spaner D. The effect of vernalization genes on earliness and related agronomic traits of spring wheat in northern growing regions. Crop Sci 2007;47:1031–9.

doi:10.2135/cropsci2006.09.0618.

[126] Dubcovsky J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L. Effect of photoperiod on the regulation of wheat vernalization genes VRN1 and VRN2. Plant Mol Biol 2006;60:469–80. doi:10.1007/s11103-005-4814-2.

[127] Dubcovsky J, Chen C, Yan L. Molecular characterization of the allelic variation at the VRN-H2 vernalization locus in barley. Mol Breed 2005;15:395–407. doi:10.1007/s11032-005-0084-6.

[128] Bonnin I, Rousset M, Madur D, Sourdille P, Dupuits C, Brunel D, et al. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. Theor Appl Genet 2008;116:383–94. doi:10.1007/s00122-007-0676-0.

[129] Li C, Dubcovsky J. Wheat FT protein regulates VRN1 transcription through interactions with FDL2. Plant J 2008;55:543–54. doi:10.1111/j.1365-313X.2008.03526.x.

[130] Li C, Lin H, Dubcovsky J. Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. Plant J 2015;84:70–82. doi:10.1111/tpj.12960.

[131] Chen F, Gao M, Zhang J, Zuo A, Shang X, Cui D. Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. BMC Plant Biol 2013;13. doi:10.1186/1471-2229-13-199.

[132] Levy J, Peterson M. Responses of spring wheats to vernalization and photoperiod. Crop Sci 1972;12:487–90.

[133] Worland AJ, Borner A, Korzun V, Li WM, Petrovic S, Sayers EJ. The influence of photoperiod genes on the adaptability of European winter wheats. Euphytica 1998;100:385–94. doi:10.1023/A:1018327700985.

[134] Košner J, Pánková K. The detection of allelic variants at the recessive vrn loci of winter wheat. Euphytica 1998;101:9–16. doi:10.1023/a:1018394222868.

[135] Christie JM, Blackwood L, Petersen J, Sullivan S. Plant flavoprotein photoreceptors. Plant Cell Physiol 2015;56:401–13. doi:10.1093/pcp/pcu196.

[136] Moglich A, Yang X, Ayers RA, Moffat K. Structure and function of plant photoreceptors. Annu Rev Plant Biol 2010;61:21–47. doi:10.1146/annurev-

arplant-042809-112259.

[137] Neff MM, Fankhauser C, Chory J. Light: an indicator of time and place. Genes Dev 2000;14:257–71. doi:10.1101/gad.14.3.257.

[138] Sharrock RA, Quail PH. Novel phytochrome sequences in Arabidopsis thaliana: structure, evolution, and differential expression of a plant regulatory photoreceptor family. Genes Dev 1989;3:1745–57. doi:10.1101/gad.3.11.1745.

[139] Mathews S, Sharrock R a. Phytochrome gene diversity. Plant Cell Environ 1997;20:666–71. doi:10.1046/j.1365-3040.1997.d01-117.x.

[140] Nathan C. Rockwell, Yi-Shin Su and JCL. Phytochome Structure and Signaling Mechanisms. Annu Rev Plant Biol 2006;57:837–58. doi:10.1146/annurev.arplant.56.032604.144208.PHYTOCHOME.

[141] Chen M, Chory J. Phytochrome signaling mechanisms and the control of plant development 2012;21:664–71. doi:10.1016/j.tcb.2011.07.002.Phytochrome.

[142] Jiao Y, Lau OS, Deng XW. Light-regulated transcriptional networks in higher plants. Nat Rev Genet 2007;8:217–30. doi:10.1038/nrg2049.

[143] Quail PH. Phytochrome regulated gene expression. J Integr Plant Biol 2007;49:11–20. doi:10.1111/j.1672-9072.2007.00422.x.

[144] Lin C, Shalitin D. C RYPTOCHROME STRUCTURE AND SIGNAL TRANSDUCTION. Annu Rev Plant Biol 2003;54:469–96. doi:10.1146/annurev.arplant.54.110901.160901.

[145] Yu X, Sayegh R, Maymon M, Warpeha K, Klejnot J, Yang H, et al. Formation of nuclear bodies of Arabidopsis CRY2 in response to blue light is associated with its blue light-dependent degradation. Plant Cell 2009;21:118–30. doi:10.1105/tpc.108.061663.

[146] Yang HQ, Wu YJ, Tang RH, Liu D, Liu Y, Cashmore a R. The C termini of Arabidopsis cryptochromes mediate a constitutive light response. Cell 2000;103:815–27. doi:10.1016/S0092-8674(00)00184-7.

[147] Cashmore AR, Jarillo JA, Wu Y-J, Liu D. Cryptochromes: Blue Light Receptors for Plants and Animals. Science 1999;284:760–5. doi:10.1126/science.284.5415.760.

[148] Lian HL, He SB, Zhang YC, Zhu DM, Zhang JY, Jia KP, et al. Blue-light-dependent interaction of cryptochrome 1 with SPA1 defines a dynamic signaling mechanism. Genes Dev 2011;25:1023–8. doi:10.1101/gad.2025111.

[149] Liu B, Zuo Z, Liu H, Liu X, Lin C. Arabidopsis cryptochrome 1 interacts with SPA1 to suppress COP1 activity in response to blue light. Genes Dev 2011;25:1029–34. doi:10.1101/gad.2025011.

[150] Liu H, Yu X, Li K, Klejnot J, Yang H, Lisiero D, et al. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. Science 2008;322:1535–9. doi:10.1126/science.1163927.

[151] Kleine T, Lockhart P, Batschauer A. An Arabidopsis protein closely related to Synechocystis cryptochrome is targeted to organelles. Plant J 2003;35:93–103. doi:10.1046/j.1365-313X.2003.01787.x.

[152] Selby CP, Sancar A. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. Proc Natl Acad Sci U S A 2006;103:17696–700. doi:10.1073/pnas.0607993103.

[153] Pokorny R, Klar T, Hennecke U, Carell T, Batschauer A, Essen LO. Recognition and repair of UV lesions in loop structures of duplex DNA by DASH-type cryptochrome. Proc Natl Acad Sci USA 2008;105:21023–7. doi:10.1073/pnas.0805830106.

[154] Laurie DA. Comparative genetics of flowering time. Plant Mol Biol 1997;35:167–77. doi:10.1023/A:1005726329248.

[155] Beales J, Turner A, Griffiths S, Snape JW, Laurie D a. A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theor Appl Genet 2007;115:721–33. doi:10.1007/s00122-007-0603-4.

[156] Khlestkina EK, Giura A, Röder MS, Börner A. A new gene controlling the flowering response to photoperiod in wheat. Euphytica 2009;165:579–85. doi:10.1007/s10681-008-9783-5.

[157] Turner A. The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. Science 2005;310:1031–4. doi:10.1126/science.1117619.

[158] Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, et al. Cloning of the Arabidopsis Clock Gene TOC1,an Autoregulatory Response Regulator Homolog. Science 2000;289:768–71. doi:10.1126/science.289.5480.768.

[159] Campoli C, Shtaya M, Davis SJ, von Korff M. Expression conservation within the circadian clock of a monocot: natural variation at barley Ppd-H1 affects circadian expression of flowering time genes, but not clock orthologs. BMC Plant Biol 2012;12:97. doi:10.1186/1471-2229-12-97.

[160] Wilhelm EP, Turner AS, Laurie DA. Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (Triticum durum Desf.). Theor Appl Genet 2009;118:285–94. doi:10.1007/s00122-008-0898-9.

[161] Cane K, Eagles HA, Laurie DA, Trevaskis B, Vallance N, Eastwood RF, et al. Ppd - B1 and Ppd - D1 and their effects in southern Australian wheat. Crop Pasture Sci 2013;64:100–14.

[162] Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, et al. Structural variation in the 5′ upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (Triticum aestivum L.), and their effect on heading time. Mol Breed 2013;31:27–37. doi:10.1007/s11032-012-9765-0.

[163] Tanio M, Kato K. Development of Near-isogenic Lines for Photoperiod-insensitive Genes, Ppd-B1 and Ppd-D1, Carried by the Japanese Wheat Cultivars and their Effect on Apical Development. Breed Sci 2007;57:65–72. doi:10.1270/jsbbs.57.65.

[164] Pruneda-Paz JL, Breton G, Para A, S.A. K. A Functional Genomics Approach Reveals CHE as a Component of the Arabidopsis Circadian Clock. Science 2009;323:1481–5.

[165] Sun H, Guo Z, Gao L, Zhao G, Zhang W, Zhou R, et al. DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (Triticum aestivum). New Phytol 2014;204:682–92. doi:10.1111/nph.12948.

[166] Worland AJ. The influence of flowering time genes on environmental adaptability in European wheats Vernalization sensitivity. Euphytica 1996;89:49–57. doi:10.1007/BF00015718.

[167] Slafer G a, Araus JL, Royo C, García Del Moral LF. Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. Ann Appl Biol 2005;146:61–70. doi:10.1111/j.1744-7348.2005.04048.x.

[168] Worland AJ, Appendino ML, Sayers EJ. The distribution, in European winter wheats, of genes that influence ecoclimatic adaptability whilst determining photoperiodic insensitivity and plant height. Euphytica 1994;80:219–28. doi:10.1007/BF00039653.

[169] Košner J, Žurková D. Photoperiodic response and its relation to earliness in wheat. Euphytica 1996;89:59–64. doi:10.1007/BF00015719.

[170] Martinić JZ. Life Cycle of Common Wheat Varietie in Natural Enviroments as Related to their Response of Shortened Photoperiod. Z Pflanzenzüchtg 1975;75:237–51.

[171] Zakhrabekova S, Gough SP, Braumann I, Müller AH, Lundqvist J, Ahmann K, et al. Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. Proc Natl Acad Sci U S A 2012;109:4326–31. doi:10.1073/pnas.1113009109.

[172] Mizuno N, Nitta M, Sato K, Nasuda S. A wheat homologue of PHYTOCLOCK 1 is a candidate gene conferring the early heading phenotype to einkorn wheat. Genes Genet Syst 2012;87:357–67. doi:10.1266/ggs.87.357.

[173] Gawroński P, Ariyadasa R, Himmelbach A, Poursarebani N, Kilian B, Stein N, et al. A distorted circadian clock causes early flowering and temperature-dependent variation in spike development in the Eps-3Am mutant of einkorn wheat. Genetics 2014;196:1253–61. doi:10.1534/genetics.113.158444.

[174] Faure S, Turner AS, Gruszka D, Christodoulou V, Davis SJ, von Korff M, et al. Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley (Hordeum vulgare) to short growing seasons. Proc Natl Acad Sci U S A 2012;109:8328–33. doi:10.1073/pnas.1120496109.

[175] Campoli C, Pankin A, Drosse B, Casao CM, Davis SJ, Von Korff M. HvLUX1 is a candidate gene underlying the early maturity 10 locus in barley: Phylogeny, diversity, and interactions with the circadian clock and photoperiodic pathways. New Phytol 2013;199:1045–59.

doi:10.1111/nph.12346.

[176] Kato, Taketa, Ban, Iriki, Murai. The influence of a spring habit gene, Vrn-D1, on heading time in wheat. Plant Breed 2001;120:115–20. doi:10.1046/j.1439-0523.2001.00586.x.

[177] Kato K, Miura H, Sawada S. QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat. Theor Appl Genet 1999;98:472–7. doi:10.1007/s001220051094.

[178] Kamran A, Iqbal M, Spaner D. Flowering time in wheat (Triticum aestivum L.): A key factor for global adaptability. Euphytica 2014;197:1–26. doi:10.1007/s10681-014-1075-7.

[179] Lewis S, Faricelli ME, Appendino ML, Valárik M, Dubcovsky J. The chromosome region including the earliness per se locus Eps-A m1 affects the duration of early developmental phases and spikelet number in diploid wheat. J Exp Bot 2008;59:3595–607. doi:10.1093/jxb/ern209.

[180] Slafer GA, Rawson HM. Intrinsic earliness and basic development rate assessed for their response to temperature in wheat. Euphytica 1995;83:175–83. doi:10.1007/BF01678127.

[181] Kato K, Wada T. Genetic analysis and selection experiment for narrow-sense earliness in wheat by using segregating hybrid progenies. Breed Sci 1999;49:233–8. doi:10.1270/jsbbs.49.233.

[182] Bullrich L, Appendino ML, Tranquilli G, Lewis S, Dubcovsky J. Mapping of a thermo-sensitive earliness per se gene on Triticum monococcum chromosome 1Am. Theor Appl Genet 2002;105:585–93. doi:10.1007/s00122-002-0982-5.

[183] Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, et al. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 2012;44:1388–92. doi:10.1038/ng.2447.

[184] Laurie DA, Pratchett N, Bezant JH, Snape JW. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (Hordeum vulgare L.) cross. Genome 1995;38:575–85.

[185] Faricelli ME, Valárik M, Dubcovsky J. Control of flowering time and spike development in cereals: The earliness per se Eps-1 region in wheat, rice, and Brachypodium. Funct Integr Genomics 2010;10:293–306. doi:10.1007/s10142-009-0146-7.

[186] Alvarez MA, Tranquilli G, Lewis S, Kippes N, Dubcovsky J. Genetic and physical mapping of the earliness per se locus Eps-Am1 in Triticum monococcum identifies EARLY FLOWERING 3 (ELF3) as a candidate gene. Funct Integr Genomics 2016:1–18. doi:10.1007/s10142-016-0490-3.

[187] Zikhali M, Wingen LU, Griffiths S. Delimitation of the Earliness per se D1 (Eps-D1) flowering gene to a subtelomeric chromosomal deletion in bread wheat (Triticum aestivum). J Exp Bot 2016;67:287–99. doi:10.1093/jxb/erv458.

[188] Lundqvist U. Scandinavian mutation research in barley - a historical review. Hereditas 2014;151:123–31. doi:10.1111/hrd2.00077.

[189] Dvorák J, Terlizzi P, Zhang HB, Resta P. The evolution of polyploid wheats: identification of the A genome donor species. Genome 1993;36:21–31. doi:10.1139/g93-004.

[190] Dvorak J, McGuire PE, Cassidy B. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 1988;30:680–9. doi:10.1139/g88-115.

[191] Dvorák J, Zhang HB. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc Natl Acad Sci U S A 1990;87:9640–4. doi:10.1073/pnas.87.24.9640.

[192] Nesbitt M, Samuel D. From staple crop to extinction? The archaeology and history of the hulled wheats. In: Padulosi S, Hammer K, Heller J, editors. Hulled wheats. Proc. First Int. Work. Hulled Wheats, vol. 4, Castelvecchio Pascoli, Tuscany, Italy; 1996, p. 41–100.

[193] Dvorak J, Luo MC, Yang ZL, Zhang HB. The structure of the Aegilops tauschii genepool and the evolution of hexaploid wheat. Theor Appl Genet 1998;97:657–70. doi:10.1007/s001220050942.

[194] Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, Dvorak J. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet 2007;114:947–59. doi:10.1007/s00122-006-0474-0.

[195] Nesbitt M. When and Where did domesticated cereals first occur in southwest Asia? In: Cappers RTJ, Bottema S, editors. Daw. farming Near East, Berlin, ex oriente; 2002, p. 113–32.

[196] Mori, N.; Liu, Y.-G., Tsunewaki K. Wheat phylogeny determined by RFLP analysis of nuclear DNA . Theor Appl Genet 1995;90:129–34.

[197] Brown TA, Jones MK, Powell W, Allaby RG. The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol Evol 2009;24:103–9. doi:10.1016/j.tree.2008.09.008.

[198] Tang H, Sezen U, Paterson AH. Domestication and plant genomes. Curr Opin Plant Biol 2010;13:160–6. doi:10.1016/j.pbi.2009.10.008.

[199] Nevo E. Evolution of wild emmer wheat and crop improvement. J Syst Evol 2014;52:673–96. doi:10.1111/jse.12124.

[200] Peng J, Ronin Y, Fahima T, Röder MS, Li Y, Nevo E, et al. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proc Natl Acad Sci 2003;100:2489–94. doi:10.1073/pnas.252763199.

[201] Peng JH, Sun D, Nevo E. Domestication evolution, genetics and genomics in wheat. Mol Breed 2011;28:281–301. doi:10.1007/s11032-011-9608-4.

[202] Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science (80-) 2017;357:93–7.

[203] Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, Gill BS, et al. Molecular characterization of the major wheat domestication gene Q. Genetics 2006;172:547–55. doi:10.1534/genetics.105.044727.

[204] Kerber ER, Rowland GG. Origin of the Free Threshing Character in Hexaploid Wheat. Can J Genet Cytol 1974;16:145–54. doi:10.1139/g74-014.

[205] Greenwood JR, Finnegan EJ, Watanabe N, Trevaskis B, Swain SM. New alleles of the wheat domestication gene *Q* reveal multiple roles in growth and

reproductive development. Development 2017:dev.146407. doi:10.1242/dev.146407.

[206] Sood S, Kuraparthy V, Bai G, Gill BS. The major threshability genes soft glume (sog) and tenacious glume (Tg), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. Theor Appl Genet 2009;119:341–51. doi:10.1007/s00122-009-1043-0.

[207] Jantasuriyarat C, Vales MI, Watson CJW, Riera-Lizarazu O. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (Triticum aestivum L.). Theor Appl Genet 2004;108:261–73. doi:10.1007/s00122-003-1432-8.

[208] Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, et al. A genetic framework for grain size and shape variation in wheat. Plant Cell 2010;22:1046–56. doi:10.1105/tpc.110.074153.

[209] Nave M, Avni R, Ben-Zvi B, Hale I, Distelfeld A. QTLs for uniform grain dimensions and germination selected during wheat domestication are co-located on chromosome 4B. Theor Appl Genet 2016;129:1303–15. doi:10.1007/s00122-016-2704-4.

[210] Hillman GC, Davies MS. Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. J World Prehistory 1990;4:157–222. doi:10.1007/BF00974763.

[211] Tanno K, Willcox G. How Fast Was Wild Wheat Domesticated? Science 2006;311:1886.

[212] Civáň P, Ivaničová Z, Brown TA. Reticulated origin of domesticated emmer wheat supports a dynamic model for the emergence of agriculture in the fertile crescent. PLoS One 2013;8. doi:10.1371/journal.pone.0081955.

[213] Talbert LE, Smith LY, Blake MK. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome 1998;41:402–7. doi:10.1139/gen-41-3-402.

[214] Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to

hexaploid wheat. Mol Biol Evol 2006;23:1386–96. doi:10.1093/molbev/msl004.

[215] Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 2005;142:169–96. doi:10.1007/s10681-005-1681-5.

[216] Semagn K, Bjørnstad Å, Ndjiondjop MN. Principles , requirements and prospects of genetic mapping in plants. African J Biotechnol 2006;5(25):2569–87. doi:10.5897/AJB2006.000-5112.

[217] Babu R, Nair SK, Prasanna BM, Gupta HS. Integrating marker-assisted selection in crop breeding–prospects and challenges. Curr Sci 2004;87:607–19.

[218] Krattinger S, Wicker T, Keller B. Genetics and Genomics of the Triticale. In: Feuillet C, Muehlbauer GJ, editors. Book, vol. 7, Springer New York; 2009, p. 337–58. doi:10.1007/978-0-387-77489-3.

[219] Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, et al. Integrating cereal genomics to support innovation in the Triticeae. Funct Integr Genomics 2012;12:573–83. doi:10.1007/s10142-012-0300-5.

[220] Staňková H. Construction of physical map of 7DS wheat chromosome arm and its use for positional cloning. 2015.

[221] Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. SNP markers and their impact on plant breeding. Int J Plant Genomics 2012:12pp. doi:10.1155/2012/728398.

[222] Jiang Z, Wang H, Michal JJ, Zhou X, Liu B, Woods LCS, et al. Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development. Int J Biol Sci 2016;12:100–8. doi:10.7150/ijbs.13498.

[223] Lander ES, Green P, Abrahamson J, Barlow H, Daly M, Lincoln S, et al. MAPMAKER:An interactive computer program for constructing genetic maps of experimental and natural populations. Genomics 1987;1:174–81. doi:DOI 10.1016/j.ygeno.2008.12.003.

[224] Lander ES, Green P. Construction of multilocus genetic linkage maps in

humans. Proc Natl Acad Sci 1987;84:2363–7. doi:10.1073/pnas.84.8.2363.

[225] Stam P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J 1993;3:739–44. doi:10.1111/j.1365-313X.1993.00739.x.

[226] Manly KF, Cudmore RH, Meer JM. Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome 2001;12:930–2. doi:10.1007/s00335-001-1016-3.

[227] Cheema J, Dicks J. Computational approaches and software tools for genetic linkage map estimation in plants. Brief Bioinform 2009;10:595–608. doi:10.1093/bib/bbp045.

[228] Buetow KH, Chakravarti A. Multipoint gene mapping using seriation. 1 General methods. Am J Hum Genet 1987;41:180–8.

[229] Lander ES, Botstein S. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 1989;121:185–99. doi:10.1038/hdy.2014.4.

[230] Zheng. Precision Mapping of Quantitative Trait Loci. Genetics 1994;136:1457–68.

[231] Zheng Z. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 1993;90:10972–6.

[232] Jansen RC, Stam P. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 1994;136:1447–55.

[233] Jansen RC. Interval mapping of multiple quantitative trait loci. Genetics 1993;135:205–11.

[234] Risch N. Genetic linkage: interpreting lod scores. Science 1992;255:803–4. doi:10.1126/science.1536004.

[235] Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994;138:963–71. doi:10.1534/genetics.107.080101.

[236] Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7. doi:10.1038/ng1195-241.

[237] Conrad LJ, Kikuchi K, Brutnell TP. Transposon Tagging in Cereal Crops. In:

Kahl G, Meksem K, editors. Handb. Plant Funct. Genomics, Weinheim: Wiley-Blackwell; 2008, p. 267–90.

[238] Voytas DF. Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 2013;64:327–50. doi:10.1146/annurev-arplant-042811-105552.

[239] Kahl G, Meksem K, editors. The Handbook of Plant funcional Genomics. Weinheim: Wiley-Blackwell; 2008.

[240] Zhang F, Puchta H, Thomson JG, editors. Advances in New Technology for Targeted Modification of Plant Genomes. Springer New York; 2015.

[241] Fu D, Uauy C, Blechl A, Dubcovsky J. RNA interference for wheat functional gene analysis. Transgenic Res 2007;16:689–701. doi:10.1007/s11248-007-9150-7.

[242] Waterworth WM, Drury GE, Bray CM, West CE. Repairing breaks in the plant genome: The importance of keeping it together. New Phytol 2011;192:805–22. doi:10.1111/j.1469-8137.2011.03926.x.

[243] Wright DA, Townsend JA, Winfrey RJ, Irwin PA, Rajagopal J, Lonosky PM, et al. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 2005;44:693–705. doi:10.1111/j.1365-313X.2005.02551.x.

[244] Puchta H, Dujon B, Hohn B. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci 1996;93:5055–60. doi:10.1073/pnas.93.10.5055.

[245] Puchta H, Dujon B, Hohn B. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res 1993;21:5034–40. doi:10.1093/nar/21.22.5034.

[246] Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, et al. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 2006;34:1–12. doi:10.1093/nar/gkl720.

[247] Pâques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther

2007;7:49–66. doi:10.2174/156652307779940216.

[248] Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, et al. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 2009;37:5405–19. doi:10.1093/nar/gkp548.

[249] Kay S, Hahn S, Marois E, Hause G, Bonas U. A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. Science 2007;318:648–51.

[250] Mak AN, Bradley P, Cernadas R a, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 2012;335:716–9. doi:10.1126/science.1216211.The.

[251] Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, et al. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science 2012;335:720–3. doi:10.1016/j.cmet.2012.08.002.

[252] Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 2010;186:756–61. doi:10.1534/genetics.110.120717.

[253] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. Science 2012;337:816–22. doi:10.1126/science.1225829.

[254] Li W, Teng F, Li T, Zhou Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat Biotechnol 2013;31:686–8. doi:10.1038/nbt.2652.

[255] Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 2013;31:691–3. doi:10.1038/nbt.2655.

[256] Li J, Aach J, Norville JE, Mccormack M, Bush J, Church GM, et al. Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. Nat Biotechnol 2013;31:688–91. doi:10.1038/nbt.2654.

[257] Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol 2014;32:947–52.

doi:10.1038/nbt.2969.

[258] Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol 2015;33:1162–4. doi:10.1038/nbt.3389.

[259] Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 2013;31:839–43. doi:10.1038/nbt.2673.High-throughput.

[260] Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N, et al. Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biol 2015;16:258–70. doi:10.1186/s13059-015-0826-7.

[261] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013;31:827–32. doi:10.1038/nbt.2647.

[262] Fu Y, Foden J a, Khayter C, Maeder ML, Reyon D, Joung K, et al. High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 2013;31:822–6. doi:10.1038/nbt.2623.

[263] Ran FA, Hsu PPD, Wright J, Agarwala V, Scott D a, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 2013;8:2281–308. doi:10.1038/nprot.2013.143.

[264] Puchta H. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr Opin Plant Biol 2017;36:1–8. doi:10.1016/j.pbi.2016.11.011.

[265] Houben A, Schubert I. Engineered plant minichromosomes: a resurrection of B chromosomes? Plant Cell 2007;19:2323–7. doi:10.1105/tpc.107.053603.

[266] Hiei Y, Ishida Y, Komari T. Progress of cereal transformation technology mediated by Agrobacterium tumefaciens. Front Plant Sci 2014;5. doi:10.3389/fpls.2014.00628.

[267] Ishida Y, Tsunashima M, Hiei Y, Komari T. Agrobacterium protocols: Third edition. In: Wang K, editor. Agrobacterium Protoc. Vol. 1, vol. 1223, Springer

Science+Business Media New York; 2014, p. 189–98. doi:10.1007/978-1-4939-1695-5.

[268] Sparks CA, Jones HD. Biolistic Transformation of wheat. In: Jones HD, Shewry PR, editors. Methods Mol. Biol. Trangenic Wheat, Barley Oats, vol. 478, Humana Press; 2009, p. 71–92. doi:10.1007/978-1-59745-379-0.

[269] Travella S, Ross SM, Harden J, Everett C, Snape JW, Harwood WA. A comparison of transgenic barley lines produced by particle bombardment and Agrobacterium-mediated techniques. Plant Cell Rep 2005;23:780–9. doi:10.1007/s00299-004-0892-x.

[270] Himmelbach A, Zierold U, Hensel G, Riechen J, Douchkov D, Schweizer P, et al. A Set of Modular Binary Vectors for Transformation of Cereals. Plant Physiol 2007;145:1192–200. doi:10.1104/pp.107.111575.

[271] Harwood WA. Advances and remaining challenges in the transformation of barley and wheat. J Exp Bot 2012;63:1791–8. doi:10.1093/jxb/err380.

[272] Salvo-Garrido H, Travella S, Billiam LJ, Harwood WA, Snape JW. The distribution of transgene insertion sites in barley determined by physical and genetic mapping. Genetics 2004;167:1371–9. doi:10.1534/genetics.103.023747.

[273] Shrawat AK, Becker D, Lörz H. Agrobacterium tumefaciens-mediated genetic transformation of barley (Hordeum vulgare L.). Plant Sci 2007;172:281–90. doi:10.1016/j.plantsci.2006.09.005.

[274] Tenea GN, Spantzel J, Lee L-Y, Zhu Y, Lin K, Johnson SJ, et al. Overexpression of Several Arabidopsis Histone Genes Increases Agrobacterium-Mediated Transformation and Transgene Expression in Plants. Plant Cell 2009;21:3350–67. doi:10.1105/tpc.109.070607.

[275] Schulze J. Improvements in Cereal Tissue Culture by Thidiazuron: A Review. Fruit, Veg Cereal Sci Biotechnol 2007;1:64–79.

[276] Dan Y, Armstrong CL, Dong J, Feng X, Fry JE, Keithly GE, et al. Lipoic acid-an unique plant transformation enhancer. Vitr Cell Dev Biol - Plant 2009;45:630–8. doi:10.1007/s11627-009-9227-5.

[277] Barro F, Martin A, Lazzeri PA, Barceló P. Medium optimization for efficient

somatic embryogenesis and plant regeneration from immature inflorescences and immature scutella of elite cultivars of wheat, barley and tritordeum. Euphytica 1999;108:161–7. doi:10.1023/A:1003676830857.

[278] Bartlet J, Alves S, Smedley M, Snape J, Harwood W. Agrobacterium-mediated barley transformation. Plant Methods 2008;4. doi:10.1186/1746-4811-4-22.

[279] Tyagi N, Dahleen LS, Bregitzer P. Candidate genes within tissue culture regeneration QTL revisited with a linkage map based on transcript-derived markers. Crop Sci 2010;50:1697–707. doi:10.2135/cropsci2009.10.0624.

[280] Nishimura A, Ashikari M, Lin S, Takashi T, Angeles ER, Yamamoto T, et al. Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. Proc Natl Acad Sci U S A 2005;102:11940–4. doi:10.1073/pnas.0504220102.

[281] El-Assal Sel D, Abd-Alla SM, El-Tarras AA, El-Awady MA. Production of early flowering transgenic barley expressing the early flowering allele of Cryptochrome2 gene. GM Crop 2011;2:50–7. doi:10.4161/gmcr.2.1.15627.

3 AIMS OF THE THESIS

The main aim of the thesis was identification and characterization of new genes/alleles responsible for determination of flowering time in wheat. This aim was accomplished by studying two different flowering pathways. The obtained results are represented by two separated parts:

I. Identification and characterization of new *Vrn-A1f-like* allele responsible for delayed flowering.

The aim of the first part of the thesis was determination and following characterization of gene responsible for difference in flowering time within doubled haploid mapping population. This population consists of individuals which contain different combination of introgression from *Triticum millitinae* in elite bread wheat cultivar Tähti.

II. Analysis of expression profile and methylation pattern of lines with three copy of *Ppd-B1a* gene as a possible cause of intragroup heading time variation.

The second part of the thesis dealt with identification of gene responsible for difference in flowering time between two bread wheat cultivars – Paragon and Kaerntner Frueher. The aim was further complemented by analysis of lines which have a common character – three copies of *Ppd-B1* gene, although they significantly differed in flowering time.

4 RESULTS

4.1 Original papers

4.1.1 Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*

4.1.2 Heritable heading time variation in wheat lines with the same number of *Ppd-B1* gene copies

4.1.3 Kvitnúť či nekvitnúť

4.1.4 Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent

4.1.5 Taxonomical classification and origin of Kamut®

4.1.1 *Characterization of new allele influencing flowering time in bread wheat introgressed from Triticum militinae*

Ivaničová Z., Jakobson I., Reis D., Šafář J., Milec Z., Abrouk M., Doležel J., Järve K. and Valárik M.

New Biotechnology, 33 (5B), 718-727, 2015

doi:10.1016/j.nbt.2016.01.008.

IF 3.199

Flowering time variation was identified within a mapping population of doubled haploid lines developed from a cross between the introgressive line 8.1 and spring bread wheat cv. Tähti. The line 8.1 carried introgressions from tetraploid *Triticum militinae* in the cv. Tähti genetic background on chromosomes 1A, 2A, 4A, 5A, 7A, 1B and 5B. The most significant QTL for the flowering time variation was identified within the introgressed region on chromosome 5A and its largest effect was associated with the *VRN-A1* locus, accounting for up to 70% of phenotypic variance. The allele of T. militinae origin was designated as *VRN-A1f-like*. The effect of the *VRN-A1f-like* allele was verified in two other mapping populations. QTL analysis identified that in cv. Tähti and cv. Mooni genetic background, *VRN-A1f-like* allele incurred a delay of 1.9–18.6 days in flowering time, depending on growing conditions. Sequence comparison of the *VRN-A1f-like* and *VRN-A1a* alleles from the parental lines of the mapping populations revealed major mutations in the promoter region as well as in the first intron, including insertion of a MITE element and a large deletion. The sequence variation allowed construction of specific diagnostic PCR markers for *VRN-A1f-like* allele determination. Identification and quantification of the effect of the *VRN-A1f-like* allele offers a useful tool for wheat breeding and for studying fine-scale regulation of flowering pathways in wheat.

4.1.2 *Heritable heading time variation in wheat lines with the same number of Ppd-B1 gene copies*

Ivaničová Z., Valárik M., Pánková K., Trávníčková M., Doležel J., Šafář J. and Milec Z.

PLoS ONE, 12(8): e0183745, 2017

doi: 10.1371/journal.pone.0183745

3.54

The ability of plants to identify an optimal flowering time is critical for ensuring the production of viable seeds. The main environmental factors that influence the flowering time include the ambient temperature and day length. In wheat, the ability to assess the day length is controlled by photoperiod (*Ppd)* genes. Due to its allohexaploid nature, bread wheat carries the following three *Ppd-1* genes: *Ppd-A1*, *Ppd-B1* and *Ppd-D1*. While photoperiod (in)sensitivity controlled by *Ppd-A1* and *Ppd-D1* is mainly determined by sequence changes in the promoter region, the impact of the *Ppd-B1* alleles on the heading time has been linked to changes in the copy numbers (and possibly their methylation status) and sequence changes in the promoter region. Here, we report that plants with the same number of *Ppd-B1* copies may have different heading times. Differences were observed among F lines derived from crossing two spring hexaploid wheat varieties. Several lines carrying three copies of *Ppd-B1* headed 16 days later than other plants in the population with the same number of gene copies. This effect was associated with changes in the gene expression level and methylation of the *Ppd-B1* gene.

4.1.3 *Kvitnúť či nekvitnúť*

Ivaničová Z.

Svet prírody, 4, 19-21, 2016

ISSN: 1339 - 5289

Reviewed journal

Pšenica je jednou z najdôležitejších hospodárskych plodín. Poskytuje zdroj potravy pre 35 % celosvetovej populácie. V dôsledku neustáleho rastu ľudskej populácie je nevyhnutné zvyšovať jej výnos. Tento cieľ je však ohrozovaný celým radom biotických a abiotických faktorov. V dôsledku zmien klimatických podmienok sme v poslednej dobe svedkami značných výkyvov počasia. Jarné mrazy môžu viesť k poškodeniu kvetných častí (Obr. 1), suché a horúce letá ohrozujú ukladanie zásobných látok do vyvíjajúcich sa obiliek. Preto je potrebné vysievať pšenice, ktoré budú lepšie prispôsobené lokálnym podmienkam. Ich kvitnutie je potrebné načasovať tak, aby nedošlo k poškodeniu kvetných častí a rastlina optimálne využila dostupné zdroje živín potrebných pre vývin a naplnenie semien. Pokiaľ sa rastlina rozhodne kvitnúť v nesprávnu dobu nevyprodukuje žiadne semená. Preto u pšenice, ktorá je jednoročnou rastlinou, je rozhodnutie kvitnúť alebo nekvitnúť otázkou života a smrti.

4.1.4 *Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent*

Civáň P., Ivaničová Z. and Brown T.A.

PLoS ONE 8(11), e81955, 2013

doi:10.1371/journal.pone.0081955

IF 3.54

We used supernetworks with datasets of nuclear gene sequences and novel markers detecting retrotransposon insertions in ribosomal DNA loci to reassess the evolutionary relationships among tetraploid wheats. We show that domesticated emmer has a reticulated genetic ancestry, sharing phylogenetic signals with wild populations from all parts of the wild range. The extent of the genetic reticulation cannot be explained by post-domestication gene flow between cultivated emmer and wild plants, and the phylogenetic relationships among tetraploid wheats are incompatible with simple linear descent of the domesticates from a single wild population. A more parsimonious explanation of the data is that domesticated emmer originates from a hybridized population of different wild lineages. The observed diversity and reticulation patterns indicate that wild emmer evolved in the southern Levant, and that the wild emmer populations in south-eastern Turkey and the Zagros Mountains are relatively recent reticulate descendants of a subset of the Levantine wild populations. Based on our results we propose a new model for the emergence of domesticated emmer. During a pre-domestication period, diverse wild populations were collected from a large area west of the Euphrates and cultivated in mixed stands. Within these cultivated stands, hybridization gave rise to lineages displaying reticulated genealogical relationships with their ancestral populations. Gradual movement of early farmers out of the Levant introduced the pre-domesticated reticulated lineages to the northern and eastern parts of the Fertile Crescent, giving rise to the local wild populations but also facilitating fixation of domestication traits. Our model is consistent with the protracted and dispersed transition to agriculture indicated by the archaeobotanical evidence, and also with previous genetic data affiliating domesticated emmer with the wild populations in southeast Turkey. Unlike other protracted models, we assume that humans played an intuitive role throughout the process.

4.1.5 *Taxonomical classification and origin of Kamut®*

Michalcová V., Dušinský R., Sabo M., Al Beyroutiová M., Hauptvogel P., Ivaničová Z. and Švec M.

Plant Syst Evol, 300(7), 1749-1757, 2014

doi:10.1007/s00606-014-1001-4

IF 1.361

Bioagriculture and healthy lifestyle are trends of the twenty-first century. Bioagriculture involves the breeding of crops without using modern synthetic substances. Kamut® brand wheat is one of the popular biocereals grown mainly in the USA and Europe. This cereal has the status of ancient wheat, not only because it has been grown since the era of the ancient Egyptian civilization, but also for its properties favourable for modern breeding programs and modern food marketing. In spite of Kamut®'s interesting history and stable place in the market, it is not a common subject of genetic studies. It is also interesting that it has not been successfully taxonomically classified yet. There are a few studies which classify this tetraploid wheat as *Triticum polonicum* L., *T. turanicum* Jakubz., *T. turgidum* L. and *T. durum* Desf. These studies are based on cytological and comparative methods. We chose molecular (transposable element resistance gene analog polymorphism, diversity arrays technology, sequencing of genes *SBEIIa*, and *ψLpx-A1_like*) and statistical methods to classify Kamut® wheat. According to our experiments we suggest that Kamut® brand wheat originated as a natural hybrid between *Triticum dicoccon* conv. *dicoccon* and *T. polonicum* and is not original ancient Egyptian wheat. We suggest that Etruscan wheat has the same parents as Kamut®

4.2 Published abstract – poster presentation

4.2.1 Different day length – different effect

4.2.2 Chasing a new flowering time gene

4.2.3 High-throughput gene resources and phenotyping - the key factors in fine mapping of new *Eps* gene

4.2.4 Identification and characterization of a new *Vrn-A1f-like* allele responsible for flowering lateness in wheat

4.2.1 *Different day length – different effect*

Ivaničová Z., Milec Z., Trávníčková M., Valárik M., Prášil I.T., Pánková K., Snape J.W. and Šafář J.

In: Abstracts of the "Olomouc Biotech 2013. Plant Biotechnology: Green for Good II". Olomouc, Czech Republic, 2013

Accelerating and delaying of flowering time was identified in Czech wheat landrace Česká přesívka under long (LD) or short days (SD) conditions, respectively. The one day (LD) or two days (SD) difference indicates a possible role of earliness *per se* gene (*QFt.cri-3B.1* locus) in determination this phenotypic trait. The presumed gene was localized within 30 cM near-centromeric region of the long arm of chromosome 3B. This locus is delineated by *barc164* and *cfa2170* microsatellites markers. Additional phenotyping and genotyping of nearly isogenic lines mapping population is currently in progress. Phenotypic effect was identified (i) in controlled conditions as well as (ii) in field experiment. The differences in apex development (i) or in heading time (ii) proved bipolar effect of assumed gene. The map around the gene was saturated with DArT, SSR, STS markers. Up to now we identified 30 new polymorphic markers which have been used to delimitate region. Fine mapping followed by positional cloning and sequencing will enable identification gene of our interest. Candidate gene analysis and its detailed characterization will provide novel information which could lead to deeper understanding of fine tuning flowering time in wheat. Furthermore, determined effect could be useful in creating new varieties, finding out new growth areas in different climate conditions.

4.2.2 *Chasing a new flowering time gene*

Ivaničová Z., Šafář J., Trávničková M., Pánková K., Bartoš J., Milec Z. and Doležel, J.

In: Book of abstracts EUCARPIA Cereal SectionITMI Conference in Weringerode, p. 156, 2014

Bread wheat (*Triticum aestivum* L.) is a staple food of 40% of the world's population. Population growth and extreme weather fluctuations call for development of new cultivars resistant to environmental stress and with increased and more stable yields. The yield can be modified by a number of ways; one of them being the optimisation of flowering time (FT) to local climatic conditions.

Apart from major FT genes, which have already been determined, there are minor genes, some of which have been identified as QTL´s. Most of them have not been discovered yet and although our current knowledge of pathways regulating FT is improving, it remains far from complete. Clearly, there is a need to identify alleles and genes which may be useful in wheat improvement.

The main objective of our work is to localise a new FT gene. We have identified 12-day difference in heading time between wheat varieties Paragon and Kaerntner Frueher. There were no differences in the main FT genes (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *vrn-B3* and *Ppd-D1b*) between these two varieties except dissimilar alleles of *Rht8* gene (*Rht8a* and *Rht8g*). However, this variation should not cause a large difference in FT. Thus, we hypothesise that there may be a novel gene/allele involved. We observed a temperature-dependent effect of the hypothetic gene in field as well as in controlled conditions (16/8h, 20/16˚C), when lower temperature prolonged FT difference up to 28 and 30 days in field and growth room, respectively. In order to localise the source of this interesting phenotype, both lines were crossed and F_2 mapping population has been developed. A preliminary localisation of this gene will be done using DArT analysis, which will be followed by QTL analysis. As there are typically 2,000-3,000 DArT polymorphism identified within a wheat genome, we expect that the high marker density will permit identification of the locus involved.

4.2.3 *High-throughput gene resources and phenotyping - the key factors in fine mapping of new Eps gene*

Ivaničová Z., Milec Z., Trávníčková M., Abrouk M., Valárik M., Prášil I.T., Pánková K., Snape J.W. and Šafář, J.

In: Sborník abstrakt, Bulletin České společnosti experimentální biologie rostlin a Fyziologické sekce Slovenské botanické společnsti.
"6. Metodické dny". Seč, Česká republika, 2014

Acceleration and delay of heading time was identified in Sandra (CP3B) substitution line under long or SD conditions, respectively. The 3B chromosome in Sandra was substituted from Czech wheat landrace Česká přesívka. The heading time difference of one day under LDs or two days under SDs may be caused by earliness *per se* (*Eps*) gene. The gene was localized within 30 cM region of long arm of 3B chromosome and designated *QFt.cri-3B.1*. The locus was delineated by *barc164* and *cfa2170* microsatellites markers.

An F_4 NILs (Nearly Isogenic Lines) mapping population was developed by crossing Sandra (CP3B) substitution line with Sandra variety in order to localize gene precisely.

274 DArT, SSR and STS markers were used for saturation of the *QFt.cri-3B.1* region. To acquire more markers we flow sorted and sequenced 3B chromosome from Česká přesívka. Obtained sequences were assembled and *QFt.cri-3B.1* region was refined.

27,388 SNPs (Single Nucleotide Polymorphisms) specific for the *QFt.cri-3B.1* region were identified using SNPLauncher pipeline (developed by Helene Rimbert – INRA Clermont-Ferrand – GDEC). The identified SNPs have been used for sequence-based and KASP markers development. From all tested markers nearly 80% were successfully converted and applied on mapping population.

For fine mapping of the *QFt.cri-3B.1* locus, precise phenotype needs to be assessed. For this purpose a NILs mapping population is being developed.

Both F_5 NILs mapping population as well as obtained SNP genotyping sources will be used for precise fine mapping of the gene which will be eventually followed by positional cloning. Candidate gene/genes analysis and its detailed characterization will provide novel information which could lead to deeper understanding of fine tuning flowering time in wheat. Furthermore, determined effect could be useful in creating new varieties and finding new growth areas in different climate conditions.

4.2.4 *Identification and characterization of a new Vrn-A1f-like allele responsible for flowering lateness in wheat*

Ivaničová Z., Jakobson I., Reis D., , Milec Z., Abrouk M., Doležel J., Järve K., Šafář J. and Valárik M.

In: Proceedings 13th International Wheat Genetics Symposium, p. 173, 2017

KEY MESSAGE

Vrn-A1f-like allele caused flowering lateness in two bread wheat mapping populations. The new allele originated from *Triticum militinae* and was identified in 80% tetraploid spring wheat varieties with AAGG genome.

ABSTRACT

Flowering is a crucial process in plant life, which impacts crop yield. Here we identify tetraploid wheat *Triticum millitinae* (2n=4x=28, AAGG) as a resource for mining new genes/alleles influencing flowering time. We observed flowering time variation in a mapping population of doubled haploid (DH) lines. It was developed from a cross between introgressive line 8.1 and elite bread wheat cv. Tähti. Line 8.1 carries introgressions from *T. militinae*, in Tähti background. The analysis of DH mapping population revealed presence of seven loci influencing flowering time. The most significant QTL for the flowering time variation was identified within the introgressed region on chromosome 5A and its largest effect was associated with the *VRN-A1* locus, covering up to 67% of phenotypic variation. The analysis of F_2 mapping population developed from cross DH81 × cv. Mooni confirmed the effect of this locus on flowering time. DNA sequence analysis revealed the origin from *T. militinae* and the allele was designated as *VRN-A1f-like* (KT696537). The allele incurred a delay of 1.9 - 18.6 days in flowering in different growing conditions. Comparison *VRN-A1f-like* allele to *VRN-A1a* sequences from the common wheat parental lines of the mapping populations revealed major mutations in the promoter region as well as in the first intron, including a MITE insertion and a large deletion. Moreover, the allele was identified to be responsible for spring habit emergence in 80% of the analyzed tetraploid wheat varieties with AAGG genome. Finally, analysis of difference in the relative expression level between parental lines of two mapping populations (DH81, cv. Tähti and cv. Mooni) was carried out. The identification and quantification of the effect of the *VRN-A1f-like* allele from *T. militinae* provides a valuable source of new alleles suitable for wheat improvement as well as for studying fine regulation of flowering pathways in wheat.

4.3 Published abstract – oral presentation

4.3.1 Characterization of new alleles influencing flowering time introgressed to bread wheat from *T. millitinae*

4.3.1 *Characterization of new alleles influencing flowering time introgressed to bread wheat from T. millitinae*

Ivaničová Z., Jakobson I., Šafář J., Milec Z., Järve K., Valárik M. and Doležel J.

In: Abstracts of the "Olomouc Biotech 2015. Plant Biotechnology: Green for Good III". Olomouc, Czech Republic, 2015

The heading time variation was identified within dihaploid (DH) lines (135 lines) mapping population (MP) developed from cross between DH8/1 line and spring wheat variety Tähti. DH8/1 carried introgressions from *Triticum millitinae* (2n=28, AAGG) on chromosomes 1A, 2A, 4A, 5A, 7A, 1B and 5B in Tähti genetic background. The highly significant loci influencing phenotypic difference between DH lines were localized on chromosomes 5A and 7A and were responsible for 55,3 or 22,8% phenotypic variation. The significant loci were identified on chromosomes 1A, 2AS and 4A and results in 23,9% of variation (3,5; 8,8 and 9,6%, respectively). To further localize the identified QTLs two MP were developed from cross between DH8/1×DH397 and DH8/1×Mooni (hexaploid wheat variety). We were able to detect 10 or 25 days difference in F_3 DH8/1xMooni MP in Olomouc (field conditions) or Tallinn (controlled conditions). The 4 days heading time difference was identified in F_2 DH8/1×DH397 MP. Following QTL analysis pointed up the presence of main QTL on chromosome 5A. The QTL peak with highest significance was located in the region of vernalization gene (*Vrn-A1*). QTL analysis showed that *Vrn-A1* is responsible for 38 or 9% of phenotypic variation in F_3 or F_2 MPs, respectively. Obtaining the promoter and whole gene sequence of *Vrn-A1* could lead to identification of structural changes behind difference in heading time.

5 CONCLUSIONS

The contribution to clarification of flowering process in wheat was the main aim of the thesis. Flowering under optimal conditions is important for protection of sensitive flower organs during pollination and for subsequent process of grain development. This can contribute to higher yield and has a great importance for humankind. Identification of genes/alleles which are responsible for tuning of flowering time is a key for understanding flowering process and using of this knowledge for development of new, better adapted cultivars. The wild and domesticated tetraploid wheat can be valuable source of new alleles. Thus, another part of the work during doctoral studies had focused on characterization of tetraploid wheat on genotypic level. This information was used for determination of place of origin of domesticated emmer wheat and illumination of evolutionary relationship within domesticated tetraploid wheat.

5.1 I. Identification and characterization of new *Vrn-A1f-like* allele responsible for flowering lateness.

The first aim of the thesis was identification of the gene which is responsible for flowering time variation within doubled haploid (DH) lines mapping population. The parental lines of mapping population were elite bread wheat cultivar Tähti and line 8.1 (contains introgressions of *Triticum millitinae* in cv. Tähti background). Overall, seven loci were identified as an effective on flowering time adjustment. The most significant effect has been located on the chromosome 5A which was later identified as vernalization gene *VRN-A1*. The sequencing of this gene from both parental lines of mapping population showed that the delayed flowering is caused by new *Vrn-A1* allele originating from *Triticum militinae*. This allele has specific features which was not present in previously identified alleles – insertion of miniature inverted-repeat transposable element and deletion with unique boarders in first intron. It was named *Vrn-A1f-like* and was responsible for 70% of phenotypic variation. This work showed existence of variation in flowering time determination within group of different alleles *Vrn-A1* conferring spring habit.

5.2 II. Analysis of expression profile and methylation pattern of lines with three copy of *Ppd-B1a* gene as a possible cause of intragroup heading time variation.

The second part of the thesis employed identification of gene responsible for difference in flowering time between two bread wheat cultivars – Paragon and Kaerntner Frueher. The analysis of F_2 mapping population revealed the locus responsible for phenotypic variation which was located on the 2B chromosome in the region of photoperiodic gene (*Ppd-B1*). The higher number of copies *Ppd-B1a* allele in Kaerntner Frueher was directly responsible for flowering earliness. However, we identified interesting phenomenon which was outlined also in previous studies of *Ppd-B1a* allele. It is related to variation in flowering time within individuals with

three copies of *Ppd-B1a*. In our work, we tried to correlate this intragroup variation with differences in expression profile and subsequently with distinct methylation pattern.

6 LIST OF ABBREVIATIONS

AFLP– Amplified Fragment Length Polymorphism

ANOVA – ANalysis Of VAriance

BAC – Bacterial Artificial Chromosomes

BC – BackCross

bp – base pair

CAPS– Cleaved Amplified Polymorphic Sequence

CIM – Composite Interval Mapping

cM – centiMorgan

CRISPR– Clustered Regularly Interspaced Short Palindromic Repeats

DArT– Diversity Arrays Technology

DNA – DeoxyriboNucleic Acid

F_x – x Filial generation

GA – Gibberelic Acid

HDD – Host Directed Duplication

HR – Homologous Recombination

IRAP – Inter-Retrotransposon Amplified Polymorphism

ISBP – Insertion Site Binding Polymorphism

ISSR – Inter-Simple Sequence Repeat

KASP – Competitive Allele Specific PCR

LD – Long Day

LOD – Likelihood Of Odds

LRS – Likelihood Ratio Statistic

miRNA – micro RNA

MITE – Miniature Inverted-repeat Transposable Element

NCBI– National Center for Biotechnology Information

NHEJ– Non-Homologous End Joining

NILs – Nearly Isogenic Lines

nm – nanometers

PCR – Polymerase Chain Reaction

PPD – PhotoPerioDic genes

QTL – Quantitative Trait Loci

REMAP– REtrotrasposon-Microsatellite Amplified Polymorphism

RFLP– Restriction Fragment Length Polymorphism

RILs – Recombinant Inbred Lines

RNA – RiboNucleic Acid

SAGE– Serial Analysis of Gene Expression

SCAR– Sequence-Characterized Amplified Region

SD – Short Day

SFE – Spring Foldback Element

SIM – Simple Interval Mapping

SNP – Single Nucleotide Polymorphism

SSR– Simple Sequence Repeats

TALEN– Transcription Activator-Like Effector Nucleases

TILLING – Targeting Induced Local Lesions In Genomes

UTR – UnTranslated Region

VIGS– Virus Gene Silencing

VRN – VeRNalization genes

7 LIST OF APPENDICES

Original papers

APPENDIX I	Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*
APPENDIX II	Heritable heading time variation in wheat lines with the same number of Ppd-B1 gene copies
APPENDIX III	Kvitnúť či nekvitnúť (in Slovak)
APPENDIX IV	Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent
APPENDIX V	Taxonomical classification and origin of Kamut®

Published abstracts – poster presentation

APPENDIX VI	Different daylength – different effect
APPENDIX VII	Chasing a new flowering time gene
APPENDIX VIII	High-throughput gene resources and phenotyping - the key factors in fine mapping of new *Eps* gene
APPENDIX IX	Identification and characterization of a new *Vrn-A1f-like* allele responsible for flowering lateness in wheat

APPENDIX I

Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*

Ivaničová Z., Jakobson I., Reis D., Šafář J., Milec Z., Abrouk M., Doležel J., Järve K. and Valárik M.

New Biotechnology, 33 (5B), 718-727, 2015

New Biotechnology • Volume 33, Number 5 • September 2016

Research Paper

Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*

Zuzana Ivaničová[1], Irena Jakobson[2], Diana Reis[2], Jan Šafář[1], Zbyněk Milec[1], Michael Abrouk[1], Jaroslav Doležel[1], Kadri Järve[2] and Miroslav Valárik[1]

[1] Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-783 71 Olomouc-Holice, Czech Republic
[2] Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

Flowering time variation was identified within a mapping population of doubled haploid lines developed from a cross between the introgressive line 8.1 and spring bread wheat cv. Tähti. The line 8.1 carried introgressions from tetraploid *Triticum militinae* in the cv. Tähti genetic background on chromosomes 1A, 2A, 4A, 5A, 7A, 1B and 5B. The most significant QTL for the flowering time variation was identified within the introgressed region on chromosome 5A and its largest effect was associated with the *VRN-A1* locus, accounting for up to 70% of phenotypic variance. The allele of *T. militinae* origin was designated as *VRN-A1f-like*. The effect of the *VRN-A1f-like* allele was verified in two other mapping populations. QTL analysis identified that in cv. Tähti and cv. Mooni genetic background, *VRN-A1f-like* allele incurred a delay of 1.9–18.6 days in flowering time, depending on growing conditions. Sequence comparison of the *VRN-A1f-like* and *VRN-A1a* alleles from the parental lines of the mapping populations revealed major mutations in the promoter region as well as in the first intron, including insertion of a MITE element and a large deletion. The sequence variation allowed construction of specific diagnostic PCR markers for *VRN-A1f-like* allele determination. Identification and quantification of the effect of the *VRN-A1f-like* allele offers a useful tool for wheat breeding and for studying fine-scale regulation of flowering pathways in wheat.

Introduction

Bread wheat (*Triticum aestivum* L.) is a staple food for 40% of the human population [1] and is one of the three most important crops worldwide. It is an allohexaploid species with a large and complex genome (1C = 17 Gb, $2n = 6x = 42$, AABBDD genome) which is composed of more than 80% of repetitive elements [2]. It is a relatively young species and its emergence, spread and adaptation to a wide range of environmental conditions are closely connected to the development of settled human society. The fast growth of the human population and changes in diet preferences have significantly increased demands for higher wheat production. It was estimated that the annual yield increase should reach about 1.7% to satisfy the needs. However, the current increase is only about 1% a year [3].

Changing environmental conditions pose a challenge to breed varieties with high and stable yield under the pressure of biotic and abiotic stress. Furthermore, intensive breeding has led to a narrowed gene pool in current elite wheat cultivars [4]. An attractive source of new genes and alleles for wheat improvement are related species, wild wheat ancestors and landraces which have not experienced bottlenecks connected to the emergence of modern cultivars [5]. One of the most important traits affecting yield is flowering time. Optimal timing of flowering is critical not only for successful flower development, but also for grain filling and for

Corresponding author: Valárik, M. (valarik@ueb.cas.cz)

http://dx.doi.org/10.1016/j.nbt.2016.01.008

GLOSSARY

Vernalization The original term 'jarovizacia' by Lysenko was replaced by the currently used term 'vernalization' by Chouard [47]. Vernalization is defined as the importance of prolonged cold exposure for initiating the flowering process. On the basis of their vernalization requirement, bread wheat varieties are classified as winter, spring and facultative varieties. Winter wheat is sown in winter and needs cold exposure for flowering, spring wheat (sown in spring) will flower without any cold period. Facultative wheat varieties can be sown either in autumn or in spring with a safe transition to generative stage.

MITE transposable element Miniature inverted-repeat transposable elements (MITEs) are non-autonomous transposable elements often associated with genes. Unlike autonomous DNA transposons, the internal sequences of MITEs are short and devoid of coding sequences. MITE sequences are frequently transcribed with plant genes.

MADS box The name refers to four of the originally identified members: MCM1, AG, DEFA and SRF of the MADS box proteins. MADS box is a conserved motif found within DNA binding protein domains. It consists of 56 amino acids, 9 of which are identical in all MADS box family members described so far. The N-terminal part is the major determinant of DNA-binding specificity whilst the C-terminal half is necessary for dimerisation. The majority of MADS-box proteins bind similar sites based on the consensus sequence $CC(A/T)_6GG$, although each protein apparently possesses a distinct binding specificity. Moreover, several MADS-box proteins specifically recruit other transcription factors into multi-component regulatory complexes. Such interactions with other proteins appear to be a common theme within this family and play a pivotal role in the regulation of target genes.

survival of plants *sensu lato*, especially for annual species. The importance of this trait is expressed in the large number of genes involved in its regulation. However, only some of these genes have been cloned, well characterized and used in breeding (reviewed in [6]).

A complex genome and wide range of environmental conditions affecting wheat during its evolution have led to the development of a robust flowering regulation which allows adaptation to various macro- and micro-environmental conditions. Many genes working in a redundant manner are involved in the regulation of flowering and, at present, three key pathways are considered [7]. The vernalization pathway is responsible for avoiding flowering before or during the cold period (*VRN* genes). The photoperiodic pathway is responsible for perception of day length (*PPD* genes) and the earliness per se (*EPS*) genes are responsible for fine tuning of flowering time in reaction to specific climatic conditions. Even a minor variation in major photoperiod and vernalization genes is connected to large differences in flowering time and thus fast adaptability of wheat to diverse locations [8]. Currently, the vernalization pathway is the most characterized and its three major genes *VRN1*, *VRN2*, and *VRN3* have been cloned [9–11]. The *VRN3* gene is orthologous to *Arabidopsis thaliana FLOWERING LOCUS T* (*FT*) [11] and seems to be the major integrator of environmental signals from cold, photoperiod, and *EPS* genes in leaves [12]. The VRN3 protein is transported through phloem to apical meristem where it triggers transition from vegetative to generative stage by inducing the expression of the *VRN1* gene [13]. The *VRN2* gene, originally believed to be the vernalization gene [9], is the most likely mediator between vernalization and photoperiodic pathways and is involved in the repression of the *VRN3* gene [14]. The *VRN1* gene is homologous to *A. thaliana* flower meristem identity gene *APETALA1* (*AP1*) [15]. It seems that *VRN1* has pleiotropic function and besides affecting the floral meristem identity in apex, it is involved in the integration of cold signals in leaves. Cold treatment leads to *VRN1* upregulation which inhibits expression of *VRN2* and releases expression of *VRN3* [14]. Additionally, *VRN1* expression has been linked to altered frost tolerance [16] and yield traits regulation [17]. *VRN1* belongs to the family of MADS box transcription factors [14,18,19] with a CArG regulatory box in promoter region [10] and an additional regulatory region in the first intron [20]. Even though it has been determined that *VRN1* is not absolutely essential for flowering in polyploids [14], its altered expression can significantly affect the flowering pathway. The significance of this gene was demonstrated by the discovery that a mutation in the first intron or promoter of the gene leads to spring growth habit [19,20]. Flowering time modulation through allelic variation of the *VRN1* gene [10,13,19–22] is a promising tool for breeding new varieties better adapted to local environmental conditions.

Attractive sources of genetic variability are wild wheat ancestors and landraces. One of them is *Triticum militinae* (Zhuk. &Migush) ($2n = 4x = 28$, A^tA^tGG genome) which originated from a hybridization event separate from emmer wheats and belongs to the *T. timopheevii* (Zhuk.) Zhuk. group. Timopheevii wheats have been used as a source of new resistance genes: four genes for resistance to leaf rust (*LrTt1, LrTt2, Lr18* and *Lr50*), three genes for resistance to stem rust (*Sr36, Sr37*, and *Sr40*) and four genes for resistance to powdery mildew (*Pm6, Pm27*, and *Pm37*) [23] have been introduced into common wheat. Recently, an introgression line 8.1 was created from a cross of *T. militinae* and elite Finnish cv. Tähti [24]. In the line, *T. militinae* introgressions were identified on seven chromosomes (1AS, 2AS, 2AL, 4AL, 5AL, 7AS, 1BS, and 5BS), and improved resistance to powdery mildew in seedlings and adult plants was found to cosegregate with the locus *QPm-tut-4A* on chromosome 4A [25]. The *QPm-tut-4A* gene is a subject of cloning, however, the analyzed mapping population showed a significant variability in flowering time. In this study, we report identification, quantitative trait loci (QTL) mapping, verification and sequence characterization of a *VRN-A1* allele introgressed into bread wheat from *T. militinae* that affects flowering time of elite spring bread wheat varieties.

Material and methods

Plant material

A mapping population of 314 doubled haploid (DH) lines was developed from a cross between introgression spring line 8.1 and

spring *T. aestivum* cv. Tähti as described by Jakobson *et al.* [24,25]. The line 8.1 contains *T. militinae* introgressions on seven chromosomes (1AS, 2AS, 2AL, 4AL, 5AL, 7AS, 1BS, and 5BS). Thus, the DH lines vary in the presence of different *T. militinae* segments. Seeds of spring wheat *T. militinae* (accession no. K-46007) were obtained from the N.I. Vavilov Institute of Plant Industry (St. Petersburg, Russia). Two verification mapping populations (VMP1 and VMP2) were developed to estimate and confirm the influence of *T. militinae* introgressions on flowering time in different genetic backgrounds. To simplify *T. militinae* based traits segregation analysis, the DH81 line was derived by dihaploidization of the original introgression line 8.1. The DH81 line contains all the introgression segments except the 2AL and 1BS introgressions and introgression on 5A was slightly shortened. To specifically address the effect of *VRN1–VRN3* genes, the DH397 line was selected from the DH mapping population described above. In the DH397 line, introgressions on 1A, 2AL, and 7A chromosomes are not present and introgressions on 5A and 5B were shorter as compared to DH81. The introgression loss or shortening included loci of the *VRN1–VRN3* genes (data not shown). VMP1 comprises 194 F_2 lines derived from a cross of the doubled haploid lines DH81 and DH397 (98 F_2 plants with DH397 as the female parent, 96 F_2 plants with DH81 as the female parent). VMP2 was developed by crossing of the DH81 line with *T. aestivum* cv. Mooni as female and comprises 94 F_2 lines. Seeds of spring bread wheat cv. Mooni were provided by Estonian Crop Research Institute (Jõgeva, Estonia).

DNA extraction and PCR amplification

Genomic DNA was extracted from 100 mg of lyophilized and homogenized leaf tissue. Homogenization was done using Tissue Lyser (Qiagen, Germany) in 96-well deep well plates with two glass beads (5 mm) in each well for 4 min at 27 Hz. DNA was purified either using the NucleoSpin® 96 Plant II kit (Macherey-Nagel, Germany) according to manufacturer's instructions, or with a modified Agencourt® Genefind v2 kit (Beckman Coulter, USA). The Genefind v2 kit was used in 96-well plate format and pipetting was done by automatic pipetting workstations Biomek NX^P (Beckman Coulter, USA). Homogenized leaf tissue was incubated in 1 ml of Lysis Buffer [0.5% sodium bisulfite; 0.1% ascorbic acid; 1% β-mercaptoethanol; 0.5 M NaCl; 0.1 M Tris–HCl; 0.05 M EDTA, pH 7.2; and 2.1 U of RNase A (Sigma–Aldrich, USA)] at 65°C for 45 min. 10 μl of Genefind v2 magnetic beads (Beckman Coulter, Beverley, USA) was mixed with 100 μl of lysate and 80 μl of isopropanol and incubated for 10 min at room temperature to bind DNA to the beads. The mixture was incubated on 96 S Super Magnet Plate (Alpaqua, USA) for 3 min to settle the beads. Supernatant was removed and the beads were washed four times with 100 μl of 70% ethanol. DNA was diluted in 50 μl of redistilled water after evaporation of the residual ethanol. PCR consisted of template DNA (~10 ng), 1× PCR buffer (10 mM Tris–HCl, pH 8.8 at 25°C; 1.5 mM $MgCl_2$; 50 mM KCl; 0.1% Triton X-100); 200 mM dNTPs; 0.5 μM of each primer; 0.5 U of HOT FIREPol® DNA Polymerase (Solis BioDyne, Estonia). PCRs were performed in the C1000 Touch™ Thermal cycler (BIO-RAD, USA). Annealing temperatures of SSR markers followed information published at Graingenes 2.0 (http://wheat.pw.usda.gov/). Amplicons were separated either on 3.5%/6% non-denaturing polyacrylamide gel using Mega-gel apparatus (C.B.S. SCIENTIFIC, USA) or on 1.5% agarose gel stained with ethidium bromide or as described by Jakobson *et al.* [22] and visualized under UV light using the InGenius imaging system (Syngene, UK).

Genotyping and sequencing

Individual plants of mapping populations were genotyped using a set of 57 polymorphic markers (Supplementary file 1, Tab. S1). Since the original mapping population (line 8.1 × Tähti) was created as near-isogenic lines (NILs), the markers used were chosen with a focus on the introgressed regions, only. In addition, markers for major flowering time genes (*VRN1, VRN3* and *PPD-A1*) were added. *VRN-A1* was mapped with publicly available markers [26] and with VRN-1A_F4 and VRN-1A_R4 primers developed in this study (Supplementary file 1, Tab. S2). *PPD-A1* and *VRN3* were mapped by genotyping-by-sequencing with markers developed in this study (Supplementary file 1, Tab. S1 and S2). A subset of markers specific for introgressions were used to determine the genomic composition of VMP lines (Supplementary file 1, Tab. S1). To identify sequence variation of the *VRN-A1* gene, a set of 25 primer pairs (Supplementary file 1, Tab. S2) covering the whole length of the *VRN-A1* gene locus (~13,000 bp) and ~ 600 bp of the promoter sequence were designed and used for amplification in all three parental lines (cv. Tähti, cv. Mooni and DH81). *VRN-A1* specific primers were designed manually using publicly available sequences of the *VRN1* gene for wheat A, B, D genome and A genome of *T. timopheevi* (GenBank accession nos. AY747601.1, AY747604.1, AY747606.1, GQ451751). Additionally, cv. Chinese Spring chromosome-derived survey sequences were used [27]. All sequences were aligned using the Geneious 5.6.4 software (http://www.geneious.com) [28]. The designed primer pairs were considered specific when at least one primer had two or more genome specific SNPs/indels in its sequence. Melting temperature was determined by Primer-BLAST [29]. DNA amplification was carried out by touchdown PCR: initial denaturation for 15 min at 95°C followed by 14 cycles at 95°C for 40 s, 65°C for 40 s (increment −0.7°C/per cycle) and 72°C for 2 min. Twenty six more cycles were performed at 95°C for 40 s, 55°C for 40 s and 72°C for 2 min. Amplicons were separated as described above. Before sequencing, 5–20 ng of PCR products were purified in a mixture of 1 U Exonuclease I/0.5 U FastAP™ Thermosensitive Alkaline Phosphatase (Thermoscientific, USA) in 7 μl of 1× PCR buffer for 20 min at 37°C. Enzymes were inactivated by incubation for 30 min at 95°C.

The sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) and purified using the Agencourt Clean SEQ kit (Beckmann Coulter, USA). The reactions were analyzed on the ABI 3730xl DNA analyzer (Applied Biosystems, USA). The resultant sequences were trimmed and assembled using Geneious 5.6.4 (http://www.geneious.com) [28]. Assemblies were verified by alignment with the reference sequence of Triple Dirk D *VRN-A1* gene (AY747601.1). Regulatory elements were delimited according to publicly available sequences [30]. Transposable elements were identified using the TriAnnot pipeline [31]. Polymorphisms in the *VRN-A1* gene between parental lines of mapping populations used (cv. Tähti, cv. Mooni, and DH81) were identified by alignment using the Geneious 5.6.4 software (http://www.geneious.com) [28].

Phenotyping of developed mapping populations

Phenotyping experiments were performed during three seasons (2009/2010, 2013 and 2014) in three different locations (Jõgeva and Tallinn – Estonia, and Olomouc – the Czech Republic) in the field as well as under controlled environmental conditions. Each phenotype experiment was designed as follows:

*Experiment 1: evaluation of the effect of **T. militinae** introgressions on flowering time variation in the DH mapping population (cv. Tähti genetic background)*

To confirm the effect of the introgressions on flowering time, experiments were established in two different environmental conditions. Experiment 1a: the full mapping population of 314 DH lines was tested in the Czech Republic. Experiment 1b: a random subset of 113 DH lines was tested in Estonia. Each plant in the experiments was genotyped using the set of 55 markers (Supplementary file 1, Tab. S1).

1a. Olomouc, the Czech Republic, season 2013

Seed germination of 314 DH lines was synchronized by soaking the seeds in tap water for 24 hours at room temperature, followed by incubation of the seeds on wet filter paper for two days at 4°C and one day at 25°C in plastic Petri dishes. Four seedlings per line were planted in jiffy pots (6 cm × 6 cm) and grown in a greenhouse at 20°C under natural light. Seedlings at the stage of three leaves were transplanted to the field in a randomized manner. Flowering time (days to emergence of half of the spike on the main stem) of each individual plant was recorded daily at the same time of a day. Mean flowering time of individual lines was used for QTL mapping.

1b. Jõgeva, Estonia, winter 2009/2010

Six seeds per line of 113 DH lines, cv. Tähti and line 8.1 were planted directly into the soil in a plastic container (35 cm × 65 cm) and grown under controlled environmental conditions (16/8 hours day/night, light intensity 168 $\mu E\ m^{-2}\ s^{-1}$, high pressure sodium lamps, temperature 20–23°C) in a greenhouse. The day when at least half of spike of at least half of plants of every line emerged was considered to be the flowering time.

Experiment 2: assignment of the flowering time variation to particular introgressions (Olomouc, 2014)

Verification of introgression effect on flowering time was done using 194 F_2 plants of VMP1 derived from reciprocal crosses of lines DH397 and DH81. Seeds were germinated, plants were grown and flowering time was recorded as described in Experiment 1a. The presence of *T. militinae* introgressions was verified by genotyping using at least one marker per introgression (Supplementary file 1, Tab. S1).

*Experiment 3: verification of the effect of **T. militinae** introgressions on flowering time variation in cv. Mooni genetic background*

F_3 plants from cross of DH81 x Mooni (VMP2, 94 lines), developed in previous experiments (Jakobson and Järve, unpublished data), were employed using $F_{2:3}$ approach described at Austin and Lee [32]. F_3 plants from the VMP2 lines were screened for flowering time in two different locations. Experiments 3a and 3b were performed in Olomouc, the Czech Republic and Tallinn, Estonia, respectively. Mean flowering time of each line was used to reconstruct the phenotype of F_2 lines. A subset of 24 markers (Supplementary file 1, Tab. S1) was used for introgression segments mapping.

3a. Olomouc, the Czech Republic, season 2014

Ten F_3 plants from 80 lines of the VMP2 were evaluated for flowering time in field conditions in Olomouc. Seeds were germinated and grown as described in Experiment 1a. Flowering time data of 80 F_3 lines was collected. Every individual F_3 plant was genotyped for 14 markers (one to five markers for every introgression (Supplementary file 1, Tab. S1) to confirm identity to the original F_2 plant.

3b. Tallinn, Estonia, season 2014

Four F_3 plants from 94 families of the VMP2 were grown in a room with controlled environmental conditions at 22–23°C, 16/8 hours day/night, humidity 50–70%, fluorescent light with intensity 150 $\mu E\ m^{-2}\ s^{-1}$. The seeds were sown directly into 13 cm × 13 cm cm pots. Plants were not ordered, but grown in a randomized manner. For each of the F_3 plants, flowering time was recorded as described in Exp. 1b.

QTL mapping and statistical analysis

QTL analysis was performed using the statistic software Map Manager QTX Version b16 [33]. Genetic distances were established using the Kosambi mapping function. The threshold value for Likelihood ratio statistic (LRS) for suggestive ($P < 0.05$), significant ($P < 0.001$) and highly significant ($P < 0.0001$) loci was calculated for each experiment using permutation test at 1000 iterations. Single marker regression was used to associate phenotype with marker in each experiment. The position of detected QTLs was determined using simple interval mapping. The confidence interval was established by bootstrap analysis with 1000 iterations using the same software. Statistical significance was determined using single factor analysis of variance (ANOVA).

Results

Identification of QTLs controlling flowering time

Using gene-specific primers (Supplementary file 1, Tab. S2) and genotyping-by-sequencing, *VRN-A1*, *VRN3* and *PPD-A1* genes were mapped within the *T. militinae* introgressions in line 8.1. *VRN-A1* was located 16.2 cM distal to *Xcfd30* and 26 cM proximal to *Xpsp3001* on chromosome 5A (Fig. 1). *VRN3* was positioned 8.8 cM distal to *Xwmc283* on chromosome 7A. *PPD-A1* was mapped in close proximity to *Xgwm71.1* loci (4.1 cM distant) on chromosome 2A (data not shown). In total, 314 and 113 DH lines generated from the progeny of the cross between line 8.1 and cv. Tähti were screened for flowering time in Olomouc (Experiment 1a) and Jõgeva (Experiment 1b). The core population analyzed in Jõgeva was included in the population tested in Olomouc. A moderate phenotypic correlation ($r = 0.46$, $P < 0.001$) between the two experiments was observed. Flowering time of line 8.1 and cv. Tähti differed by 3.5 days (89.3 versus 92.8) in Olomouc and by 5 days in Jõgeva (72 versus 77 days). In both experiments, line 8.1 flowered later than cv. Tähti. Flowering time of individual plants in mapping populations ranged from 81 to 98 and from 65 to 87 days in

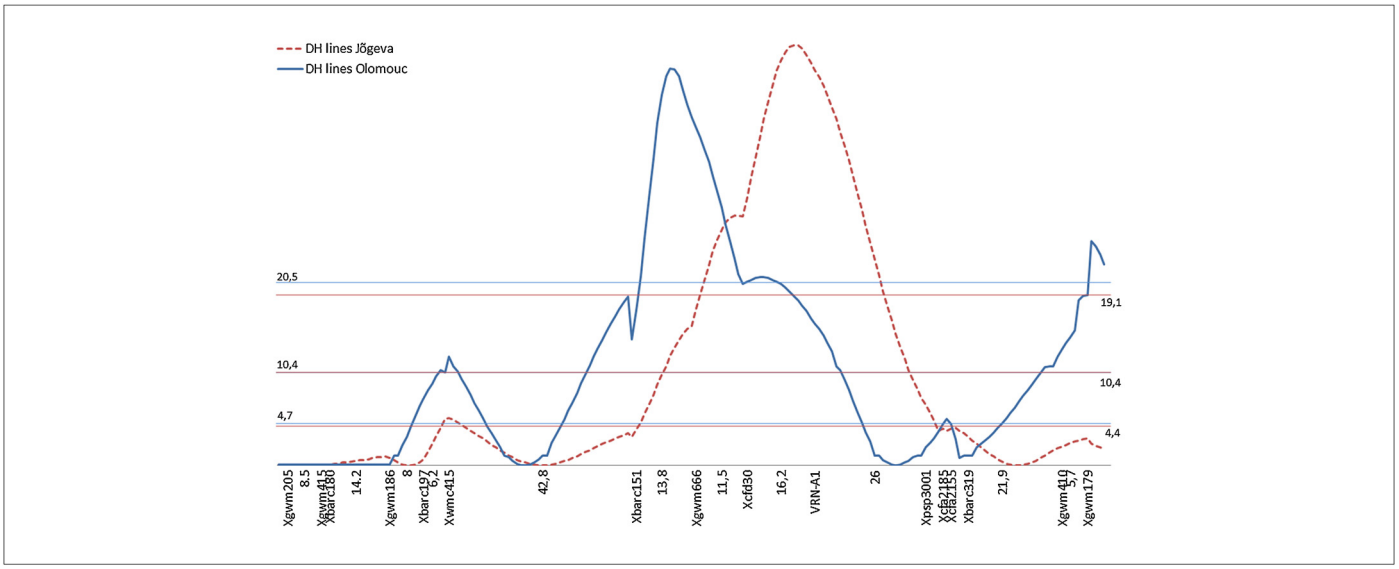


FIGURE 1

Location of the flowering time QTL on chromosome 5A. Simple interval mapping analyses of the DH mapping population based on heading time data from Experiment 1a (introgression line 8.1 × cv. Tähti, Olomouc, solid line) and Experiment 1b (introgression line 8.1 × cv. Tähti, Jõgeva, dashed line) localised the flowering time QTL in the region between markers *Xbarc151* and *Xcfa2155*, with the peak at the *Xgwm666/VRN-A1* locus. The suggestive threshold line (likelihood ratio statistic LRS: 4.7/4.4), the significant threshold line (LRS: 10.4/10.4) and highly significant threshold line (LRS: 20.5/19.1) are displayed for Experiments 1a/1b, respectively. Threshold values are shown on the left (Experiment 1a) and on the right side (Experiment 1b) of the graph. The threshold lines of respective QTL lines are marked with the same colour. The distance in cM between markers is shown on *x* axis.

Experiment 1a and Experiment 1b, respectively (Table 1). Marker regression analysis showed that at least 29 markers were associated with phenotypic variation on a suggestive level in at least one experiment (Supplementary file 1, Tab. S3). Significant QTLs were identified on five chromosomes: 1A, 2A, 4A, 5A and 7A; however, highly significant loci responsible for variation in at least one experiment were determined only on chromosomes 5A and 7A. The QTL localised on 7A, delimited by the *Xbarc70–Xpsp3050* markers and peaking at *Xwmc283* was associated with *VRN3* [34,35] and was significant only in Olomouc (Supplementary file 1, Tab. S3). The QTL localised on 5A detected in both experiments was delimited by the *Xbarc151–Xcfa2155* markers and overlapped with the *VRN-A1* gene locus (Fig. 1). The *T. militinae* allele of the *VRN-A1* gene accounted for 1.9 and 5.8 days delay in flowering time in Experiment 1a and 1b, respectively. Large differences in the number of significant QTLs and flowering time suggested additional environmental and genetic influences in Experiments 1a and 1b, which necessitated further verification.

Confirmation of flowering time difference using verification mapping populations (VMPs)

Quantitative trait loci responsible for phenotypic variance in the primary mapping population often do not operate in the same manner in a different genetic background [36]. To verify loci identified in previous experiments, VMP1 and VMP2 were created and used. A difference of 6.5 days ($P < 0.001$) was identified between the parental lines of VMP1 (DH397 × DH81) in Experiment 2 (Tab. 1). Marker regression analysis showed highly significant contribution of the *VRN-A1* locus on chromosome 5A (Supplementary file 1, Tab. S4) to flowering time variation.

TABLE 1

Statistical analysis of phenotyping data.Phenotyping experiments were performed under controlled environmental conditions (greenhouse; 1b, 3b), or in field conditions (1a, 2, 3a). The experiment 1 was conducted on the DH mapping population generated from the cross between the introgression line 8.1 and cv. Tähti; the experiment 2 was performed on VMP1 population (DH397 x DH81) and experiment 3 was performed on VMP2 population (Mooni x DH81). Data are given for all lines of the mapping populations. FT - mean heading time in days. MIN - minimum value; MAX - maximum value; SD - standard deviation, *VRN1-A1f-like* contribution - contribution of the *VRN1-A1f*-like allele to the phenotype variation, FT difference between parental lines. Note that in all experiments the *T. militinae* allele caused a delay in flowering time. All parameters are in days.

Experiment	Parental lines	Background	Location	Greenhouse/ field	FT	MIN	MAX	SD	VRN1-A1f-like contribution	FT difference between parents
1a	8.1/Tähti	Tähti	O	f	90	81	98	3	+1.9	+3.5
1b	8.1/Tähti	Tähti	J	g	76	65	87	5	+5.8	+5. 5
2	DH397/DH81	Tähti	O	f	86	82	94	2	+1.7	+6.5
3a	DH81/Mooni	Mooni	O	f	84	78	90	2	+3.3	+9.8
3b	DH81/Mooni	Mooni	T	g	66	52	89	9	+18.6	+29.9

g-greenhouse; f – field; O – Olomouc; T – Tallinn; J - Jõgeva

The locus was delimited by markers *Xgwm639-Xcfd2155* (Supplementary file 2, Fig. S1) and responsible for 9% of the trait variation. The effect of replacing the Tähti allele with the *T. militinae* allele in the *VRN-A1* locus was almost the same as in Experiment 1a. The phenotypic analysis of the parental genotypes of VMP2 (DH81 × cv. Mooni) revealed an approximately 10 days and 30 days difference in flowering time in Experiment 3a and Experiment 3b, respectively (Tab.1). A moderate phenotypic correlation ($r = 0.63$, $P < 0.001$) was detected between these two experiments. Flowering time analysis of the F_3 generation was used for reconstruction of the F_2 phenotype followed by QTL analysis. The analysis identified the same loci for both experimental locations (Supplementary file 1, Tab. S5; Supplementary file 2, Fig. S5). The QTL with highest significance was located on chromosome 5A in the region delimited by markers *Xgwm666–Xgwm982*, with the peak in the position of the *VRN-A1* gene (Supplementary file 2, Fig. S2). The *VRN-A1* locus was responsible for 38% and 70% of flowering time variation in Experiment 3a and Experiment 3b, respectively (Supplementary file 1, Tab. S5; Supplementary file 2, Fig. S5). In different locations, the *T. militinae VRN-A1* allele caused a significant delay in flowering time in both backgrounds. However, the trait variance in tested populations was highly affected by growth conditions (greenhouse/field), and to a lesser extent, by the background of the *VRN-A1* allele.

Sequence analysis of VRN-A1 *alleles*

QTL analyses of the DH mapping population as well as VMPs indicated that different *VRN-A1* alleles were responsible for the major part of flowering time variation, and therefore the variability of the gene sequence in the parental lines (cv. Tähti, DH81 and cv. Mooni) was examined. The gene was amplified using 25 primer pairs (Supplementary file 1, Tab. S2) that cover the whole length of the *VRN-A1* gene (~13,000 bp), including approximately 600 bp of the promoter sequence. Comparison of the obtained sequences with the sequence of *VRN-A1* allele of the Triple Drk D (TDD) (GenBank accession no. AY747601.1), which was taken as reference, confirmed the completeness of the gene sequences. The sequences of *VRN-A1* alleles (cv. Mooni, cv. Tähti and DH81) have been deposited in the NCBI database (as KT696535, KT696536, KT696537, respectively) and contain the whole gene sequence, including the 340 bp and 625 bp segments of the promoter region in line DH81 and cvs. Tähti and Mooni, respectively, and a 276 bp segment downstream of the stop codon. The acquired sequences were aligned with the sequence of TDD and three types of mutations were identified which could be classified as small indels (1–10 bp), large indels (>10 bp), and SNPs. The promoter region of the *T. militinae VRN-A1* gene (from line DH81) was highly similar to the previously published *VRN-A1f* promoter (GQ451751) from *Triticum timopheevii* [37]. Only one 1 bp indel difference was identified (Supplementary file 2, Fig. S3) between these sequences in the HDD region as defined in [19] and hence the *T. militinae VRN1* allele from line DH81 was named *VRN-A1f-like*. No sequence variation in the *VRN-A1* locus was identified between the wheat cultivars Tähti and Mooni and they differ from TDD only by two single-base indels in the first intron and one SNP in intron 2 (Supplementary file 2, Fig. S4). On the other hand, *VRN-A1f-like* showed 38 SNPs, 6 small indels and 3 large indels as compared to TDD. In the promoter region, 5 SNPs were identified accompanied by a small (8 bp) deletion in the position −128 bp, and a large (50 bp) deletion in the position −63 bp, both from the start codon, compared to TDD (Fig. 2; Supplementary file 2, Fig. S3). In contrast to TDD, the Spring foldback transposable element (SFE) insertion [19] is absent in the *T. militinae VRN-A1f-like* allele (Fig. 2; Supplementary file 2, Fig. S3). None of these mutations is present in the known regulatory regions (CArG box, TATA box) of the *VRN-A1* alleles.

A majority of remaining mutations were identified in the first intron. The mutations included 24 SNPs, 2 small indels, a large deletion (2753 bp) and a large insertion (433 bp) of the miniature inverted-repeats transposable element (MITE) (Fig. 3; Supplementary file 2, Fig. S4). The newly identified MITE consisted of 424 bp including 53 bp inverted repeats at the ends and 9 bp host duplication (TAAAAAATA). The sequence was found to be 98% identical to DXX-MITE_3bSeqIt2-B-G2821-Map7. Six insertions of MITE

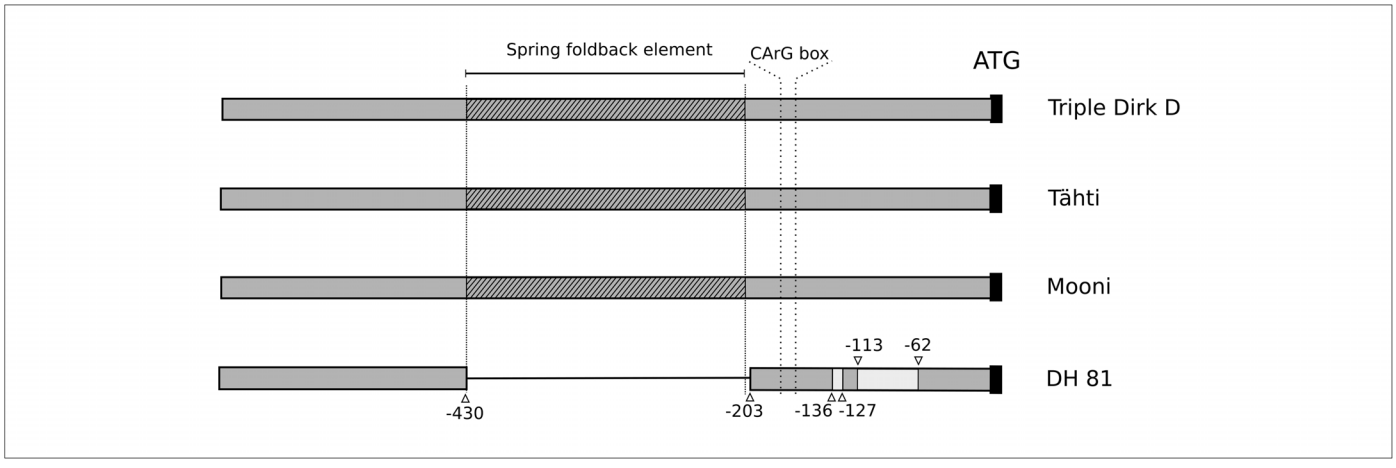


FIGURE 2

Comparison of the variation in VRN-A1 promoter region among parental lines. Sequencing the promoter region in parental lines of mapping populations showed sequence variation of *T. militinae VRN-A1f-like* allele from DH81 line in comparison with reference sequence of *VRN-A1a* promotor of Triple dirk D (GenBank accession no. AY747601.1). The differences comprised the absence of Spring fold element (SFE) insertion and two deletions (8 bp and 50 bp) located downstream of the CArG box. The exact position of indels is marked by white triangles and numbers represent flanking bases counted upstream from the START codon. Light grey boxes represent deletions, and the box with diagonal stripes represents the insertion of SFE. For sequence details see alignment (Supplementary file 2, Fig S3).

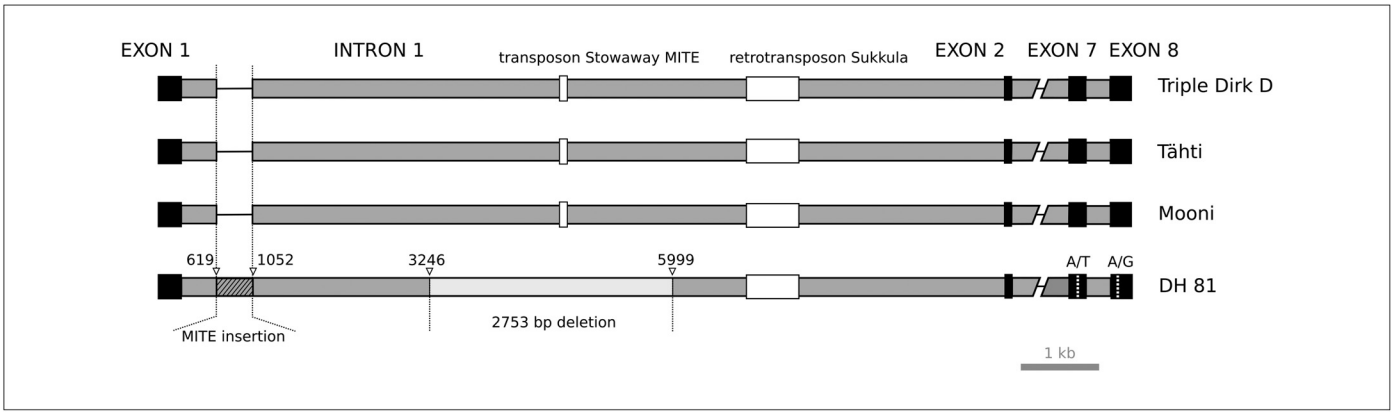


FIGURE 3

Comparison of the variation in *VRN-A1* allele among parental lines. Sequencing the *VRN-A1* gene from parental lines of mapping populations, and comparisons with the reference sequence of the *VRN-A1a* gene of Triple dirk D (AY747601.1) revealed insertion of MITE transposon (424 bp) with (9 bp) host duplication and large deletion (2753 bp) within the first intron of the *VRN-A1f-like* allele from DH81 line. Moreover, two synonymous SNPs were found in exon 7 and exon 8. The positions of indels are marked by white inverted triangles. Light grey box represents deletion; black boxes represent exons; the box with diagonal stripes represents MITE insertion and white boxes represent known retroelement insertions. Dashed vertical line shows single nucleotide substitution in exons only.

were identified on wheat chromosome 3B using the TRIANNOT pipeline [31]. In all six cases, the host duplications resemble TATA-box sequence. The identified MITE is homologous to four sequences in barley deposited at NCBI (www.ncbi.nlm.nih.gov) where it has been detected in close proximity to or directly associated with genes (either promoters or introns, data not shown).

In addition, six and one SNPs were identified in introns 2 and 4, respectively. In introns 4 and 8, the remaining 2 and 1 small indels were found, respectively. Of all the exons, only exons 7 and 8 had one SNP each. The SNP in exon 7 causes no amino acid change. The SNP in exon 8 causes a neutral amino acid substitution (A**A**C→A**G**C; ASN→SER; Fig. 3; Supplementary file 2, Fig. S4). Besides sequence polymorphisms, differences in amplicon size can be used as markers. Ten of the 25 primer pairs can be used as markers without sequencing (Supplementary file 1, Tab. S2). Three of them (VRN1AF/VRN_R1_2a, VRN1A_F3/R3 and F4/R4) are codominant markers showing size difference between the *VRN-A1a* and *VRN-A1f-like* alleles. Another five markers (VRN1_F2a/R2, F6/R6, F7/R7, F8/R8, and F9/R9a) are dominant for cvs. Tähti and Mooni, and two (VRN1_F2/R2 and F6/R9a) are dominant for *T. militinae*. The VRN1A_primer pairs F3/R3 and F4/R4 detect a 433 bp size variation caused by insertion of the MITE to the first intron of the *VRN-A1f-like* allele. VRN1 markers _F2a/R2 and F2/R2 detect the presence/absence of the 50 bp deletion in the promoter of the *VRN-A1f-like* allele. The remaining dominant markers detect the 2753 bp deletion in the first intron of the *VRN-A1f-like* allele (Supplementary file 1, Tab. S2).

Discussion

Allelic variation of vernalization (*VRN)* and photoperiod (*PPD)* genes is considered to be the main source of flowering time plasticity which enables wheat to grow and successfully flower at different altitudes and latitudes. The variability in flowering time is of high economic importance as it affects wheat reproductive success and in elite breeding material strongly affects yield components. Variability in the *VRN* and *PPD* genes can be defined either as copy number variation [38,39] or mutations in coding regions, promoters, or regulatory elements [19,20]. Extensive studies in cultivated and wild wheats were performed to characterize mutations in the promoter region or the first intron of the *VRN1* gene and to identify new *VRN1* alleles [13,19,30,37,40]. Moreover, the contribution of *VRN1* homoeologous alleles in hexaploid wheat and their effect on vigor and yield components has been estimated.

Tetraploid spring wheat *T. militinae* belongs to the Timopheevi group of wheat originating from a hybridization event separate from emmer wheats [41]. *T. militinae* is a promising source of new alleles useful in breeding [24], including new functional alleles of genes affecting flowering time. Diagnostic markers specific for these alleles are a valuable addition to the toolkit available for optimizing flowering time. Here we report interspecific introgression of *T. militinae* alleles of vernalization genes *VRN1, VRN3* and possibly *VRN2*, as well as the photoperiod-related gene *PPD1*, into the genome of spring wheat cv. Tähti (line 8.1). Using gene-specific primers and genotyping-by-sequencing, *VRN-A1, VRN-A3* and *PPD-A1* genes were mapped in the line 8.1 within the regions of *T. militinae* introgressions on chromosomes 5A, 7A and 2A, respectively. Furthermore, in at least one experiment, QTL analysis detected significant QTLs in the regions of these genes, indicating that the *T. militinae* alleles of *VRN-A1, VRN3, PPD-A1*, and possibly also *VRN-A2* may affect flowering time in *T. aestivum.*

Flowering time QTL analysis identified several minor QTLs at a significant level. A highly significant major QTL was identified in the locus of *VRN-A1* gene on chromosome 5A. The effect of the major QTL was confirmed in two additional mapping populations and, depending on genetic background and environmental conditions, it explained 5–70% of phenotype variability (Fig. 1, Supplementary file 1, Tab. S3–5). The *T. militinae VRN-A1* allele was designated *VRN-A1f-like* as its promoter region resembles the *VRN-A1f* allele from *T. timopheevii* [37]. The replacement of the Tähti *VRN-A1* allele with the *VRN-A1f-like* allele in the DH mapping population delayed flowering time by 5.8 and 1.9 days under controlled and field conditions, respectively. A similar effect was observed during subsequent verification of the effect (Experiment 2), where sister doubled haploid introgression lines DH81

and DH397 (VMP1) differed in flowering time by 6.5 days ($P < 0.001$) (Tab. 1). Line DH81 carries the *VRN-A1f-like* allele while DH397 does not, and the allele contributed to flowering time delay by 1.7 days (Tab. 1; Supplementary file 1, Tab. S4). Replacement of the allele in cv. Mooni caused an 18.6 and a 3.3 day flowering time delay under controlled and field conditions, respectively. Thus, the *VRN-A1f-like* allele can cause a delay of approximately 3 days in flowering time in field conditions. The approximately 3–6 fold increase in flowering time delay in controlled environmental conditions as compared to the field conditions could be explained by the effect of lower light quality and quantity, and/or limited root space. The controlled environmental conditions also caused slight alteration from normal distribution of flowering time within lines of mapping populations as compared to field conditions (Supplementary file 2, Fig. S5). Significant alteration of phenotype between field and greenhouse conditions has been described before, but the causes of this phenomenon remain largely unknown [42–44].

Nevertheless, a stable difference in flowering time due to the *VRN-A1* allele in spring wheat cultivars was not described before. The fact that no recombination suppression in the *VRN-A1f-like* locus was observed and that the *VRN-A1f-like* allele provides difference in flowering time in different genetic backgrounds of spring wheat, could be used as additional source of flowering time variation in breeding with negligible linkage drag. Such phenotypic variability is often associated with alteration in gene sequence, which facilitates development of markers tightly linked with to the trait and suitable for marker assisted selection.

To identify the possible causes of phenotypic variation on DNA sequence level, *VRN-A1* genes of *T. militinae* (in DH81), cv. Tähti and Mooni were sequenced. The promoter region of *T. militinae VRN-A1f-like* allele was found almost identical to the previously described *VRN-A1f* allele from *T. timopheevii* [37]. The only difference between the two sequences was one 1 bp indel in the HDD region of the *VRN-A1* promoter (Supplementary file 2, Fig. S3). However, the incomplete sequence of the *VRN-A1f* allele did not permit precise sequence comparison to confirm that these two alleles were the same. Comparison of the *VRN-A1* promoter region from spring wheat cvs. Mooni and Tähti using TDD as a reference revealed that cvs. Tähti and Mooni carry a *VRN-A1a* promoter region identical to that of the reference sequence. The *T. militinae VRN-A1f-like* allele differed mainly by three indels, one small and two large (Fig. 2). The main difference in the *VRN-A1f-like* allele was an absence of a Spring foldback element (SFE) insertion, which was observed in the promoter of *VRN-A1a* in spring wheat cultivars [19]. The element was also not found in the *VRN-A1b–VRN-A1h* alleles of hexaploid wheat [19,37]. A similar situation was also observed in related species *T. monococcum*, *T. boeticum*, and *T. urartu* [13] and *T. dicoccoides* [21]. In the *VRN-A1a* allele, the SFE insertion is often accompanied by promoter duplication [19]. Additionally, duplication of the whole *VRN1* gene causing a delay in flowering has been reported [30,39], however, no obvious duplication in the *VRN-A1f-like* allele was observed in the present study. The second large indel was a 50 bp deletion close to the start codon (Fig. 2). The deletion was reported only in the promoter of the *VRN-A1f* allele [37]. The small indel of 8 bp found between this deletion and the CArG-box (Fig. 2, Supplementary file 2, Fig. S3) is more frequent and common for the promoters of the *VRN-A1f, g, h, VRN-B1a* and *VRN-G1a* alleles [37]. On the basis of currently available data, it seems that the 50 bp deletion is only found in *T. timopheevii* group.

Comparison of the *VRN-A1* exon and intron regions from spring wheat cvs. Mooni and Tähti to TDD as the reference revealed that cvs. Tähti and Mooni are highly similar to the reference and the only differences in intronic regions involved two single base indels and one SNP. These mutations are outside the known regulatory regions and probably have no effect on the *VRN-A1a* gene expression. The *VRN-A1f-like* sequence differs from cvs. Mooni, Tähti, and TDD more significantly. The majority of the variability comprising 24 SNPs, 2 small indels, one large deletion (2753 bp) and one large insertion (433 bp) located in the first intron. This observation is not surprising since the first intron represents 64.4% of the gene. Only two SNPs were found in exon sequence: one causing no amino acid change and one causing a neutral amino acid change. The major mutation of the first intron comprised a 2753 bp deletion (Fig. 3, Supplementary file 2, Fig. S4). The second large indel was insertion of a MITE transposon in 5′ direction from the deletion (Fig. 3, Supplementary file 2, Fig. S4). The MITE has not been described before, however its homologue (98% identical) was identified on wheat chromosome 3B and designated as DXX-MITE_3bSeqIt2-B-G2821-Map7 [45]. The insertion of a repetitive element close to the 5′ end of the first intron was described in the *VRN-D1a* allele [46]. However, the effect of such an insertion on gene expression is unknown and requires further examination.

Despite the fact that the introgression line 8.1 flowered later than cvs. Tähti and Mooni in all experiments, not all *T. militinae* alleles in the detected QTLs increased flowering time. While the *VRN-A1f-like* allele as well as (to a lesser extent) the *T. militinae* allele of *VRN3* delayed flowering in the common wheat background, the *T. militinae* alleles of *PPD-A1* as well as the significant QTLs flanking the major QTL of *VRN-A1* on chromosome 5A (Fig. 1), showed an opposite effect, accelerating flowering by 1-2 days each (Supplementary file 1, Tab. S3). However, the minor QTLs accelerating flowering were only detected in the DH population. Further investigations involving larger populations and/or other genetic backgrounds are needed to prove these QTLs which, if verified, could find application in wheat breeding.

Conclusions

QTL analysis of a doubled haploid mapping population led to the identification of five loci originating from *T. militinae* and affecting flowering time variation. The QTL with the most significant effect was located on chromosome 5A in the region of *VRN-A1* gene. Sequence analysis of the *T. militinae VRN-A1* allele (designated as *VRN-A1f-like*) revealed significant variation in the promoter and regulatory regions of the first intron. Moreover, a new insertion of a transposable element MITE to the first intron was identified. The phenotypic effect of the allele was tested and verified in two elite spring cultivars of bread wheat, and in both backgrounds, the *VRN-A1f-like* allele significantly delayed flowering. Moreover, significant variation in flowering time of spring cultivars caused by *VRN-A1* allele has not been reported before. This new finding opens additional way to manipulate flowering time and breed cultivars better adapted to specific environmental niches. Identification of the novel *VRN-A1f-like* allele with the MITE insertion

will allow future studies to quantify its effect on fine tuning of the flowering time. The specific set of eleven PCR markers developed here for the *VRN-A1f-like* allele will enable its tracking in future studies and broadens the toolkit of diagnostic molecular assays available for *VRN1* alleles in temperate cereal species. Characterization of the variation in flowering time loci together with improved knowledge of molecular mechanisms controlling flowering time will facilitate breeding improved varieties better adapted to changing environmental conditions and human needs.

Author's contributions

Z.I. carried out genotyping and phenotyping in Olomouc, sequencing and sequence processing and drafted the manuscript. I.J. performed genotyping, development of mapping populations, QTL and statistical analysis, participated in study design and drafted the manuscript. D.R. carried out phenotyping and genotyping in Tallinn. J.S. contributed to design of the study and drafting of the manuscript. Z.M. participated in drafting the manuscript and sequence processing. M.A. contributed to sequence analysis annotation and bioinformatics analysis. J.D. participated in the design and coordination of study and preparation of the manuscript. K.J. participated in the design and coordination of the study. M.V. conceived the study participated in its design and coordination and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work has been supported by the Operational Program Education for Competitiveness — European Social Fund (project CZ.1.07/2.3.00/20.0165), the award LO1204 from the National Program of Sustainability I, by the Czech Science Foundation (grant award 14-07164S), by the Czech Ministry of Education, Youth and Sports (OP VK CZ.1.07/2.3.00/20.0165) and by the Estonian Ministry of Agriculture.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.nbt.2016.01.008.

References

[1] Bushuk W. Wheat breeding for end-product use. Euphytica 1998;100:137–45.
[2] Devos KM, Doležel J, Feuillet C. Genome organization and comparative genomics. In: Carver BF, editor. Wheat sci. trade. Oxford: Wiley–Blackwell; 2009. p. 327–67. http://dx.doi.org/10.1002/9780813818832.ch15.
[3] Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL, Chalhoub B, et al. A workshop report on wheat genome sequencing: international genome research on wheat consortium. Genetics 2004;168:1087–96. http://dx.doi.org/10.1534/genetics.104.034769.
[4] Feuillet C, Langridge P, Waugh R. Cereal breeding takes a walk on the wild side. Trends Genet 2008;24:24–32. http://dx.doi.org/10.1016/j.tig.2007.11.001.
[5] Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci U S A 2013;110:8057–62. http://dx.doi.org/10.1073/pnas.1217133110.
[6] Wang M, Wang S, Xia G. From genome to gene: a new epoch for wheat research? Trends Plant Sci 2015;1–8. http://dx.doi.org/10.1016/j.tplants.2015.03.010.
[7] Milec Z, Valárik M, Bartoš J, Šafář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv 2014;32:200–14. http://dx.doi.org/10.1016/j.biotechadv.2013.09.008.
[8] Gororo N. Photoperiod and vernalization responses in *Triticum turgidum* × *T. tauschii* synthetic hexaploid wheats. Ann Bot 2001;88:947–52. http://dx.doi.org/10.1006/anbo.2001.1531.
[9] Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 2004;303:1640–4. http://dx.doi.org/10.1126/science.1094305.
[10] Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci U S A 2003;100:6263–8. http://dx.doi.org/10.1073/pnas.0937399100.
[11] Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, et al. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci U S A 2006;103:19581–86. http://dx.doi.org/10.1073/pnas.0607142103.
[12] Li C, Distelfeld A, Comis A, Dubcovsky J. Wheat flowering repressor VRN2 and promoter CO_2 compete for interactions with NUCLEAR FACTOR-Y complexes. Plant J 2011;67:763–73. http://dx.doi.org/10.1111/j.1365-313X.2011.04630.x.
[13] Dubcovsky J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L. Effect of photoperiod on the regulation of wheat vernalization genes VRN1 and VRN2. Plant Mol Biol 2006;60:469–80. http://dx.doi.org/10.1007/s11103-005-4814-2.
[14] Chen A, Dubcovsky J. Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. PLoS Genet 2012;8. http://dx.doi.org/10.1371/journal.pgen.1003134.
[15] Li C, Dubcovsky J. Wheat FT protein regulates VRN1 transcription through interactions with FDL2. Plant J 2008;55:543–54. http://dx.doi.org/10.1111/j.1365-313X.2008.03526.x.
[16] Dhillon T, Pearce SP, Stockinger EJ, Distelfeld A, Li C, Knox AK, et al. Regulation of freezing tolerance and flowering in temperate cereals: the VRN-1 connection. Plant Physiol 2010;153:1846–58. http://dx.doi.org/10.1104/pp.110.159079.
[17] Kamran A, Iqbal M, Spaner D. Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. Euphytica 2014;197:1–26. http://dx.doi.org/10.1007/s10681-014-1075-7.
[18] Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J. Regulation of VRN-1 vernalization genes in normal and transgenic polyploid wheat. Plant Physiol 2005;138:2364–73. http://dx.doi.org/10.1104/pp.105.064287.
[19] Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J. Allelic variation at the VRN-1 promoter region in polyploid wheat. Theor Appl Genet 2004;109:1677–86. http://dx.doi.org/10.1007/s00122-004-1796-4.
[20] Fu D, Szucs P, Yan L, Helguera M, Skinner JS, Von Zitzewitz J, et al. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Mol Genet Genomics 2005;273:54–65. http://dx.doi.org/10.1007/s00438-004-1095-4.
[21] Shcherban AB, Strygina KV, Salina Ea. VRN-1 gene- associated prerequisites of spring growth habit in wild tetraploid wheat *T. dicoccoides* and the diploid A genome species. BMC Plant Biol 2015;15:1–13. http://dx.doi.org/10.1186/s12870-015-0473-x.
[22] Santra DK, Santra M, Allan RE, Campbell KG, Kidwell KK. Genetic and molecular characterization of vernalization genes Vrn-A1, Vrn-B1, and Vrn-D1 in spring wheat germplasm from the pacific northwest region of the U.S.A. Plant Breed 2009;128:576–84. http://dx.doi.org/10.1111/j.1439-0523.2009.01681.x.
[23] Leonova IN, Budashkina EB, Kalinina NP, Salina EA. *Triticum aestivum – Triticum timopheevii* introgression lines as a source of pathogen resistance genes. Czech J Genet Plant Breed 2011;49–55.
[24] Jakobson I, Peusha H, Timofejeva L, Järve K. Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* × *Triticum militinae* hybrid line. Theor Appl Genet 2006;112:760–9. http://dx.doi.org/10.1007/s00122-005-0181-2.
[25] Jakobson I, Reis D, Tiidema A, Peusha H, Timofejeva L, Valárik M, et al. Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat–*Triticum militinae* introgression line. Theor Appl Genet 2012;125:609–23. http://dx.doi.org/10.1007/s00122-012-1856-0.
[26] Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH. Allelic variation at the vernalization genes Vrn-A1, Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. Crop Sci 2008;48:458–70. http://dx.doi.org/10.2135/cropsci2007.06.0355.
[27] IWGSC. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 2014;345:1251788. http://dx.doi.org/10.1126/science.1251788.
[28] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012;28:1647–9. http://dx.doi.org/10.1093/bioinformatics/bts199.
[29] Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012;13:134. http://dx.doi.org/10.1186/1471-2105-13-134.
[30] Muterko a, Balashova F, Fayt Ia, Sivolap VIYM. Molecular-genetic mechanisms of regulation of growth habit in wheat. Cytol Genet 2015;49:58–71. http://dx.doi.org/10.3103/S0095452715010089.
[31] Leroy P, Guilhot N, Sakai H, Bernard A, Choulet F, Theil S, et al. TriAnnot: a versatile and high performance pipeline for the automated annotation of plant genomes. Front Plant Sci 2012;3:1–14. http://dx.doi.org/10.3389/fpls.2012.00005.
[32] Austin DF, Lee M. Comparative mapping in F2:3 and F6:7 generations of quantitative trait loci for grain yield and yield components in maize. Theor Appl Genet 1996;92:817–26. http://dx.doi.org/10.1007/BF00221893.
[33] Manly KF, Cudmore RH, Meer JM. Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome 2001;12:930–2. http://dx.doi.org/10.1007/s00335-001-1016-3.

[34] Somers DJ, Isaac P, Edwards K. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 2004;109:1105–14. http://dx.doi.org/10.1007/s00122-004-1740-7.
[35] Bonnin I, Rousset M, Madur D, Sourdille P, Dupuits C, Brunel D, et al. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. Theor Appl Genet 2008;116:383–94. http://dx.doi.org/10.1007/s00122-007-0676-0.
[36] Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 2005;142:169–96. http://dx.doi.org/10.1007/s10681-005-1681-5.
[37] Golovnina Ka, Kondratenko EY, Blinov AG, Goncharov NP. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. BMC Plant Biol 2010;10:168. http://dx.doi.org/10.1186/1471-2229-10-168.
[38] Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL. Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genet 2015;16:96. http://dx.doi.org/10.1186/s12863-015-0258-0.
[39] Díaz A, Zikhali M, Turner AS, Isaac P, Laurie Da. Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 2012;7. http://dx.doi.org/10.1371/journal.pone.0033234.
[40] Milec Z, Tomková L, Sumíková T, Pánková K. A new multiplex PCR test for the determination of Vrn-B1 alleles in bread wheat (*Triticum aestivum* L.). Mol Breed 2012;30:317–23. http://dx.doi.org/10.1007/s11032-011-9621-7.
[41] Gill BS, Chen PD. Role of cytoplasm-specific introgression in the evolution of the polyploid wheats. Proc Natl Acad Sci U S A 1987;84:6800–4. http://dx.doi.org/10.1073/pnas.84.19.6800.
[42] Zeller SL, Kalinina O, Brunner S, Keller B, Schmid B. Transgene × environment interactions in genetically modified wheat. PLoS One 2010;5:e11405. http://dx.doi.org/10.1371/journal.pone.0011405.
[43] Liu Y, Song L, Yu W, Hu Y, Ma X. Light quality modifies camptothecin production and gene expression of biosynthesis in *Camptotheca acuminata* Decne seedlings. Ind Crop Prod 2015;66:137–43. http://dx.doi.org/10.1016/j.indcrop.2014.12.046.
[44] Park Y, Muneer S, Jeong B. Morphogenesis, flowering, and gene expression of *Dendranthema grandiflorum* in response to shift in light quality of night interruption. Int J Mol Sci 2015;16:16497–513. http://dx.doi.org/10.3390/ijms160716497.
[45] Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. Science 2014;345:1249721. http://dx.doi.org/10.1126/science.1249721.
[46] Muterko A, Balashova I, Cockram J, Kalendar R, Sivolap Y. The new wheat vernalization response allele Vrn-D1s is caused by DNA transposon insertion in the first intron. Plant Mol Biol Rep 2014;294–303. http://dx.doi.org/10.1007/s11105-014-0750-0.
[47] Chouard P. Vernalization and its relations to dormancy. Annu Rev Plant Physiol 1960;11:191–238. http://dx.doi.org/10.1146/annurev.pp.11.060160.001203.

APPENDIX II

Heritable heading time variation in wheat lines with the same number of *Ppd-B1* gene copies

Ivaničová Z., Valárik M., Pánková K., Trávníčková M., Doležel J., Šafář J. and Milec, Z.

PLoS ONE; E 12(8): e0183745, 2017

PLOS ONE

RESEARCH ARTICLE

Heritable heading time variation in wheat lines with the same number of *Ppd-B1* gene copies

Zuzana Ivaničová[1], Miroslav Valárik[1], Kateřina Pánková[2], Martina Trávníčková[2,3], Jaroslav Doležel[1], Jan Šafář[1], Zbyněk Milec[1]*

1 Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ Olomouc, Czech Republic, **2** Crop Research Institute, Drnovská 507, Prague, Czech Republic, **3** Czech University of Life Sciences Prague, Kamýcká 129, Prague 6, Czech Republic

* milec@ueb.cas.cz

Citation: Ivaničová Z, Valárik M, Pánková K, Trávníčková M, Doležel J, Šafář J, et al. (2017) Heritable heading time variation in wheat lines with the same number of *Ppd-B1* gene copies. PLoS ONE 12(8): e0183745. https://doi.org/10.1371/journal.pone.0183745

Editor: Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA

Received: December 21, 2016

Accepted: August 10, 2017

Published: August 28, 2017

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (award LO1204 from the National Program of Sustainability I), Czech Science Foundation (grant no. P501-12-G090) and supported by Ministry of Agriculture of the Czech Republic, Project No. MZE ČR RO0417. The funders had no role in study design, data collection

Abstract

The ability of plants to identify an optimal flowering time is critical for ensuring the production of viable seeds. The main environmental factors that influence the flowering time include the ambient temperature and day length. In wheat, the ability to assess the day length is controlled by photoperiod (*Ppd)* genes. Due to its allohexaploid nature, bread wheat carries the following three *Ppd-1* genes: *Ppd-A1*, *Ppd-B1* and *Ppd-D1*. While photoperiod (in)sensitivity controlled by *Ppd-A1* and *Ppd-D1* is mainly determined by sequence changes in the promoter region, the impact of the *Ppd-B1* alleles on the heading time has been linked to changes in the copy numbers (and possibly their methylation status) and sequence changes in the promoter region. Here, we report that plants with the same number of *Ppd-B1* copies may have different heading times. Differences were observed among F_7 lines derived from crossing two spring hexaploid wheat varieties. Several lines carrying three copies of *Ppd-B1* headed 16 days later than other plants in the population with the same number of gene copies. This effect was associated with changes in the gene expression level and methylation of the *Ppd-B1* gene.

Introduction

The day length plays a crucial role in the plant life cycle and mainly impacts the decision of when to flower. Certain plants require a long day (LD), and certain plants require a short day (SD) before flowering; however, there are also plants that are insensitive to the day length. This behaviour is called a photoperiod response and is genetically controlled. Bread wheat (*Triticum aestivum* L.) is one of the world's most important staple food sources and provides nutrition to 30% of the human population. Originally, wheat was an LD plant that required at least 14 hours of daylight to flower. When grown under SD conditions, flowering was significantly delayed. However, the introduction of photoperiod-insensitive wheat cultivars facilitated the spreading of wheat cultivation to regions with favourable conditions (e.g., water availability, appropriate temperature) and short daylight.

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

In wheat, the photoperiod response is controlled by photoperiod (*Ppd*) genes that are located on chromosomes 2A, 2B and 2D. These genes belong to a pseudo-response regulator family [1]. Of these genes, the *Ppd-D1a* allele on chromosome 2D has a major effect on the photoperiod response. The allele causing the photoperiod insensitivity carries a 2,089 bp deletion upstream of the coding region [1]. Another photoperiod-insensitive allele, *Ppd-A1a*, is located on chromosome 2A and differs from the wild type by a 1,085 bp deletion. Interestingly, the deletion occurs at the same gene region as that in the *Ppd-D1a* allele [2]. In contrast, two different mechanisms were identified for the *Ppd-B1* allele. Díaz et al. [3] described the impact of different numbers of *Ppd-B1* copies on the photoperiod response as follows: while the presence of only one copy of *Ppd-B1* causes sensitivity to day length, an increased copy number of *Ppd-B1* is associated with day length insensitivity and, consequently, earlier flowering. The authors also reported different haplotypes of the *Ppd-B1a* allele based on a combination of the presence/absence of amplicons derived from the junctions between the individual *Ppd-B1a* copies. Zhang et al. [4] showed that different haplotypes are responsible for the variation in the heading time, and varieties with *Ppd-B1_Hapl-VI* demonstrated the earliest heading. Furthermore, Nishida et al. [2] described a 308 bp insertion in the promoter region of the winter variety Winter-Abukumawase. This insertion was associated with photoperiod insensitivity, and the allele carrying this mutation was designated *Ppd-B1a.1*. The Winter-Abukumawase variety has only one *Ppd-B1* copy (H. Nishida, pers. comm.); therefore, similarly to the *Ppd-D1* allele, the earlier heading appears to be associated with DNA sequence changes in the promoter region.

The copy number variation (CNV) involves deletions or duplications of DNA regions that are typically larger than 1 kb [5,6]. Such DNA rearrangements are often associated with changes in the phenotype. CNVs have been widely studied in humans [7,8] and have been linked with various diseases, such as Parkinson´s disease, which is associated with a triplication of the α-synuclein gene [9]. However, CNVs also played a significant role in the human genome evolution with positive effects on beneficial traits, such as endurance running, which is likely associated with a duplication of the *aquaporin 7* gene [10]. CNVs have also been studied in plants either at the whole genome level [11] or in analyses focusing on the expression of particular genes [3]. Francia et al. [12] reported an increased copy number of genes encoding C-repeat Binding Factor (*HvCBF4* and *HvCBF2*) in barley, which resulted in greater frost tolerance. Another study in barley showed that higher number of *HvFT1* copies (homologue of *FLOWERING LOCUS* T in barley) can also accelerate the flowering time [13].

Variation at the DNA sequence level is not the only factor that influences gene expression. Epigenetic modifications, including DNA methylation and histone modifications, play an important role and may lead to significant changes in the phenotype. In plants, DNA methylation occurs at several particular sites as follows: CG, CHG and CHH, where H stands for A, T or C, while in mammals, mainly the CG site is involved [14,15]. DNA methylation is usually associated with a lower gene expression and gene silencing [16,17]. Nevertheless, Sun et al. [18] observed a higher level of methylation of the *Ppd-B1* gene, which was linked to a higher number of copies of the gene. The authors suggested that the higher level of DNA methylation leads to an increased expression level of *Ppd-B1* and, consequently, an accelerated heading time. This effect might be attributed to reduced binding of a putative repressor to the promoter region, which results in an increased gene expression. An increase in the level of lysine 4 trimethylation in histone H3 (H3K4me3) was implicated in the upregulation of *TaVRN1* and *TaFT1* genes during the vernalization of winter wheat. This upregulated state was maintained throughout plant development and was reset only in the next generation [19]. Similarly, in *Arabidopsis*, the trimethylation of the lysine residue in histone H3 (H3K27me3) leads to a repression of the *FLOWERING LOCUS C* (*FLC*) gene during vernalization and results in the ability to flower [20,21].

In this study, we report differences in heading time among recombinant inbred lines (RIL) F_7 individuals that carry the same copy number of the *Ppd-B1* allele. These lines were derived by crossing two spring hexaploid wheat varieties with differing *Ppd-B1* copy numbers. Although the parental varieties carry the same set of alleles that control the flowering time (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *VRN-A3*, *vrn-B3*, *Ppd-D1b* and *Ppd-A1b*), several lines with three *Ppd-B1* copies differed significantly in the heading time. The difference was associated with changes in the *Ppd-B1* expression level. The role of the sequence and epigenetic differentiation of the *Ppd-B1* gene and its promoter in the heading time variation was investigated.

Materials and methods

Plant material

Seeds of the bread wheat (*Triticum aestivum* L.) cultivars Kaerntner Frueher (KF) and Paragon (P) were obtained from Gene Bank (https://grinczech.vurv.cz/gringlobal) of Crop Research Institute (Prague, Czech Republic). KF is a photoperiod-insensitive spring hexaploid wheat that carries the *Ppd-B1a* allele (three copies of *Ppd-B1*), while P is photoperiod sensitive and carries the *Ppd-B1b* allele (one copy of *Ppd*-B1). The remaining alleles influencing the flowering time are the same in both cultivars (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *vrn-B3*, *Ppd-D1b* and *Ppd-A1b* [22] and *VRN-A3*). The two cultivars were crossed, and an F_2 population was obtained. Then, RILs were created using the single seed descent method to obtain the F_7 generation. Plants in the F_7 population were grown in ten replicates under a controlled regime (20˚C/16˚C–day/night temperature) in a growth chamber under an artificial LD of 16 hours. Three biological replicates of each line were used to analyse the gene expression and DNA methylation status of *Ppd-B1*.

QTL analysis

Five seeds of each parental cultivar and all 130 individuals in the F_2 mapping population were incubated for four days at 6˚C to synchronise germination. Seedlings were grown under controlled conditions (20˚C/16˚C–day/night temperature) under artificial light with a 16-hour LD. The heading date was recorded at the emergence of half of the spike. For the QTL analysis, 90 F_2 individuals were selected to fit one 96-well PCR plate. DNA was extracted from the leaves using the Invisorb Spin Mini Plant Kit (Stratec Molecular, Berlin, Germany) according to the manufacturer´s instructions. The mixed model QTL analysis [23], genotyping according to the DArT markers [24] and the genetic map construction were carried out by Diversity Arrays Technology (DArT, Canberra, Australia http://www.diversityarrays.com).

VRN-A3 sequencing

The *VRN-A3* gene was sequenced to identify variations between the parental cultivars. In total, six primer pairs (Table 1) were used to obtain the complete gene sequence and the 1,750 bp and 143 bp flanking sequences from the 5´ and 3´ end of the gene, respectively.

The *VRN-A3* gene-specific primers were designed using publicly available sequences (GenBank accession nos. DQ890162, EF428115 and EF428119) and chromosome-specific survey sequences of cv. Chinese Spring [25]. All sequences were aligned using Geneious 5.6.4 software [26]. The primer pairs were considered specific when at least one primer contained three or more genome-specific SNPs/indels. The melting temperatures were determined using Primer-BLAST [27]. The PCR reaction mix (15 μl volume) comprised 10 ng of the template DNA, 200 μM of each dNTP, 2 mM of $MgCl_2$, 0.2 μM of both primers, 1 × PCR buffer B1 (Solis BioDyne, Tartu, Estonia) and 0.03 U of Hot FirePol *Taq* polymerase (Solis BioDyne, Tartu, Estonia). This mixture was consecutively used for the methylation status determination, cloning

and copy number assessment. The DNA amplification was carried out by touchdown PCR as follows: initial denaturation for 13 min at 95˚C, followed by 14 cycles at 95˚C for 40 s, 65˚C for 40 s (increment of -0.7˚C/per cycle) and 72˚C for 2 min. An additional 19 cycles were performed at 95˚C for 40 s, 55˚C for 40 s and 72˚C for 2 min with a final elongation at 72˚C for 10 min. The purification of the product and sequencing were performed as previously described by Ivaničová et al. [28].

Ppd-B1 promoter sequencing

The promoter region and a portion of the first exon (from—1545 bp to + 251 bp) of the parental cultivar KF and the F_7 lines 32_2 and 37_4 were sequenced to determine whether there were sequence variations that could possibly cause the difference in the heading time. The primers (Table 1) were designed based on the Sonora 64 sequence (JF946486.1) using Primer-BLAST [27]. The DNA amplification and purification and the sequencing of the PCR product were performed in same manner as described above for the *VRN-A3* gene.

RNA extraction and expression analysis

The expression of the *Ppd-B1* gene was assessed in lines of the F_7 mapping population and the parental cultivars. RNA was extracted from the leaves of 20-day-old seedlings using the RNasy

Table 1. Primer pairs used for the sequencing of the *VRN-A3* and *Ppd-B1* genes.

Gene	Name	Sequence (5´- 3´)	Product size (bp)	Reference
Vrn-A3	VRN3_F4	CTAAATAGCAAGACGCCACTAT	545	this study
	VRN3_R4	CCTCCTAGAAACTGCCACACA		
	VRN3_F5	TAATGGACCTCCATAGCTAGC	909	this study
	VRN3_R5	GTGGCTTCTAGGCCGTGCC		
	VRN3_F6	CGCAGCTCATACCTTGGACTA	1365	this study
	VRN3_R6	CATCGAGAATCATCTTCCCAC		
	VRN3_F7	ACGTCCACAGAACCAATTCA	604	this study
	VRN3_R7	CATCGAGAATCATCTTCCCAC		
	VRN3_F8	AAATGGCCGGGAGGGAC	552	this study
	VRN3_R8	ACGTAGAGAGTACTACGTGC		
	VRN3_F9	GTTCTGGCAAGCACACGAC	597	[28]
	VRN3_R9a	AATTTGCTGACTTTGCGGGC		
Ppd-B1	PromF1	AAGTGTACGTGGTTAACATTAG	623	this study
	PromR1	GGAGTTATCTTAACACTTGC		
	PromF2	GTGCTAAGATAACTTTGTC	546	this study
	PromR2	GAAAGGAAAGAAAGAAGC		
	PromF3	AAATATGCGCTGTATGTC	539	this study
	PromR3	CGTGAACAAGACCAGGACCAG		
	PromF4	GTATAGAGTCAGAAGGAGGGC	522	this study
	PromR4	TGAGTGCCAGATCCAAAAGCTG		
	Ex1_F1	TGCCATATAGATCCTTTCTGATA	448	this study
	Ex1_R1	TTGACACCAACAGCTTCCAG		
	Ppd-B1_101F	CGCCACTGCATGTACCAAGTTA	96	this study
	Ppd-B1_101R	CTGTCAGAACAAGGTCGATGTTG		

A set of six primer pairs was used to sequence the 2,732 bp region, which includes the *VRN-A3* gene (839 bp) and flanking regions (1,750 and 143 bp from the 5´ and 3´ end of the gene). Amplification was performed using a touchdown PCR protocol. For the *Ppd-B1* gene, a set of five primer pairs was used to sequence the promoter region together with a portion of the first intron (from—1,545 bp to + 251 bp).

https://doi.org/10.1371/journal.pone.0183745.t001

Plant Mini Kit (Qiagen, Hilden, Germany). The leaves were collected from all lines in three biological replicates three hours after dawn when the *Ppd-B1* expression reached the maximum level [3]. DNA was removed during the RNA purification using the RNase-Free DNase Set (Qiagen, Hilden, Germany). cDNA was synthesized using the Transcription High Fidelity cDNA Synthesis Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer´s protocol with 2 μg of total RNA and anchored-oligo $(dT)_{18}$ primers.

The gene expression level was determined using reverse transcription-qPCR (RT-qPCR). RT-qPCR was performed using qPCR 2x SYBR Master Mix (Top-Bio, Prague, Czech Republic) on the CFX96TM Real-Time PCR Detection System (Bio-Rad, Hercules—California, USA). The expression level of *Ppd-B1* was standardized against the reference gene *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) according to Sun et al. [18].

The data were analysed using the $2^{-\Delta\Delta Cq}$ method with CFX Manager 3.0 software (Bio-Rad, Hercules—California, USA). Three replicate PCR amplifications were performed for each sample. The transcript level of the target gene *Ppd-B1* in cultivar KF was designated 1.0. The primers for *Ppd-B1* (Ppd-B1_101F: CGCCACTGCATGTACCAAGTTA, Ppd-B1_101R: CTGTCAGAACAAGGTCGATGTTG) were designed with Primer Express® Software v3.0.1 (Thermo Fisher Scientific, Waltham—Massachusetts, USA). The primers' efficiency and correlation coefficient were E = 102.8% and R^2 = 0.991, respectively.

Methylation status and DNA sequence comparison

Genomic DNA was extracted in three biological replicates from young leaves of F_7 individuals and the parental cultivars using the Invisorb Spin Mini Plant Kit (Stratec Molecular, Berlin, Germany) according to the manufacturer´s instructions. To determine the methylation status, the DNA was bisulfite converted using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine—California, USA). The amplification of the converted DNA was carried out using conditions and primers for the "Region II" as described by Sun et al. [18]. This region is associated with significant differences in the methylation of CpG islands between photoperiod-insensitive and photoperiod-sensitive varieties. The primers spanned the region from -1,250 to -778 bp from the start codon. Since the primers used by Sun et al. [18] for the amplification of the unconverted genomic DNA did not perform reliably in our materials, new primers were designed (PromF2: GCCTTACGCACATCATCAGC and PromR2: GGTGACGTGGACGAAATGGA) that spanned the region from -1,220 to -659 bp. The primer specificity was verified using DNA amplified from chromosomes 2A, 2B and 2D of both parental cultivars as the template for the PCR. The chromosomes were purified by flow cytometric sorting after labelling GAA microsatellites using FISHIS [29], and their DNA was amplified according to Šimková et al. [30].

PCR was performed at 95˚C for 13 min for the initial denaturation, followed by 34 cycles of 95˚C for 30 s, 60˚C for 30 s and 72˚C for 1 min, and a final step at 72˚C for 5 min. All amplifications were performed using the C1000 Touch TM Thermal cycler (Bio-Rad, Hercules—California, USA).

The PCR amplicons from both the original genomic DNA and the bisulfite converted genomic DNA of the parental cultivars and F_7 lines were purified using Agenocourt AMPure XP (Beckman Coulter, Brea—California, USA) and cloned using TOPO® TA Cloning® Kit for Sequencing (Invitrogen, Carlsbad—California, USA) according to the manufacturer's instructions. Inserts from the DNA clones were amplified in a 20 μl reaction mixture comprising a single bacterial colony as the template. The reaction conditions were as follows: initial denaturation at 95˚C for 13 min, followed by 35 cycles of 95˚C for 35 s, 55˚C for 35 s, and 72˚C for 50 s with a final extension for 10 min at 72˚C. The PCR products were purified and sequenced as previously described [28]. The sequences were trimmed and aligned using Geneious 5.6.4

(http://www.geneious.com, [26]). The methylation status was determined with the Kismeth online tool, which is available at http://katahdin.mssm.edu/kismeth/revpage.pl.

Copy number assessment and SNP marker development

The *Ppd-B1* gene copy number in the F_2 population and both parents was determined by iDNA Genetics (Norwich, UK) using the Taqman® assay as previously described by Díaz et al. [3]. To assess the CNV in the F_7 plants, a method based on SNP polymorphisms was used. The size of the *Ppd-B1* locus and a higher inter-copy identity prevents the sequencing of individual *Ppd-B1* copies from a single chromosome. Therefore, the sequence of *Ppd-B1* gene of the cultivar Sonora 64 (JF946486.1), which has three copies of the gene similarly to KF, was used to develop the markers. We identified four SNPs between Sonora 64 and P (DQ885762). A primer pair (GACTCCTGCCATGAGTTTTGATG and ACCGCAGTGTGACTTCGATTATC) was used to identify the Sonora 64-like allele and designed to overlap the SNPs located at positions -10,427 bp (G-A) and -10,656 bp (A-G) from the START codon. The marker was named *owm1001*. The PCR amplification was performed in a 15 μl reaction mixture under the following cycling conditions: initial denaturation at 95˚C for 13 min, followed by 35 cycles of 95˚C for 35 s, 55˚C for 35 s and 72˚C for 50 s with a final elongation of 10 min at 72˚C. The amplicons were purified and sequenced as previously described by Ivaničová et al. [28]. The sequences were trimmed and aligned using the software Geneious 5.6.4 [26].

Results

Mapping and QTL analysis

Ninety individuals of the F_2 mapping population (30 from each of the early, intermediate and late flowering lines) were genotyped using the DArT technology [23]. In total, 5,069 DArT markers were found to be polymorphic and used to construct the genetic map. The heading times of the F_2 lines ranged from 50 to 78 days. The QTL analysis revealed only one significant peak (LOD 15.6) on chromosome 2BS (Fig 1). Marker *owm1001*, which was derived from the *Ppd-B1* promoter region, mapped onto the peak, thus confirming that the heading time difference was mediated by the *Ppd-B1* gene. There was also one marker located on chromosome 7A (Fig 1), but further analyses showed that it was not significant.

VRN-A3 gene and *Ppd-B1* promoter sequence variation

The sequence of the *VRN-A3* gene (839 bp) and the adjacent 5′ end (1,750 bp) and 3′ end (143 bp) sequences were obtained from both parental cultivars. No sequence variations were identified. By comparing to the reference sequence of the cultivar Chinese Spring *FT* (*VRN-A3*) gene (EF428115), only one SNP (T → C) was shown, which was located 506 bp from the START codon in the first intron. This comparison confirmed that the putative marker association identified by the QTL analysis on chromosome 7A was most likely a bias.

A comparison of the sequence promoter and a portion of the first exon of the *Ppd-B1* gene (from—1,545 bp to + 251 bp) in the F_7 lines 32_2 and 37_4 and the parental cultivars showed no variations.

Copy number determination

Using the Taqman® assay as previously described by Díaz et al. [3], we found that the early flowering cultivar KF carried three copies of *Ppd-B1* (designated the *Ppd-B1a* allele), while the late flowering cultivar P carried only one copy (designated the *Ppd-B1c* allele) per haploid genome. The same result was observed in the F_2 individuals as follows: 30 early heading plants

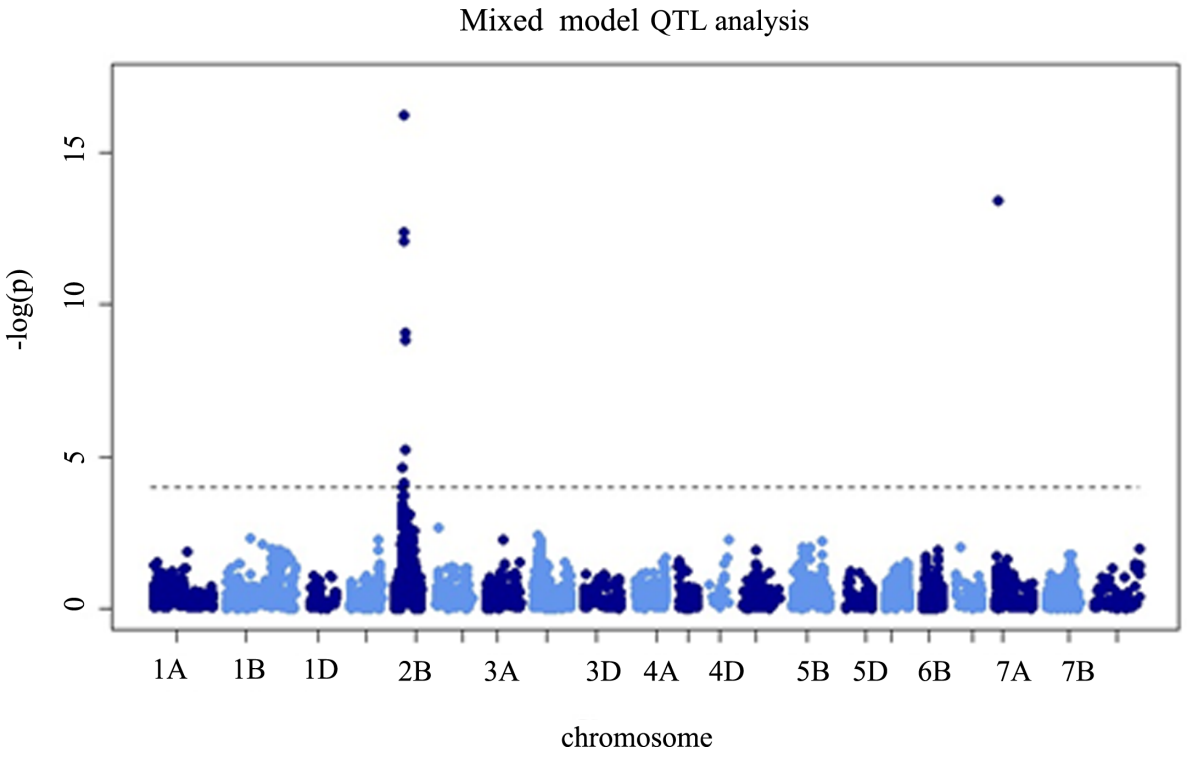


Fig 1. Whole genome QTL analysis of the heading time variation. Ninety F_2 individuals were genotyped using the DArT markers. A mixed model QTL analysis revealed only one significant peak on short arm of chromosome 2B (2BS).

https://doi.org/10.1371/journal.pone.0183745.g001

carried three copies, and 30 intermediate plants that are heterozygous in the *Ppd-B1* locus. Surprisingly, not all late plants had one copy as expected. Of these 30 late heading plants, 28 carried one copy of *Ppd-B1*, but two plants had three copies.

Heading time

The mean heading time for KF was 50.6 ± 1.2 days, while the mean heading time for P was 86.2 ± 1.5 days. The heading time variation between the parents was observed under both SD and LD conditions. Offspring (generations F_2—F_7) derived from crossing the two cultivars were grown under LD conditions only to shorten the time to harvest the seeds and accelerate the attainment of the F_7 generation with a homogenous background. The heading time of the F_2 lines varied between 50–78 days and corresponded with the respective *Ppd-B1* CNV for a majority of plants (88 out of 90), i.e., plants with one copy of *Ppd-B1* headed later than plants with three copies. Surprisingly, we also observed heading time variations among the F_2 lines with three copies of the gene. These lines (32_2 and 37_4) headed significantly later (17 days on average) than the early flowering parent KF with three gene copies. This phenomenon was observed repeatedly across the generations; however, a heading time analysis was performed only for the F_6 and F_7 generations. In the F_6 generation (field conditions, natural LD), the mean heading times of lines 32_2 and 37_4 were 90.8 and 90.5 days, respectively, while the early parent KF headed in 78 days, and the late parent P headed in 96.1 days. These late lines and one early line (11_6, serving as a control) carried three *Ppd-B1* copies and were, therefore, selected for further analyses in the F_7 generation.

A phenotypic analysis of the F_7 lines (controlled regime, artificial LD) confirmed the results obtained in the F_2 mapping population. The early line 11_6 headed at 52.2 ± 2.2 days on

average, while the late lines 37_4 and 32_2 headed at 66.6 ± 4.7 and 74.1 ± 5.2 days on average, respectively. As previously mentioned, all lines (both early and late) had the *Ppd-B1a* allele (three copies). The early parent KF carries the same allele, and thus, late flowering was unexpected in these two lines. To clarify this observation, we assessed the *Ppd-B1* expression level and methylation status of the lines to determine the reason for the large differences in the heading times (16–24 days) between the lines with the same *Ppd-B1* copy number. Prior to the expression analysis, the lines were genotyped with markers (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *vrn-B3*, *Ppd-D1b* and *Ppd-A1b*) to exclude the possibility of outcrossing during the development of the F_7 generation.

Ppd-B1 expression

The *Ppd-B1* expression level (Fig 2) was determined using a RT-qPCR assay with *Ppd-B1* specific primers. As expected, plants in the early flowering line had expression levels that were similar to those of KF, while the plants in the late flowering lines had a decreased expression level.

The differences between the late lines 32_2 and 37_4 and late parent P were statistically significant at the $P<0.05$ level. The expression levels in the early line 11_6 and early parent KF were not significantly different, while the differences between the early line 11_6 and late lines 32_2 and 37_4 were also significant at the $P<0.05$ level. The lower expression of *Ppd-B1* in P correlated with the delayed heading time ($r = -0.955$, $P = 0.00001$). However, the gene expression levels in lines 32_2 and 37_4 did not correlate with the respective *Ppd-B1* copy number.

Methylation status of the *Ppd-B1* promoter region

The amplification of the unconverted genomic DNA with the newly designed primers resulted in an amplicon size of 561 bp, while the amplicon size of Region II using the primers designed

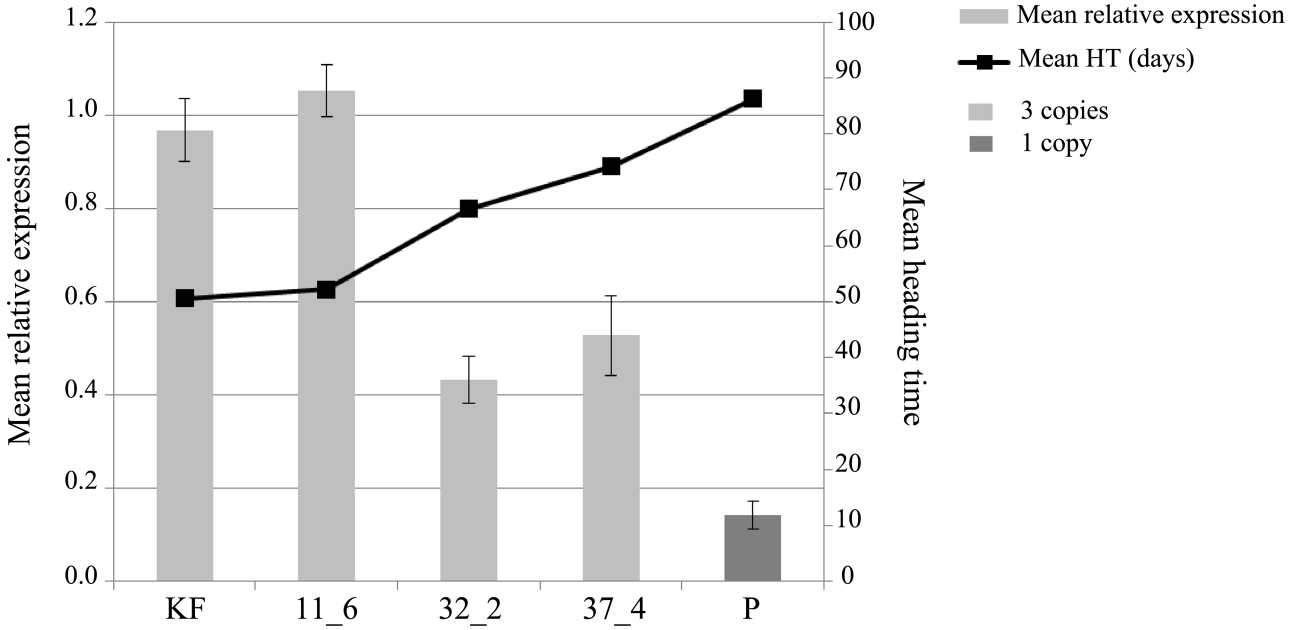


Fig 2. Mean relative expression of the *Ppd-B1* allele and the mean heading times of the parental cultivars and the F_7 lines. 11_6—early line; 32_2 and 37_4—late lines; KF—early parental cultivar Kaerntner Frueher; P—late parental cultivar Paragon; HT—heading time. The data were derived from three biological replicates. The mean standard error of the relative expression is shown for each line.

https://doi.org/10.1371/journal.pone.0183745.g002

Table 2. Methylation status of the CpG islands in the early and late flowering lines (F_7 generation).

Line	Plant	Methylation status (%)				*Ppd-B1* copies
		CG	Mean CG with SEM	CHG	CHH	
Early parent KF	121	85.69	87.17 ± 0.76	14.69	1.6	3
	123	87.62		12.37	2.69	
	129	88.21		13.24	2.09	
11_6	7	88.62	90.22 ± 1.02	16.64	0.70	3
	12	92.12		19.77	1.22	
	19	89.92		15.21	1.13	
32_2	101	74.08	70.61 ± 1.79	17.26	1.95	3
	104	69.66		11.82	1.14	
	109	68.10		9.81	1.13	
37_4	111	92.73	91.73 ± 0.79	19.48	1.65	3
	112	90.16		16.99	1.62	
	117	92.30		15.47	1.09	
Late parent P	134	11.6	9.97 ± 0.86	11.11	2.35	1
	135	8.68		12.31	1.68	
	136	9.64		14.54	1.95	

11_6—early line; 32_2 and 37_4—late lines; KF—early parental cultivar Kaerntner Frueher; P—late parental cultivar Paragon; SEM—standard error of mean

https://doi.org/10.1371/journal.pone.0183745.t002

by Sun et al. [18] was 472 bp. The combination of the sequencing data from the two amplicons resulted in a common 442 bp region that was used to assess the methylation status. The sequencing of at least 70 subclones per line revealed that the early parental variety KF with three copies of *Ppd-B1* had 85.69% of CG methylated, while the late P variety with one copy had only 9.97% of CG methylated (Table 2, Fig 3). The early line 11_6 showed a methylation status that was similar to that of KF (mean value 90.22%, Table 2, S1 Fig). Late line 32_2 had a slightly decreased methylation level of CG (mean value 70.61%, Table 2, S1 Fig), and this difference was statistically significant ($P<0.01$). Late line 37_4 showed a repeatedly high level of methylation (mean value 91.73%, Table 2, S1 Fig), which did not correspond with either the late heading or the lower expression level of *Ppd-B1*. Therefore, we have sequenced the promoter regions of the late lines 37_4 and 32_2 and parental cultivar KF. No polymorphisms were identified by comparing the promoter region sequences with parental cultivar P.

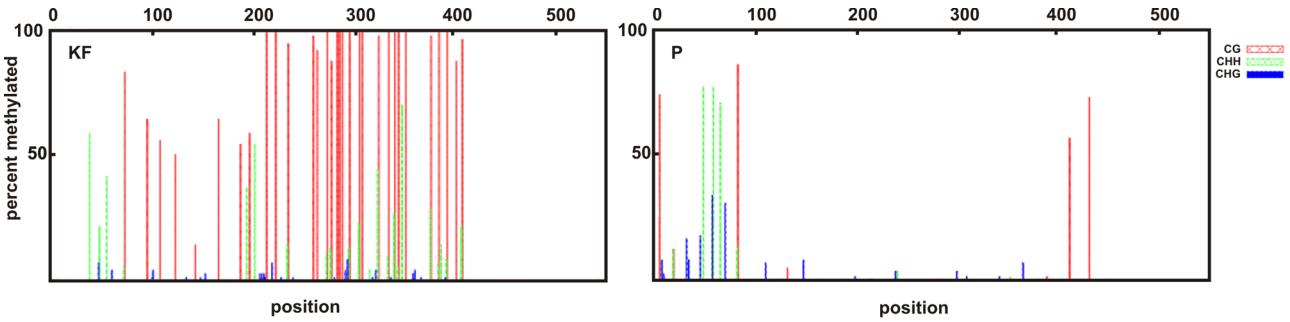


Fig 3. Comparison of the methylation status of the parental cultivars. The methylation level in the promoter region (442 bp) of the *Ppd-B1* gene from parental cultivars KF and P was analysed. The x-axis shows the cytosine positions in the analysed region, and the y-axis shows the percent of methylated CpG islands. A major difference was observed in the CG sites, and cultivar P had an 8.8-fold lower methylation level. Red lines represent the methylation of CG sites, the green lines represent the CHG sites and the blue lines represent the CHH sites.

https://doi.org/10.1371/journal.pone.0183745.g003

Discussion

Heading time and *Ppd-B1* CNV

The difference in the heading time between parental cultivars KF and P was first detected in 2010 during experiments focused on heading date variations due to a novel *Vrn-B1* allele [22]. The difference ranged from 20 to 40 days depending on the growth conditions (LD or SD) and was reproducible in several replications in both field and growth room trials. Surprisingly, both varieties have the same set of the following major flowering time alleles: *Vrn-A1*, *Vrn-B1c*, *vrn-D1*, *Vrn-A3*, *vrn-B3*, *Ppd-A1b* and *Ppd-D1b*. Even with identical major flowering time genes (vernalization and photoperiod) or saturation of their requirements, variations in the flowering time could still be observed. Such changes or variations are usually very subtle and represent a fine-tuning of the flowering time by so-called earliness *per se* genes. Earliness *per se* genes usually fine-tune the flowering time based on environmental conditions that are different from the vernalization and photoperiod. The earliness *per se* genes do not contribute to the flowering time variations dramatically, but under certain conditions, their contribution could be significant. The *Eps-A1* gene from chromosome $1A^m$ of *T. monococcum* is a good example of such a gene [31,32]. This gene under field conditions contributes to the flowering time variation by four to six days, but under fluorescent light and a temperature of 16˚C or 23˚C, the flowering time difference could be 80 and 50 days, respectively, [31]. This gene has a pleiotropic effect and affects all developmental phases from the double ridge stage to the heading stage [33].

In the present work, the QTL analysis of the F_2 mapping population identified a single significant peak (LOD 15.6) on chromosome arm 2BS at the position of the *Ppd-B1* gene. Over 5,000 DArT markers were used in the QTL analysis, which provided enough sensitivity to exclude the effect of other loci, including the earliness *per se* genes. However, by assessing the CNV of the *Ppd-B1* alleles [3], we confirmed that the early flowering cultivar KF carried three copies of *Ppd-B1* (designated the *Ppd-B1a* allele), while the late flowering cultivar P carried one copy only (designated the *Ppd-B1c* allele) per haploid genome.

Similarly, Díaz et al. [3] and Cane et al. [34] observed variations in the heading time due to differences in the copy number of the *Ppd-B1* gene. Earlier flowering corresponded to a higher level of *Ppd-B1* gene expression in the lines with the three gene copies than in those with only one copy of the *Ppd-B1* gene; thus, a shorter time was necessary for the plants to achieve a critical amount of the Ppd-B1 protein and an earlier initiation of flowering. In our work, we observed a significant heading time variation in the F_2 mapping population (derived from crossing the cultivars KF and P) with the same number of *Ppd-B1* gene copies. Cane et al. [34] observed a similar effect—plants with two copies had a longer heading time than plants carrying only one copy. The authors speculated that they identified a different *Ppd-B1* allele than the allele studied by Díaz et al. [3]. According to Díaz et al. [3], the cultivar Chinese Spring with three intact and one truncated *Ppd-B1* copies had a longer flowering time than plants with two copies of the gene. The variability in the heading time in the lines with the same copy number of flowering time gene (*HvFT1*) was also detected in barley [13]. Nitcher et al. observed that the heading time in F_2 individuals of BGS213 × IMC mapping population with five copies of the *HvFT1* gene ranged from 30 to 39 days, while plants with three copies headed between 40 and 79 days, and plants with one copy headed between 100 and 125 days.

The effect of the CNV of the *Ppd-B1* gene on the heading date was studied recently by Würschum et al. [35], who showed the importance of the CNV for the heading time and wheat adaptation to different environmental conditions. The authors screened a panel of 1,100 winter wheat varieties to determine the frequency and geographic distribution of the CNVs at the *Ppd-B1* and *Vrn-A1* loci along with their effect on the flowering time. Their study confirmed

that an increase in the number of *Ppd-B1* gene copies reduced the number of days required for heading. Interestingly, in contrast to the general trend, varieties from the Balkan region flowered later and had higher *Ppd-B1* copy numbers. We observed the same effect in the line with the fixed genomes that were isogenic for the major flowering time genes. Thus, we speculated that the difference in the flowering time in the line with the same *Ppd-B1* copy number was due to an altered *Ppd-B1* gene expression. To validate this hypothesis, we used the F_7 generation of the mapping population for further analyses.

Ppd-B1 expression and promoter methylation

The expression of *Ppd-B1* increases quite rapidly, peaks three hours after dawn, and then decreases and oscillates close to the level maintained during the dark period [3]. This phenomenon can be observed in photoperiod-sensitive and photoperiod-insensitive plants. However, the difference in the gene expression between both types of plants is relatively small. We analysed the *Ppd-B1* expression in both parents and in early and late lines with three *Ppd-B1* copies and found a robust correlation between the heading time and the expression level ($P = 0.00001$). As expected, the late cultivar P had a lower relative expression than the early cultivar KF. The early line 11_6 with three copies had the same expression as KF, while the late lines (32_2 and 37_4) had a decreased expression ($P<0.05$) that was still higher than that of the late parental cultivar P (Fig 2).

The increase in the copy number of the *Ppd-B1* gene and the higher methylation level in its promoter appear to be associated with a higher expression, which, in turn, is associated with an acceleration of flowering and even a loss of photoperiod sensitivity. Sun et al. [18] reported a link between the methylation of the *Ppd-B1* allele, increased CNV and a higher level of *Ppd-B1* expression. Wheat cultivars with three copies of the *Ppd-B1* gene have more methylated CpG islands than varieties with one gene copy [18]. The authors described the following six regions in the *Ppd-B1* gene: regions I—IV are located in the promoter (upstream the coding region) and regions V and VI are located in the coding region, spanning exons 1–3. The most significant difference in the CpG methylation between the photoperiod-insensitive and photoperiod-sensitive cultivars was observed in Region II. The authors characterized the "*a*" and "*b*" methylation haplotypes as follows: the *a* haplotype (higher methylation) was associated with a higher gene copy number and exhibited a higher expression of *Ppd-B1* than the varieties with the "*b*" haplotype (lower methylation) and lower CNV. Nevertheless, several cultivars with a low methylation level and one *Ppd-B1* gene copy had a higher expression than cultivars with three or even four copies of the gene and a high methylation status. Although DNA methylation is usually connected with gene silencing [36], in the case of *Ppd-B1*, it appears to act in the opposite way. Sun et al. [18] also noted that methylated *Ppd-B1* promoter regions cover the same area, which is deleted in the *Ppd-A1a* and *Ppd-D1a* alleles; these alleles have a higher expression than their respective alleles without the deletion. The authors concluded that a deletion or methylation in this region may block putative repressor function and, therefore, lead to an enhanced *Ppd-B1* expression.

Identity of copies and suggestive effect of paramutation

A general understanding of the "gene copy number" implies that extra copies are the same as the original, which may be true shortly after a duplication event. However, the individual gene copies may gradually accumulate differences, which may change gene expression or even function [37]. In addition, an increased copy number may compromise the stability of homeostasis, and the extra gene copies may be downregulated unless they present an evolutionary advantage.

The variation in the heading time between the wheat lines with different numbers of *Ppd-B1* gene copies may occur due to differences in the DNA sequences between the individual gene copies. Such differences were reported in the cultivar Chinese Spring, which has four copies of the *Ppd-B1* gene distributed over a 100 kb region, with truncation of the first copy [3]. As previously mentioned, Cane et al. [34] identified the *Ppd-B1* allele with two copies, but in contrast to the observations by Díaz et al.[34], this allele caused a longer flowering time. Cane et al. [34] explained the discrepancy by the possible existence of two different alleles. Würschum et al. [35] also suggested that individual copies of the *Ppd-B1* gene may differ in DNA sequences and are non-functional or have different functional properties. In fact, indel-related variation was found to have a more significant effect on the flowering time than the variation in the copy number [38].

To clarify this issue, we sequenced a portion of the promoter region and *Ppd-B1a* allele from both parental lines, and no differences in the DNA sequence were found. This result implied that the effect on the flowering time was largely due to the differences in the copy number. A lower level of DNA methylation corresponds to lower gene expression and, subsequently, the later flowering time observed in line 32_2 with three copies of *Ppd-B1* than that of the early flowering cultivar KF with the same number of *Ppd-B1* copies. However, line 37–4, which also has three copies of *Ppd-B1*, had a higher methylation level than the early flowering cultivar KF and also flowered even later than line 32_2 (Table 2). This observation suggests that the optimal methylation status of the promoter activation would be close to values observed in the early cultivar KF (87%) and line 11_6 (90%) (Table 2). In this case, approximately 71% (line 32_2) may not be enough for promoter activation, and the 92% (line 37_4) may have an inhibitory effect on expression.

The delay in the heading time detected in the plants with three copies of the *Ppd-B1* gene might be explained by a paramutation effect. Paramutation is an allelic interaction in which one allele, which is referred to as paramutagenic (in our study *Ppd-B1b*), causes a heritable change in the expression of a homoeologous paramutable (in our study *Ppd-B1a*) allele [39]. One can speculate that the decreased level of DNA methylation observed in line 32_2 (three copies) might be due to paramutation. Different epialleles may be responsible for the difference in the gene expression patterns and the following variable range of phenotypes (reviewed in [40]). The irregularities we identified in the heading times may be explained by the weak stability of epigenetic modifications at the DNA level [41].

In conclusion, we have described late flowering wheat lines carrying the same number of *Ppd-B1* copies (three) as the early flowering parental variety. A late flowering time was associated with a lower *Ppd-B1* expression than that in a parent with the same gene copy number. While the molecular mechanisms underlying this interesting phenomenon remain unclear, our results indicate that the CNV may not be responsible for complete gene penetrance. The CNV of the *Ppd-B1* allele has a clear impact on the heading date and, thus, should be employed as a source of variability in breeding programmes. The insensitive allele of *Ppd-B1* was identified recently, and thus, its mode of function is not well understood. Further investigations are needed to fully explain the behaviour of the *Ppd-B1* alleles and their effect on flowering time pathways.

Supporting information

S1 Fig. Comparison of the methylation status of the F_7 lines. The methylation level in the promoter region (442 bp) of the *Ppd-B1* gene from F_7 lines was analysed. The x-axis shows the cytosine positions in the analysed region, and the y-axis shows the percent of methylated CpG islands. Red lines represent the methylation of CG sites, the green lines represent the CHG

sites and the blue lines represent the CHH sites.A-C: early line 11_6; D-F: late line 32_2; G-I: late line 37_4.
(PDF)

Acknowledgments

We thank Dr. Andrzej Kilian for the DArT genotyping and data analysis. We thank our colleagues J. Vrána for the chromosome sorting and R. Šperková and M. Seifertová for the excellent technical assistance.

Author Contributions

Conceptualization: Miroslav Valárik, Jan Šafář, Zbyněk Milec.

Formal analysis: Zuzana Ivaničová, Zbyněk Milec.

Funding acquisition: Miroslav Valárik, Kateřina Pánková, Jaroslav Doležel.

Investigation: Zuzana Ivaničová, Kateřina Pánková, Martina Trávníčková, Zbyněk Milec.

Project administration: Miroslav Valárik, Jan Šafář, Zbyněk Milec.

Resources: Jaroslav Doležel.

Supervision: Miroslav Valárik, Jaroslav Doležel, Jan Šafář.

Validation: Zuzana Ivaničová, Miroslav Valárik, Zbyněk Milec.

Writing – original draft: Zuzana Ivaničová, Miroslav Valárik, Kateřina Pánková, Martina Trávníčková, Jaroslav Doležel, Jan Šafář, Zbyněk Milec.

Writing – review & editing: Zuzana Ivaničová, Miroslav Valárik, Martina Trávníčková, Jaroslav Doležel, Jan Šafář, Zbyněk Milec.

References

1. Beales J, Turner A, Griffiths S, Snape J, Laurie D. A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theor Appl Genet. 2007; 115: 721–733. https://doi.org/10.1007/s00122-007-0603-4 PMID: 17634915
2. Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, et al. Structural variation in the 5′ upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (Triticum aestivum L.), and their effect on heading time. Mol Breed. 2012; 31: 27–37. https://doi.org/10.1007/s11032-012-9765-0
3. Díaz A, Zikhali M, Turner A, Isaac P, Laurie D. Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One. 2012; 7: e33234. https://doi.org/10.1371/journal.pone.0033234 PMID: 22457747
4. Zhang X, Gao M, Wang S, Chen F, Cui D. Allelic variation at the vernalization and photoperiod sensitivity loci in Chinese winter wheat cultivars (Triticum aestivum L.). Front Plant Sci. 2015; 6: 1–10. https://doi.org/10.3389/fpls.2015.00001
5. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006; 444: 444–54. https://doi.org/10.1038/nature05329 PMID: 17122850
6. Zmieńko A, Samelak A, Kozłowski P, Figlerowicz M. Copy number polymorphism in plant genomes. Theor Appl Genet. 2014; 127: 1–18. https://doi.org/10.1007/s00122-013-2177-7 PMID: 23989647
7. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36: 949–951. https://doi.org/10.1038/ng1416 PMID: 15286789
8. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2012; 464: 704–712. https://doi.org/10.1038/nature08516.Origins

9. Singleton AB, Farrer M, Johnson J, Singleton AB, Hague S, Kachergus J, et al. -Synuclein Locus Triplication Causes Parkinson's Disease. Science (80-). 2003; 302: 841–841. https://doi.org/10.1126/science.1090278 PMID: 14593171
10. Dumas M, Sadick N, Noblesse E, Juan M, Lachman-Weber N, Boury-Jamot M, et al. Hydrating skin by stimulating biosyntheis of aquaporins. J Drugs Dermatology. 2008; 6(6 Suppl): 20–24.
11. Saintenac C, Jiang D, Akhunov ED. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol. BioMed Central Ltd; 2011; 12: R88. https://doi.org/10.1186/gb-2011-12-9-r88 PMID: 21917144
12. Francia E, Morcia C, Pasquariello M, Mazzamurro V, Milc JA, Rizza F, et al. Copy number variation at the HvCBF4–HvCBF2 genomic segment is a major component of frost resistance in barley. Plant Mol Biol. Springer Netherlands; 2016; 161–175. https://doi.org/10.1007/s11103-016-0505-4 PMID: 27338258
13. Nitcher R, Distelfeld A, Tan C, Yan L, Dubcovsky J. Increased copy number at the HvFT1 locus is associated with accelerated flowering time in barley. Mol Genet Genomics. 2013; 288: 261–75. https://doi.org/10.1007/s00438-013-0746-8 PMID: 23591592
14. Ehrlich M, Gama-Sosa M, Huang L, Midgett R, Kenneth C, Mccune R, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acid Res. 1982; 10: 2709–2721. PMID: 7079182
15. Henderson IIRI, Chan SSRS, Cao X, Johnson L, Jacobsen SSESSE. Accurate sodium bisulfite sequencing in plants. Epigenetics. 2010; 5: 47–49. https://doi.org/10.4161/epi.5.1.10560 PMID: 20081358
16. Bird A. The essentials of DNA methylation. Cell. 1992; 70: 5–8. https://doi.org/10.1016/0092-8674(92)90526-I PMID: 1377983
17. Keshet I, Lieman-Hurwitz J, Cedar H. DNA methylation affects the formation of active chromatin. Cell. 1986; 44: 535–543. https://doi.org/10.1016/0092-8674(86)90263-1 PMID: 3456276
18. Sun H, Guo Z, Gao L, Zhao G, Zhang W, Zhou R, et al. DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (Triticum aestivum). New Phytol. 2014; 1–11. https://doi.org/10.1111/nph.12948 PMID: 25078249
19. Diallo A, Ali-Benali M, Badawi M, Houde M, Sarhan F. Expression of vernalization responsive genes in wheat is associated with histone H3 trimethylation. Mol Genet Genomics. 2012; 287: 575–90. https://doi.org/10.1007/s00438-012-0701-0 PMID: 22684814
20. Finnegan E, Dennis E. Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. Curr Biol. 2007; 17: 1978–83. https://doi.org/10.1016/j.cub.2007.10.026 PMID: 17980595
21. Sung S, Schmitz R, Amasino R. A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis. Genes Dev. Cold Spring Harbor Laboratory Press; 2006; 20: 3244–3248. https://doi.org/10.1101/gad.1493306 PMID: 17114575
22. Milec Z, Tomková L, Sumíková T, Pánková K. A new multiplex PCR test for the determination of Vrn-B1 alleles in bread wheat (Triticum aestivum L.). Mol Breed. 2012; 30: 317–323. https://doi.org/10.1007/s11032-011-9621-7
23. Piepho HP. A mixed-model approach to mapping quantitative trait loci in Barley on the basis of multiple environment data. Genetics. 2000; 156: 2043–2050. PMID: 11102394
24. Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, et al. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet. 2006; 113: 1409–1420. https://doi.org/10.1007/s00122-006-0365-4 PMID: 17033786
25. IWGSC, Mayer K. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science. 2014; 345: 1251788. https://doi.org/10.1126/science.1251788 PMID: 25035500
26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367
27. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012; 13: 134. https://doi.org/10.1186/1471-2105-13-134 PMID: 22708584
28. Ivanicova Z, Jakobson I, Reis D, Safar J, Milec Z, Abrouk M, et al. Characterization of new allele influencing flowering time in bread wheat introgressed from Triticum militinae. N Biotechnol. 2016; 33: 718–727. https://doi.org/10.1016/j.nbt.2016.01.008 PMID: 26899284
29. Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S. FISHIS: Fluorescence In Situ Hybridization in Suspension and Chromosome Flow Sorting Made Easy. PLoS One. 2013; 8. https://doi.org/10.1371/journal.pone.0057994 PMID: 23469124

30. Šimková H, Svensson JT, Condamine P, Hribová E, Suchánková P, Bhat PR, et al. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics. 2008; 9: 294. https://doi.org/10.1186/1471-2164-9-294 PMID: 18565235

31. Bullrich L, Appendino M, Tranquilli G, Lewis S, Dubcovsky J. Mapping of a thermo-sensitive earliness per se gene on Triticum monococcum chromosome 1Am. Theor Appl Genet. 2002; 105: 585–593. https://doi.org/10.1007/s00122-002-0982-5 PMID: 12582508

32. Valárik M, Linkiewicz A, Dubcovsky J. A microcolinearity study at the earliness per se gene Eps-A(m)1 region reveals an ancient duplication that preceded the wheat-rice divergence. Theor Appl Genet. 2006; 112: 945–957. https://doi.org/10.1007/s00122-005-0198-6 PMID: 16432738

33. Lewis S, Faricelli ME, Appendino ML, Valárik M, Dubcovsky J. The chromosome region including the earliness per se locus Eps-Am1 affects the duration of early developmental phases and spikelet number in diploid wheat. J Exp Bot. 2008; 59: 3595–607. https://doi.org/10.1093/jxb/ern209 PMID: 18836186

34. Cane K, Eagles H, Laurie D, Trevaskis B, Vailance N, Eastwood R, et al. Ppd—B1 and Ppd—D1 and their effects in southern Australian wheat. Crop Pasture Sci. 2013; 64: 100–114.

35. Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL. Multiply to conquer: Copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genet. BMC Genetics; 2015; 16: 96. https://doi.org/10.1186/s12863-015-0258-0 PMID: 26219856

36. Newell-Price J, Clark AJL, King P. DNA methylation and silencing of gene expression. Trends Endocrinol Metab. 2000; 11: 142–148. https://doi.org/10.1016/S1043-2760(00)00248-4 PMID: 10754536

37. Force a, Force a, Lynch M, Lynch M, Postlethwait J, Postlethwait J. Preservation of duplicate genes by subfunctionalization. Am Zool. 1999; 39: 0. 10101175

38. Kiss T, Balla K, Veisz O, Láng L, Bedő Z, Griffiths S, et al. Allele frequencies in the VRN-A1, VRN-B1 and VRN-D1 vernalization response and PPD-B1 and PPD-D1 photoperiod sensitivity genes, and their effects on heading in a diverse set of wheat cultivars (Triticum aestivum L.). Mol Breed. 2014; 34: 297–310. https://doi.org/10.1007/s11032-014-0034-2 PMID: 25076837

39. Della Vedova C, Cone K. Paramutation: The Chromatin Connection. Plant Cell. 2004; 16: 1358–1364. https://doi.org/10.1105/tpc.160630 PMID: 15178748

40. Pilu R. Seminars in Cell & Developmental Biology Paramutation phenomena in plants. Semin Cell Dev Biol. Elsevier Ltd; 2015; 44: 2–10. https://doi.org/10.1016/j.semcdb.2015.08.015 PMID: 26335267

41. Herman JJ, Spencer HG, Donohue K, Sultan SE. How stable "should" epigenetic modifications be? Insights from adaptive plasticity and bet hedging. Evolution (N Y). 2014; 68: 632–643. https://doi.org/10.1111/evo.12324 PMID: 24274594

APPENDIX III

Kvitnúť či nekvitnúť

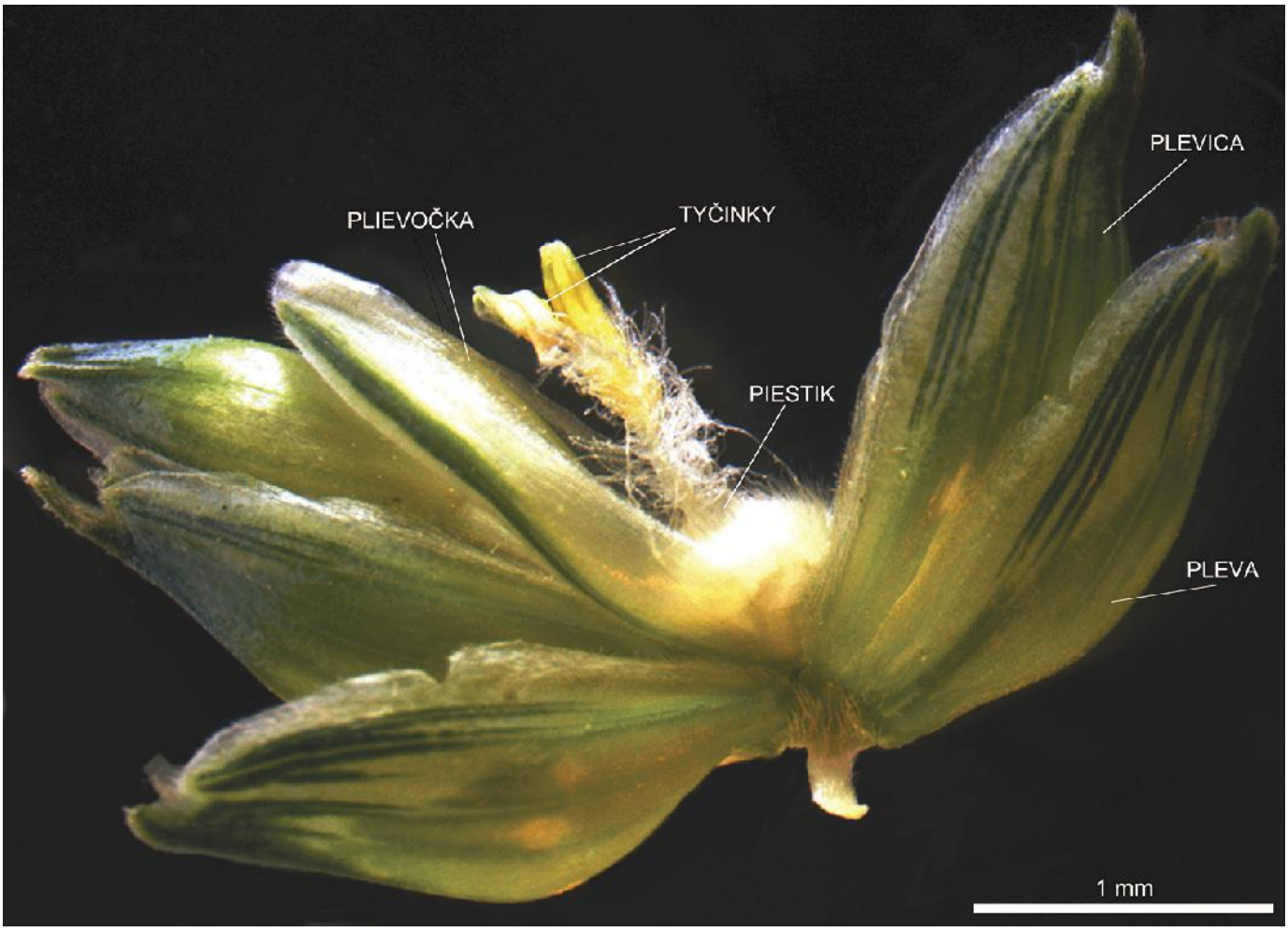
Ivaničová Z.

Svet prírody, 4, 2016

Kvitnúť či nekvitnúť?

Zuzana Ivaničová

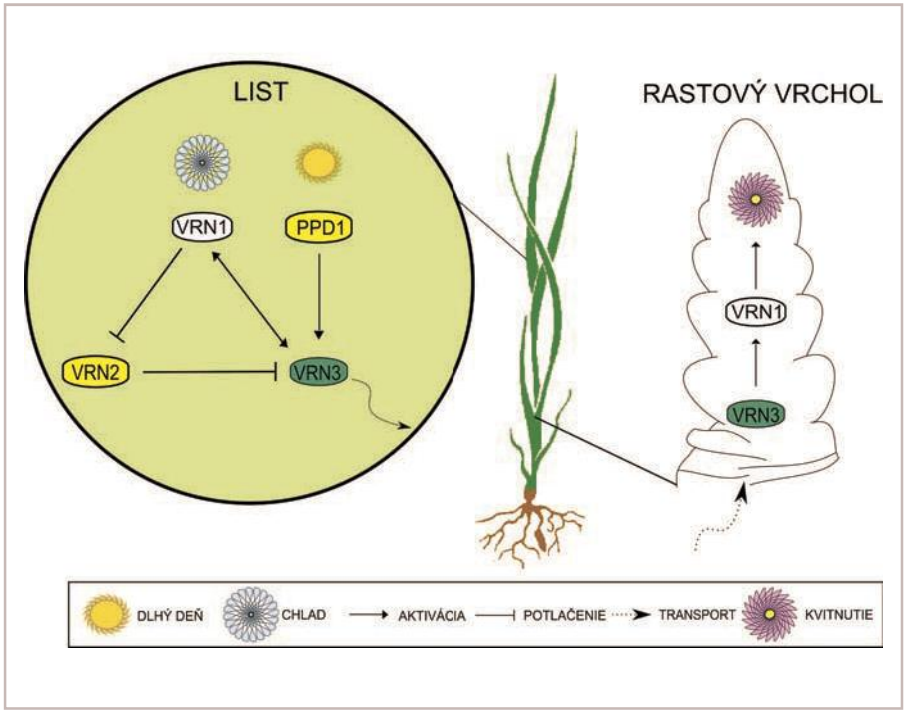
Pšenica je jednou z najdôležitejších hospodárskych plodín. Poskytuje zdroj potravy pre 35 % celosvetovej populácie. V dôsledku neustáleho rastu ľudskej populácie je nevyhnutné zvyšovať jej výnos. Tento cieľ je však ohrozovaný celým radom biotických a abiotických faktorov. V dôsledku zmien klimatických podmienok sme v poslednom čase svedkami značných výkyvov počasia. Jarné mrazy môžu viesť k poškodeniu kvetných častí (obr. 1), suché a horúce letá ohrozujú ukladanie zásobných látok do vyvíjajúcich sa obiliek. Preto je potrebné vysievať pšenice, ktoré budú lepšie prispôsobené lokálnym podmienkam. Ich kvitnutie je potrebné načasovať tak, aby nedošlo k poškodeniu kvetných častí a rastlina optimálne využila dostupné zdroje živín potrebných na vývin a naplnenie semien. Pokiaľ sa rastlina rozhodne kvitnúť v nesprávny čas, nevyprodukuje žiadne semená. Preto pri pšenici, ktorá je jednoročnou rastlinou, je rozhodnutie kvitnúť alebo nekvitnúť otázkou života a smrti.



OBRÁZOK 1. STAVBA KVETU PŠENICE LETNEJ. KVET JE TVORENÝ PIESTIKOM A TROMA TYČINKAMI. Z VNÚTORNEJ STRANY JE CHRÁNENÝ PLIEVOČKOU, Z VONKAJŠEJ PLEVICOU A PLEVOU.

Kvitnutie je teda jedným z najdôležitejších procesov nielen pre rastlinu, ale aj pre človeka. Odhalenie spleti génov, ktoré ho ovplyvňujú, je jedným z cieľov nášho výskumu. Úplné porozumenie tohto deja tak, aby sme ho vedeli cielene ovplyvniť, je však veľmi ťažká úloha. Na určení optimálneho času kvitnutia sa totiž zúčastňuje veľké množstvo génov. Väčšina týchto génov ovplyvňuje proces kvitnutia len v malej miere a až ich súhrou dochádza k iniciácii celého procesu. Zapojenie veľkého počtu génov však umožňuje relatívne presné a spoľahlivé načasovanie kvitnutia v širokej škále prírodných podmienok. Avšak, identifikácia všetkých zúčastnených génov a ich vzájomných interakcií, rovnako aj interakcií s vonkajšími podmienkami prostredia, je zložitý proces. Pri pšenici je táto úloha skomplikovaná veľkosťou a komplexnosťou jej dedičnej informácie (genómu). Pšeničný génom je približne šesťkrát väčší ako u človeka a až na 90 % je tvorený opakujúcimi sa úsekmi DNA. Navyše má pšenica letná za sebou komplikovanú históriu, keďže sa na jej pôvode podieľali tri odlišné druhy pôvodnej pšenice. Jej dedičná informácia je teda zložená z troch príbuzných génomov. Hľadanie jediného génu v tejto spleti by sme mohli porovnať s hľadaním ihly v kope sena až na to, že v tomto prípade sú tie kopy tri.

Vráťme sa však k pôvodnému problému. Akým spôsobom je pšenica schopná určiť, že nastali ideálne podmienky na kvitnutie? Ako dokáže rozlíšiť medzi teplým počasím s dostatočnou vlhkosťou v jesenných a jarných mesiacoch? Ako „predpovedá“, že po zdanlivo výhodných podmienkach na jeseň nasledujú chladné zimné mesiace, zatiaľ čo podmienky po jari sú vhodné na tvorbu ďalšej generácie? Veľmi jednoduchou odpoveďou by bolo: dlhodobým pozorovaním. Pri určovaní vhodnosti obdobia využila jeden dôležitý jav a to, že príchod jarných a letných mesiacov v miernom pásme koreluje s predlžovaním dňa a so zvyšovaním teploty. Naopak, počas jesene sa dĺžka dňa postupne zmenšuje. Chladná zima je následne vystriedaná obdobím s teplotne vhodnými podmienkami. Prekonanie chladného obdobia spolu s predlžujúcim sa dňom je preto dostatočným signálom, že prichádzajú optimálne podmienky na kvitnutie a vývin semien.



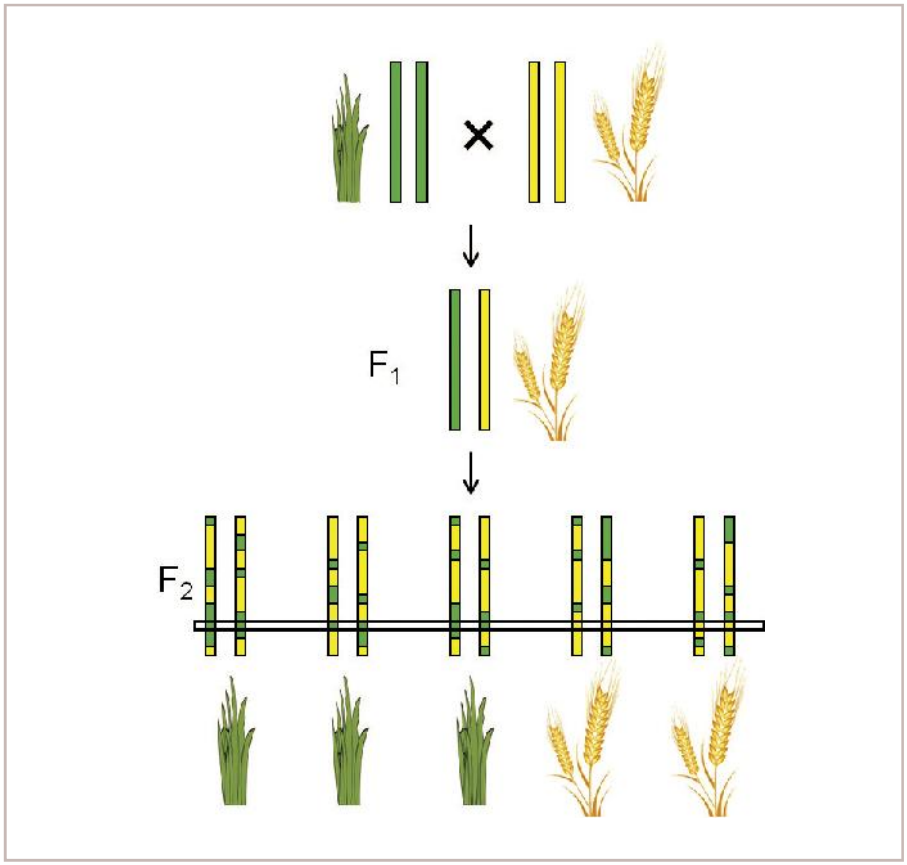
OBRÁZOK 2. MOLEKULÁRNE DRÁHY OVPLYVŇUJÚCE KVITNUTIE PŠENICE LETNEJ. DLHÝ DEŇ A VERNALIZÁCIA SÚ VNÍMANÉ V LISTE. SÚ OVPLYVŇOVANÉ PREDOVŠETKÝM GÉNMI PPD1 A VRN1.TIETO GÉNY NÁSLEDNE VPLÝVAJÚ NA ĎALŠIE ČASTI DRÁHY A V KONEČNOM DÔSLEDKU NA VRN3. VRN1 POTLÁČA NEGATÍVNY EFEKT VRN2 NA VRN3 (FLORIGÉN) A ZÁROVEŇ ZVYŠUJE JEHO EXPRESIU. FLORIGÉN JE TRANSPORTOVANÝ CEZ FLOÉM DO RASTOVÉHO VRCHOLU. V RASTOVOM VRCHOLE UMOŽŇUJE VZNIK VRN1, KTORÝ NÁSLEDNE VEDIE K SPUSTENIU GÉNOV ZODPOVEDNÝCH ZA VÝVIN KVETNÝCH ČASTÍ, A TEDA KVITNUTIE. VRN1 MÁ TEDA V PROCESE KVITNUTIA VIACERO DÔLEŽITÝCH ÚLOH.

Ako rastlina určuje teplotu a dĺžku dňa z genetického hľadiska?

Prekonanie chladného obdobia dokáže pšenica identifikovať vďaka trom hlavným vernalizačným génom (*VRN1*, *VRN2* a *VRN3*). Proces iniciácie kvitnutia chladným obdobím nazývame vernalizácia (jarovizácia) a je potrebný na kvitnutie ozimných pšeníc. Pred zimným obdobím je kvitnutie potlačené inhibíciou funkcie génu *VRN3*. Toto sa deje vďaka funkcii génu *VRN2* – inhibítora kvitnutia. Tieto vplyvy sú odstránené práve pôsobením chladu a krátkych dní počas zimného obdobia. Dlhé obdobie chladu a krátkeho dňa (väčšinou stačí len jeden z nich po obdobie 6 - 10 týždňov) spustí funkciu génu *VRN1* v listoch. Tento gén začne aktivovať gén *VRN3*. *VRN1* však na to sám nestačí, ale potrebuje podporu génov spustených dostatočne dlhým dňom (tým sa zabráni, aby sa kvitnutie spustilo počas dlhej zimy s krátkymi dňami). Vnímanie dĺžky dňa majú na starosti fotoperiodické gény (*PPD*). Predlžujúci sa deň má za následok postupné zvyšovanie hladiny proteínov, ktoré vznikajú prepisom týchto génov. Len čo sa dosiahne určité hraničné množstvo týchto proteínov, začnú ovplyvňovať ďalšie časti dráhy a v spolupráci s *VRN1* spustia produkciu *VRN3*. Keď je hladina *VRN3* dostatočná, natrvalo zablokuje *VRN2*, a tým spustí proces kvitnutia.

Informácie o dĺžke dňa a vhodnosti teplotných podmienok sú teda spracovávané v listoch a integrované prostredníctvom génu *VRN3*. Proteín vznikajúci prepisom tohto génu sa nazýva florigén. Florigén je z listov transportovaný do rastového vrcholu, kde umožňuje zvýšenie prepisu *VRN1*. V rastovom vrchole má *VRN1* odlišnú funkciu ako v listoch. Nie je zodpovedný za vnímanie chladu, ale vedie k spusteniu génov zodpovedných za prechod z vegetatívneho do generatívneho štádia a tvorbu kvetných častí – má pleiotropný účinok (Obr. 2).

Fotoperiodická a vernalizačná dráha, ktoré sú tvorené týmito génmi, sú zodpovedné za „hrubé“ určenie času kvitnutia a schematické znázornenie oboch dráh je uvedené na obrázku č. 2. Za veľmi jemné zmeny v čase kvitnutia sú zodpovedné takzvané gény *Earliness per se*. Tie umožňujú doladenie kvitnutia v závislosti od lokálnych podmienok prostredia nezávisle od vernalizácie a fotoperiódy. Preto sú pre nás veľmi zaujímavé, avšak veľmi malý efekt týchto génov na kvitnutie,



OBRÁZOK 3. SCHÉMA URČENIA PRIBLIŽNEJ POZÍCIE GÉNU ZODPOVEDNÉHO ZA ROZDIEL V ČASE KVITNUTIA. RODIČMI MAPOVACEJ POPULÁCIE JE OZIMNÁ PŠENICA (ZELENÁ) A JARNÁ PŠENICA (ŽLTÁ). FAREBNÉ OBDĹŽNIKY ZODPOVEDAJÚ GENETICKEJ INFORMÁCII. POTOMKOVIA KRÍŽENIA (F1) ZDEDILI POLOVICU GENETICKEJ INFORMÁCIE OD KAŽDÉHO RODIČA. Z GENETICKÉHO HĽADISKA SÚ TOTOŽNÉ, A PRETO MAJÚ AJ ROVNAKÝ ČAS KVITNUTIA. JEDNOTLIVÉ JEDINCE SÚ NÁSLEDNE SAMOOPELENÉ. POČAS TVORBY POHLAVNÝCH BUNIEK DOCHÁDZA K TZV. CROSSING-OVERU - VÝMENE ÚSEKOV MEDZI RODIČOVSKÝMI CHROMOZÓMAMI. JEDNOTLIVÉ CHROMOZÓMY SÚ NÁHODNOU KOMBINÁCIOU ÚSEKOV (ŽLTÉ A ZELENÉ) PÔVODNÝCH RODIČOVSKÝCH CHROMOZÓMOV. POTOMKOVIA (F2) SA ODLIŠUJÚ (SEGREGUJÚ) VO VIACERÝCH ZNAKOCH VRÁTANE ČASU KVITNUTIA. ÚSEK DEDIČNEJ INFORMÁCIE POCHÁDZAJÚCEJ Z OZIMNÉHO RODIČA POTOM BUDE PRÍTOMNÝ U VŠETKÝCH JEDINCOV, KTORÉ BUDÚ NA VYKVITNUTIE VYŽADOVAŤ VERNALIZÁCIU A NAOPAK. TAKÝMTO SPÔSOBOM IDENTIFIKUJEME ČASŤ DNA, V KTOREJ JE PRÍTOMNÝ HĽADANÝ GÉN.

ktorý sa môže pohybovať v rádoch hodín až dní, a jeho závislosť od zmien vonkajšieho prostredia veľmi sťažujú ich identifikáciu. V kontexte prirovnania kôp so senom môže naša ihla miestami pripomínať steblo a nájsť určité steblo v kope sena je takmer nemožné. Avšak napriek malým efektom týchto génov viedli ich rozdielne variácie k vznikom odrôd, ktoré vykvitnú a dozrejú pri odlišných podmienkach. Práve táto flexibilita umožnila pestovanie pšenice od polárneho kruhu až po rovník (67 ° szš po 45 °jzš).

Odlišnosti vo vernalizačných génoch sú zodpovedné za vznik dvoch typov pšeníc: jarín a ozimín. Obdobie chladu je potrebné na kvitnutie ozimných pšeníc, avšak jariny vernalizáciu nepotrebujú. Inhibícia génu *VRN1* bola v prípade týchto pšeníc odstránená, a to v dôsledku zmien v sekvencii DNA jedného alebo viacerých génov *VRN*.

Zmeny v sekvencii DNA fotoperiodických génov taktiež viedli k rozlíšeniu pšeníc: na citlivé a necitlivé na dĺžku dňa. Pri citlivých je kvitnutie zrýchlené vplyvom dlhého dňa. Naopak, necitlivé vykvitnú bez ohľadu na dĺžku dňa. Pri necitlivých odrodách sú fotoperiodické gény funkčné bez ohľadu na dĺžku dňa. Okrem toho môže byť fotoperiodická necitlivosť spôsobená aj zvýšeným počtom kópií týchto génov. Odrody pšeníc necitlivých na dĺžku dňa majú výhodu v oblastiach, kde sa optimálna teplota a vlhkosť prelína s krátkou dĺžkou dňa, čo v niektorých oblastiach sveta umožňuje získanie úrody dvakrát do roka alebo zabraňuje ohrozeniu úrody suchými a horúcimi letnými mesiacmi (napr. Kalifornia, Mexiko, Izrael, Sýria).

Ako sme uviedli na začiatku, je potrebné zvýšiť výnos pšenice, v ideálnom prípade spojenom so zachovaním až zmenšením poľnohospodársky využívanej plochy. To však nie je možné s využitím dnes dostupných odrôd. Vyšľachteniu variet schopných lepšie odolávať nepriaznivým vplyvom počasia predchádza identifikácia génov ovplyvňujúcich žiadané vlastnosti. Princípom je vyhľadávanie odrôd, ktoré sa odlišujú napr. časom kvitnutia a určenie génu, ktorý je za sledovaný znak zodpovedný (Obr. 3). Prvým krokom je ich kríženie a následné kombinovanie ich genetickej informácie. Rôzne kombinácie génov potomkov majú za následok rozdiel v čase kvitnutia. Následne je pri nich potrebné zaznamenať čas kvitnutia a zloženie dedičnej informácie – čiže určiť, ktoré časti sú od jedného, a ktoré od druhého rodiča. Obe informácie nám umožnia identifikovať oblasť genómu, ktorý je prítomný pri všetkých jedincoch so žiadaným časom kvitnutia. V tomto úseku sa s najväčšou pravdepodobnosťou nachádza aj gén, ktorý hľadáme. Ďalším krokom je určenie sekvencie DNA nášho génu a charakterizácia jeho funkcie v procese kvitnutia.

Tento postup navonok vyzerá veľmi jednoducho, avšak nie je vôbec triviálny. Od nájdenia zaujímavého fenotypu (vonkajší znak, v našom prípade čas kvitnutia) po identifikáciu génu môže prejsť niekoľko rokov. Ešte dlhší čas bude trvať vyšľachtenie variety s daným génom a jej následné využitie v poľnohospodárstve.

Je fascinujúce, akým spôsobom dokáže pšenica určiť, kedy nastal najvhodnejší čas kvitnutia, pretože je to rozhodovanie o bytí a nebytí. Zároveň je detailné poznanie génov zúčastňujúcich sa na tomto procesu len ďalším krokom na získanie účinnejších nástrojov použiteľných pri tvorbe odrôd s vyšším výnosom, ktoré budú lepšie prispôsobené lokálnym podmienkam a potrebám ľudstva.

ZUZANA IVANIČOVÁ

je doktorandkou v Ústave experimentálnej botaniky, Centrum regiónu Haná pre biotechnologický a poľnohospodársky výskum, Šlechtitelů 31, CZ-783 71 Olomouc – Holice, Česká republika

APPENDIX IV

Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent

Civáň P., Ivaničová Z. and Brown T.A.

PLoS ONE 8(11), e81955, 2013

OPEN ACCESS Freely available online

PLOS ONE

Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent

Peter Civáň[1,2,3], Zuzana Ivaničová[2,4], Terence A. Brown[1*]

1 Manchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom, **2** Department of Genetics, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia, **3** Centre of Marine Sciences, University of Algarve, Faro, Portugal, **4** Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic

Abstract

We used supernetworks with datasets of nuclear gene sequences and novel markers detecting retrotransposon insertions in ribosomal DNA loci to reassess the evolutionary relationships among tetraploid wheats. We show that domesticated emmer has a reticulated genetic ancestry, sharing phylogenetic signals with wild populations from all parts of the wild range. The extent of the genetic reticulation cannot be explained by post-domestication gene flow between cultivated emmer and wild plants, and the phylogenetic relationships among tetraploid wheats are incompatible with simple linear descent of the domesticates from a single wild population. A more parsimonious explanation of the data is that domesticated emmer originates from a hybridized population of different wild lineages. The observed diversity and reticulation patterns indicate that wild emmer evolved in the southern Levant, and that the wild emmer populations in south-eastern Turkey and the Zagros Mountains are relatively recent reticulate descendants of a subset of the Levantine wild populations. Based on our results we propose a new model for the emergence of domesticated emmer. During a pre-domestication period, diverse wild populations were collected from a large area west of the Euphrates and cultivated in mixed stands. Within these cultivated stands, hybridization gave rise to lineages displaying reticulated genealogical relationships with their ancestral populations. Gradual movement of early farmers out of the Levant introduced the pre-domesticated reticulated lineages to the northern and eastern parts of the Fertile Crescent, giving rise to the local wild populations but also facilitating fixation of domestication traits. Our model is consistent with the protracted and dispersed transition to agriculture indicated by the archaeobotanical evidence, and also with previous genetic data affiliating domesticated emmer with the wild populations in southeast Turkey. Unlike other protracted models, we assume that humans played an intuitive role throughout the process.

Citation: Civáň P, Ivaničová Z, Brown TA (2013) Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent. PLoS ONE 8(11): e81955. doi:10.1371/journal.pone.0081955

Editor: Randall P. Niedz, United States Department of Agriculture, United States of America

Received May 10, 2013; **Accepted** October 18, 2013; **Published** November 29, 2013

Funding: The work was supported by the following: Natural Environment Research Council NE/E015948/1 www.nerc.ac.uk>; and Slovak Research and Development Agency APVV-0661-10, APVV-0197-10 http://www.apvv.sk/agentura?lang=sk>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: terry.brown@manchester.ac.uk

Introduction

Shortly after the Younger Dryas (12,800–11,600 BP) – the closing cold and dry echo of the last glaciation – the nomadic hunter-gatherer communities of southwest Asia adopted a sedentary lifestyle. The reasons for this cultural innovation and the accompanying changes in subsistence strategy have been widely debated, with underlying causes sought among factors as diverse as labor productivity [1], climatic response [2,3], predator-prey relationships [4,5], human intuition [6] and a changing human worldview [7]. Whatever the drivers, the beginning of agriculture was a central component of the set of changes associated with the Neolithic, and is viewed as the major transition in the human past, the period when humans first began to exert a degree of control over their food resources [8].

Agricultural origins in southwest Asia are traditionally associated with eight founder crops including three cereals, einkorn wheat (*Triticum monococcum* L.), emmer wheat (*T. turgidum* L.) and barley (*Hordeum vulgare* L.) [9]. Archaeobotanical and genetic analysis of these crops, especially the cereals, is increasingly being used as a means of studying the human dimension to the adoption of agriculture [8]. Initially, much of this work was influenced by experimental

studies which showed that if appropriate husbandry practices were applied, then the period required for a wild cereal to undergo the suite the phenotypic changes associated with domestication might be as short as a few decades [10]. The attractive idea that a single group of enlightened people could have been responsible for the domestication of one or more staple crops within a few human generations [11] was supported by the first comprehensive genetic comparison of wild and cultivated cereal genotypes [12], which was interpreted as indicating a rapid domestication of einkorn in the Karaca Dağ region of southeast Turkey [6,13,14]. Attempts to extend this rapid, localized model to other crops were initially successful [15], but the paradigm was challenged by computer simulations which showed that the tree-building algorithms used to analyze the genetic datasets could not distinguish crops that are truly monophyletic from ones resulting from multiple independent domestications [16,17]. Archaeological research also began to provide conflicting evidence in the form of archaeobotanical data suggesting that cereal domestication was a protracted process that began with a lengthy period of wild plant management before a slow and piecemeal emergence of the domestication phenotypes, the whole process taking several millennia [8,18,19].

The conflict between these opposing views of the origins of agriculture is exemplified by the work carried out with emmer and other tetraploid wheats. In the first major genetic study, Özkan et al. [20] used distance-based tree building to compare variations at 204 amplified fragment length polymorphisms (AFLPs) in 43 domesticated lines and 99 wild populations, and identified a single origin for tetraploid wheat domestication near Karaca Dağ. A subsequent examination of chloroplast microsatellite haplotypes, including accessions from areas neglected in the AFLP study, concluded that emmer was domesticated in the northwestern edge of the Fertile Crescent (referred to as Kartal Dağı by the authors), some 250 km west of Karaca Dağ [21]. However, two distinct chloroplast lineages were identified in the domesticated plants, suggesting at least a biphyletic origin. AFLPs were then analyzed in the additional wild accessions [22], but the results did not confirm the chloroplast data and instead located the closest wild relatives of domesticated emmer in the Karaca Dağ and Sulaymaniyah (Iraq/Iran border) regions. Luo et al. [23] attempted to solve the puzzle by analyzing restriction fragment length polymorphisms (RFLPs) in 227 wild and 241 domestic tetraploid wheats. A significant proportion of the domesticated lines showed equally strong genetic affinity with wild populations from the Diyarbakir region and southern Levant, leading the authors to conclude that emmer was either domesticated independently in these two regions, or was domesticated in the Diyarbakir region and subsequently acquired additional diversity by gene flow from wild populations in the southern Levant and other parts of the Fertile Crescent. The latter model was supported by a reanalysis of the AFLP data [24], but the former – separate domestications in Turkey and the Levant – agrees with the outcome of an independent study of glutenin alleles in 185 domestic and 59 wild tetraploid wheats [25].

The contradictory scenarios arising from the genetic analyses contrast with the outcomes of archaeobotanical studies. Preserved emmer spikelets with rough abscission scars indicative of nonshattering ears, looked on as the key domestication phenotype [26], appear simultaneously in the pre-pottery Neolithic B (PPNB, c.10,000 BP) layers of archaeological sites in the southern Levant (Jordan Valley), northern Syria and southeast Turkey [15,24,27,28]. This first emergence of the domestication phenotype was, however, merely one step in the process that led to the fully domesticated crop [8]. For at least 1000 years previously, wild emmer had been cultivated in both the southern and northern Levant [18,29], as revealed by stored assemblages that contain the seeds of weeds associated with arable cultivation. Furthermore, after their first appearance in the archaeobotanical record, the domestication traits rise to dominance only slowly, with different phenotypes following independent dynamics over a period of some 3000 years, in parallel in different parts of the Fertile Crescent [18,26,27,30].

The above summary of the outcomes of research into the origins of domesticated emmer raises a question which, put bluntly, is why do different genetic analyses of a single crop give such inconsistent results, and why do none of these results agree with the archaeobotanical evidence? Part of the problem lies with the assumption, implicit in the use of phylogenetic methods to analyze genetic data from modern crops, that the evolution of those crops since domestication has been treelike, when in reality there is likely to have been gene flow and hybridization between different crop lineages [16–18]. A second issue that has been less explored is the possibility that domesticated crops have a reticulate rather than linear relationship with their wild progenitor populations. Reticulation refers to the pattern arising when different parts of a genome have different genealogical histories due, for example, to introgression, incomplete lineage sorting (syn. deep coalescence), or hybrid speciation [31]. If a genetic dataset contains incongruent signals resulting from these processes, then a network rather than a tree is a more appropriate representation of the genealogy [32,33]. Should a dataset used to study the origins of domesticated emmer contain such incongruent signals, then these will be suppressed if forced into a single tree, which will show only a single scenario of a pseudo-divergent genealogy. The tree will therefore hide the incompatible signals and not provide the correct interpretation of the domestication process, and different sets of accessions and genetic markers will yield different phylogenetic results.

To re-assess the origins of domesticated emmer we developed a novel typing method based on detection of DNA polymorphisms associated with the insertion of long terminal repeat (LTR) retrotransposons in the repetitive 5S and 5.8S ribosomal RNA (rRNA) gene arrays. Arrays of 5S rRNA gene-spacer units are located on homeologous wheat chromosomes 1 and 5 [34], each array containing thousands of units [35]. The 5.8S rRNA genes lie within the main multicopy rDNA arrays, which are located independently of the 5S arrays on wheat chromosomes 1A, 1B and 6B [36]. Individual rDNA units are not subject to selective pressure [37], allowing the accumulation of mutations including transposable element insertions.

To obtain independent evidence regarding the relationship between wild and domesticated emmer, we also re-examined previously-reported sequence data [38] for 21 loci in tetraploid wheats. Based on a tree constructed from the concatenated sequence data matrix, the previous report concluded that domesticated emmer has a monophyletic origin. We show that this conclusion lacks statistical support because of extensive gene-tree conflicts. With both the retrotransposon and nuclear gene datasets we examine the scale of the phylogenetic incongruence with the aid of filtered supernetworks [52] and interpret the phylogeographical data with respect to the revealed reticulation. The results enable us to propose a dynamic model for agricultural origins based on a human driven dispersal of wild plants prior to domestication. The model offers an explanation for the observed patterns of diversity and reticulation, is consisted with the archaeological evidence for domestication as a protracted and dispersed process, and assigns an active role to the early farmers in shaping the geographic distribution and genetic constitution of emmer wheat.

Materials and Methods

Wheat Samples and DNA Extraction

The sample set (Table A in File S1) comprised 227 accessions including tetraploid wheats with both the BA^u and GA^u genome constitutions. The former, covering all geographical regions, were 70 wild emmers (*T. turgidum* L. subsp. *dicoccoides* (Korn. ex Asch. & Graebn.) Thell.), 99 hulled domesticates (*T. turgidum* L. subsp. *dicoccum* (Schrank ex Schübl.) Thell., *T. ispahanicum* Heslot, *Triticum turgidum* L. subsp. *paleocolchicum* Á. & D. Löve), and 36 free-threshing domesticates including *T. turgidum* subsp. *durum* (Desf.) Husn. and examples of rarely analyzed ancient subspecies of *T. turgidum* (subsp. *carthlicum* (Nevski) Á. & D. Löve, subsp. *turanicum* (Jakubz.) Á. & D. Löve, subsp. *turgidum*, subsp. *polonicum* (L.) Thell.). The GA^u tetraploids comprised 16 wild accessions (*Triticum timopheevii* (Zhuk.) Zhuk. subsp. *armeniacum* (Jakubz.) Slageren) and 6 domesticates (*Triticum timopheevii* (Zhuk.) Zhuk. subsp. *timopheevii*). Maps depicting the geographical locations of accessions were drawn with ArcGIS 10 (ESRI).

To extract DNA, 2–5 grains were crushed to powder and mixed with approximately 400 µl extraction buffer (100 mM Tris-HCl, 20 mM Na_2EDTA, 1.4 M NaCl, 2% cetyl trimethylammonium bromide, 0.3% (v/v) β-mercaptoethanol) in a 2.0 ml tube. After 1 h incubation at 65°C, samples were centrifuged for 2 min at 13,000×*g* and the supernatant mixed in a 1:5 ratio with binding buffer (High Pure PCR Product Purification Kit, Roche). Extracts were purified according to the manufacturer's instructions and DNA quality and quantity assessed by electrophoresis in 1% agarose gels.

Genotyping of LTR Retrotransposon Insertions in rRNA Gene Arrays

PCR was used to detect polymorphic LTR retrotransposon insertions in 5S rRNA arrays, using combinations of primers specific to the 5S rRNA gene repeat and the LTRs of different classes of wheat retrotransposon (Figure 1). The 5S primers were designed by identifying conserved sequences in the wheat 5S genes present in GenBank, avoiding possible cross-annealing to *Cassandra* retrotransposons which carry a 5S gene-like sequence in their LTRs [39]. The LTR primers were designed from wheat retrotransposons present in the TREP database [40]. LTR sequences were aligned and family-specific annealing sites identified for outward facing primers. A similar method was used to detect retrotransposon insertions in the vicinity of 5.8S rRNA genes within the main rDNA arrays. Primer design was aided by FastPCR [41] and Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/); primer sequences are listed in Table B in File S1. Preliminary tests on a small set of emmer samples showed that PCRs with LTR primers specific for the *Jeli* and *BARE1*/*Wis*/*Angela* retrotransposon families gave polymorphic products of the expected size range. Since it is known that the *Jeli* and *Angela* families have been actively transposing since the formation of tetraploid wheat [42], the latter primer was further adjusted to detect only *Angela* retrotransposons. The specificity of PCRs with this primer was checked by sequencing products of different lengths obtained from several einkorn samples.

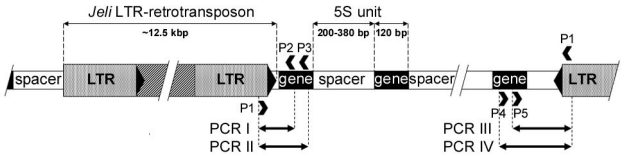


Figure 1. PCR system for detection of LTR retrotransposon insertions in the 5S and 5.8S rDNA loci. The example shown is for detection of a Jeli insertion in a 5S array. Primer P1 is specific for the distal region of the LTR, and primers P2-P5 anneal at different positions with the 5S gene. Depending on its orientation, a Jeli sequence is detected by PCRs with primer combinations P1-P2 and P1-P3, or P1-P4 and P1-P5. Two different PCRs are carried out for each detection to reduce false-positive results. A similar strategy is used to detect other types of LTR retrotransposon and to identify insertions adjacent to 5.8S genes.

doi: 10.1371/journal.pone.0081955.g001

The tetraploid sample set was genotyped with six primer combinations targeting *Jeli* and *Angela* insertions (Table B in File S1) in 12.5 µl PCRs containing 1 x reaction buffer, 1.75 mM $MgCl_2$, 0.25 µM each primer, 0.2 mM each dNTP, 0.025 U/µl AmpliTaq Gold DNA polymerase (Applied Biosystems) and ~30 ng template DNA. Cycling conditions were: 5 min at 95°C; 35 cycles of 45 s at 95°C, 45 s at 57°C, 30 s at 72°C; 7 min at 72°C. All PCR products were analyzed by electrophoresis for 1.5 h at 5V/cm in 1.75% agarose gels; the products of *Jeli* PCR-screening were additionally genotyped using an Applied Biosystems 3730 DNA Analyzer with 500 LIZ size standard to resolve alleles of similar size.

Products of the *Jeli*–5.8S PCR, which were of uniform size (~720 bp), were sequenced (Macrogen) for 186 accessions (GenBank accession numbers JX470351-JX470368). These sequences were aligned with Geneious 5.1.7 [43] and a MJ

network constructed in Network 4.6 [44]. Each *Jeli*-5S and *Angela*-5S amplicon of distinct size was regarded as an independent insertion. Insertions detected in only one sample were discarded as phylogenetically uninformative, and a few markers that were difficult to score (due to poor amplicon synthesis for some accessions) were omitted. The remaining insertions were depicted as a virtual gel in Geneious.

The sequence and binary retrotransposon markers were also analyzed in combination. The *Jeli*-5.8S alignment was dissolved into individual polymorphic positions (splits) and all such splits imported into SplitsTree4 [45] as newick tree vectors. Individual binary markers of the corresponding sample set were loaded into SplitsTree4 in the same way and the resulting set of 43 splits used to construct a filtered supernetwork [32]. The minSupportingTrees parameter was set to 31, meaning that only the splits congruent with >31 other splits are shown in the network. This reduction of network distortion is expected to filter out conflicting 'noise' resulting from splits which may be present in the input datasets due to homoplasy and deep coalescence [33,46]. Other parameters of the supernetwork construction were left at the default values.

Supernetwork Analysis of Published Sequence Data

Sequence data from 21 loci (*11B*, *91A*, *AapA*, *AlperA*, *Bp3B*, *Bp2A*, *Bp5A*, *ChsA*, *Gsp1A*, *Gsp1B*, *HgA*, *HiplA*, *MdhA*, *Mdh4B*, *Mp7A*, *MybA*, *MybB*, *NrpA*, *PsyA*, *ZdsB*) of tetraploid wheats (*T. turgidum* subsp. *dicoccoides*, *dicoccum*, *durum*; *T. timopheevii*) published by Haudry et al. [38] were downloaded from GenBank, recoded according to geographic origin (Table C in File S1) and aligned in Geneious. The sequence alignments were edited by extracting polymorphic positions, removing sequences or columns with multiple unrecognized bases, and reducing indels >1 bp to single positions. A concatenated data matrix was created from all 21 genes and the most parsimonious (MP) tree was searched in dnapars of PHYLIP 3.67 [47] with 10 jumble runs and the remaining options at the default values. A bootstrap analysis with 1000 resampled datasets was also conducted and a majority-rule consensus tree was constructed. The same approach was used to construct an MP tree for each individual locus. If several equally parsimonious trees were identified for a given gene, a strict consensus tree was constructed. For five loci (*11B*, *91A*, *Mdh4B*, *MybB*, *NrpA*), outgroup data (*T. timopheevii*) were not available. The topology of the trees for these loci could therefore be misleading and confound the supernetwork construction by introducing false incongruence among the gene trees. Therefore, a reduced dataset of 15 loci – omitting the five loci with no outgroup as well as the *ChsA* locus (extensive missing data) – was processed alongside the full dataset consisting of all 21 loci. From the resulting sets of 21 and 15 partial trees (representing 28–65 accessions and 37–65 accessions, respectively) filtered supernetworks were produced in SplitsTree4. The minimal set of partial trees required to support a split (minSupportingTrees parameter) was adjusted to eight and six in the full and reduced datasets, respectively. The remaining settings were left as default. The partial trees from the reduced dataset were also analysed in pairs with the autumn algorithm implemented in Dendroscope 3 [48] to identify conflicting tree pairs and the minimal number of hybridizations required to explain the given conflicts.

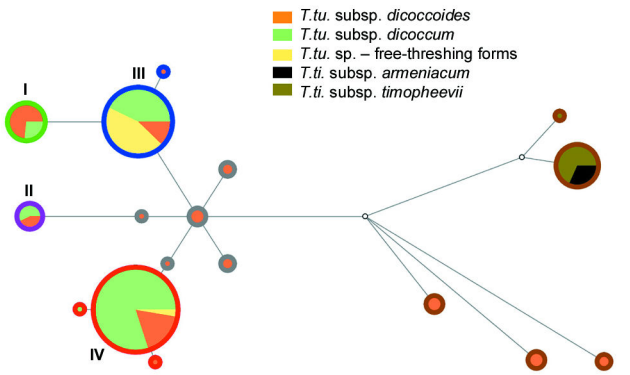


Figure 2. MJ network constructed from 5.8S-*Jeli* sequence data. Node sizes are proportional to the number of accessions displaying that allele, and the edge lengths are proportional to the number of substitutions between pairs of allele sequences. The taxonomic content of each node is indicated as a pie chart. The color coding of the outer circle of each node relates to the symbols used for different groups of accessions in Figure A in File S1.

doi: 10.1371/journal.pone.0081955.g002

Results

Jeli–5.8S Amplicon Sequences

We typed polymorphic retrotransposon insertions in a collection of tetraploid wheats (Table A in File S1) by carrying out PCRs with combinations of primers that targeted conserved regions within the 5S and 5.8S rRNA genes and the LTRs of different groups of wheat retrotransposons (Figure 1, Table B in File S1). The PCR system comprising one primer specific for the 5.8S gene and one specific for the LTR of the *Jeli* group of retrotransposons gave a product of uniform size (~720 bp), which was sequenced for each of 186 accessions. The first 55 bp of this sequence aligns with the wheat 5.8S gene and the remainder appears to be an atypical LTR sequence.

Seventeen alleles were identified in the 186 accessions ($\pi = 5.68 \times 10^{-3}$). In a median joining (MJ) network those alleles shared between wild and domesticated forms of *T. turgidum* fell into four distinct groups (clusters I–IV), which made three independent connections with the remainder of the network (Figure 2). Wild emmers with alleles from the smaller clusters I and II are geographically localized in the southern Levant and the NW Fertile Crescent, respectively (Figure A part A in File S1). Those from the larger clusters are geographically dispersed, from the Levant to Iraq/Iran (cluster III), and from the Jordan Valley to Diyarbakir region (cluster IV) (Figure A part B in File S1).

The MJ network topology located ten alleles outside of clusters I–IV and hence absent in domesticated emmer. These alleles could be divided in two groups of five, one group comprising alleles basal to the domesticate clusters (circled in

gray in Figure 2), and the second group made of more 'early-diverging' alleles (circled in brown), the latter including two alleles identified only in *T. timopheevii*. Accessions containing the early-diverging alleles were broadly distributed, with *T. turgidum* in the west arm and *T. timopheevii* in the east arm of the Fertile Crescent, but those *T. turgidum* accessions with basal alleles were restricted to the southern Levant (Figure A part C in File S1). Of the 14 sequence types found in *T. turgidum* subsp. *dicoccoides*, only two were detected in wild emmer from southeast Turkey or Iraq/Iran, and neither of these two alleles were unique to these locations (Figure A part B in File S1).

Angela- and Jeli–5S Binary Data

We identified eleven distinct insertional polymorphisms of *Angela* and *Jeli* retrotransposons in 5S arrays of tetraploid wheat (designated as *A-li-100*, *A-li-174*, *J-Id-158*, *J-Id-184*, *J-Id-236*, *J-Id-323*, *J-li-47*, *J-li-66*, *J-li-131*, *J-li-272*, and *J-li-447*; Figure 3). None of these insertions were present in the *T. timopheevii* and *T. turgidum* accessions with early-diverging *Jeli*–5.8S alleles, implying that the invasion of *Angela* and *Jeli* retrotransposons into 5S arrays has occurred since the emergence of BA^u tetraploids, less than 0.5 MYA [49,50]. Typically, an accession gave 1–4 amplicons (<550 bp) with each primer combination (see Figure 1). Two of the *Jeli*–5S insertions (*J-li-66* and *J-Id-236*) were present only in wild emmer, and two others (*J-Id-184* and *J-li-447*) only in domesticated accessions. The latter presumably originated <10,000 years ago, indicating that *Jeli* retrotransposons are still active in the wheat genome [42]. Among the shared insertions, *J-li-272* was present in 71% of the hulled domesticates and 33% of the free-threshing ones, but only in five wild emmer accessions, three of these from the Iranian Zagros mountains. The other shared insertions were present in wild samples from all over the Fertile Crescent (*A-li-100*, *J-Id-158*, *J-li-131*), or appeared concentrated to the west of the Euphrates (*J-Id-323*, *J-li-47*, *A-li-100*). Again, the wild emmers from southeast Turkey, Iraq and Iran posses only a subset of variability found in the Levant, missing the markers *J-li-47*, *A-li-100* and *J-Id-236*.

When insertion profiles were considered, clear distinctions were seen between wild and domesticated accessions (Figure 3). For example, the individual insertions *A-li-100*, *J-li-47*, *J-Id-158*, *J-li-272*, and *J-Id-323* were present in both wild and domesticated emmer, but the combinations *A-li-100/J-li-272, J-Id-158/J-Id-323* and *J-li-47/J-li272*, although frequent in domesticated emmer (11%, 11%, and 9%, respectively), were not found in wild accessions. The insertion profiles distinguished between the five subspecies of naked tetraploids, although with some signs of gene flow, and divided the hulled emmer domesticates into five broad phylogeographic groups, described as 'North-eastern', 'Iranian *dicoccum*/*ispahanicum*', '*abyssinicum*, *serbicum* and south-eastern', 'Inland' and 'Mediterranean'.

Combined Retrotransposon Data

The *Jeli*–5.8S, *Angela*–5S and *Jeli*–5S datasets were combined and conflicting phylogenetic signals analyzed by

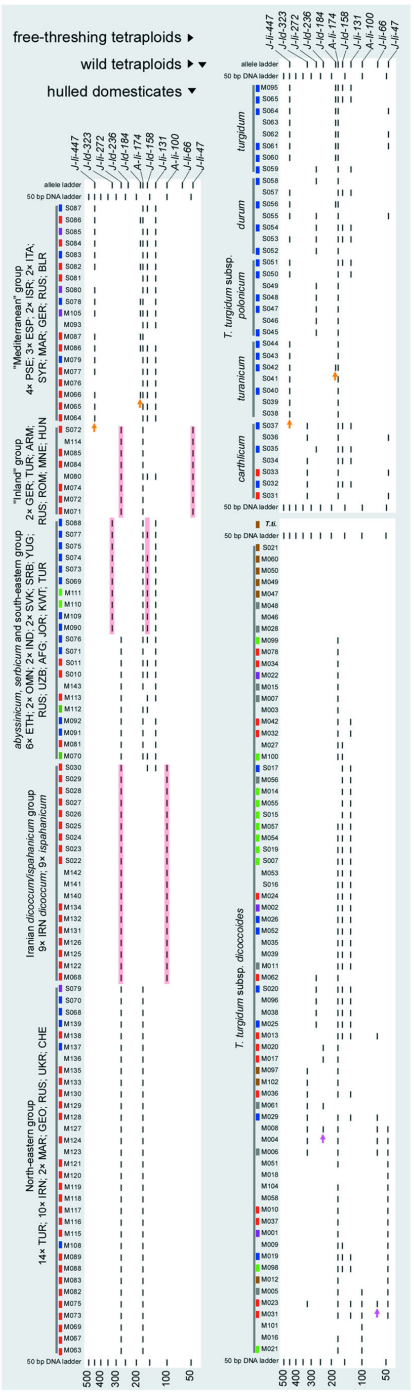


Figure 3. Virtual gel of the retrotransposon insertions detected in the 5S arrays. Accessions are identified by the codes given in Table A in File S1. All *T.timopheevii* accessions displayed the same profile marked here as *T.ti*. Pink and orange arrows indicate insertions detected only in wild and domesticated tetraploid wheats, respectively. Combinations absent in wild populations but frequent in domesticated accessions are highlighted with a pink background. Accessions for which 5.8S-*Jeli* sequences were obtained are color-coded in the same way as the node outer circles in Figure 2.

doi: 10.1371/journal.pone.0081955.g003

constructing a filtered supernetwork (Figure 4). The part of the network containing the *T. timopheevii* samples and the early diverging wild emmers was free of reticulation. The majority of wild emmers were included in the reticulated part of the network, together with three clusters of domesticated samples. The first cluster (cluster A) contained all but two of the free-threshing samples (exceptions being two *T. turgidum* subsp. *carthlicum* accessions placed in cluster C) suggesting a single origin for the *durum*, *turanicum, turgidum* and *polonicum* subspecies, in accordance with the reported reduction of nucleotide diversity in the *dicoccum-durum* line [38]. Cluster A also contained hulled tetraploids, mainly *dicoccum* samples from Ethiopia, Turkey, India and Oman as well as the *serbicum* varieties. Most of the wild emmers in cluster A came from the southern Levant and Iraq/Iran, with one accession from Karaca Dağ. The second cluster (B) contained four geographically dispersed *dicoccum* samples closely associated with wild emmer from the Gaziantep region. The third cluster (C) included hulled domesticates from Iran and Transcaucasia, as well as most of the *dicoccum* samples from the Mediterranean and inland Europe. Affiliated wild samples mostly came from the Karaca Dağ region, but also from the Levant. These three clusters were interconnected through deeper hybridization links originating within a core of Levantine wild emmers.

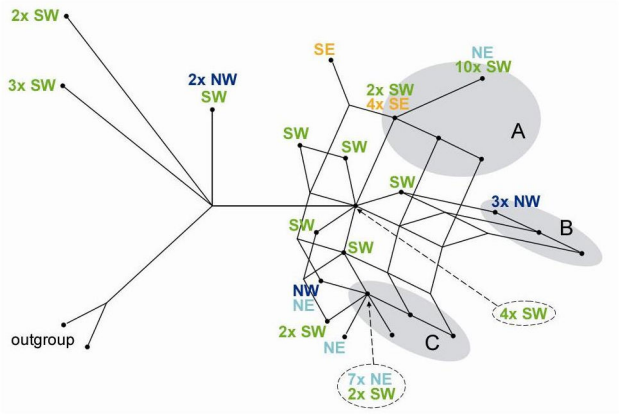


Figure 4. Supernetwork of combined retrotransposon data. Network nodes containing samples are marked by black dots. The geographic locations of the wild emmer accessions are indicated next to the corresponding nodes as follows: *SW* (green), southern Levant (Israel, Lebanon, Jordan, SW Syria); *NW* (dark blue), northwest (western part of the Syria–Turkey border, here approximated as 'Gaziantep region'); *NE* (light blue), northeast (vicinity of Karaca Dağ); *SE* (orange), southeast (parts of Iraq and Iran). Domesticated emmer is found in the nodes within the grey background areas. Cluster A includes all but two free-threshing samples (22) together with *dicoccum* accessions from Ethiopia (6), Turkey (5), Oman (2), India (2), Slovakia (2), former Yugoslavia (1), Morocco (1), Palestine (1) and Germany (1). Cluster B includes four *dicoccum* samples (Palestine, Morocco, Turkey and Scandinavia). Cluster C includes all Iranian *dicoccum* (15) and *T. ispahanicum* (8) accessions, *dicoccum* samples from Turkey (8), Russia (4), Spain (3), Italy (2), Hungary (2), Palestine (2), Jordan (2), Germany (2), Switzerland (1), Armenia (1), Georgia (1), Ukraine (1), Belarus (1), Romania (1), Serbia (1), Israel (1), Morocco (1), Eritrea (1), Uzbekistan (1), and two free-threshing samples. The outgroup comprises the *T. timopheevii* accessions.

doi: 10.1371/journal.pone.0081955.g004

Reanalysis of Available Sequence Data

To obtain independent evidence regarding the relationship between wild and domesticated emmer, we re-examined previously-reported sequence data [38] for 21 loci in *T. turgidum* subsp. *dicoccoides* and *T. turgidum* subsp. *dicoccum* (Table C in File S1). Sequence alignments revealed 55 sequence types for the 21 loci in the domesticated emmers, of which 37 were also identified in wild accessions. As some of the 18 remaining, apparently *dicoccum*-specific sequences might be present in unsampled wild populations, it seems likely that post-domestication divergence has played only a minor role in generating diversity. In most cases, therefore, allelic sequences in domesticated emmer are identical to those in wild emmer, but they often appear in different combinations, a clear sign of reticulation. For example, the EF108894 allele of the *GdhA* gene (which is present in domesticates and wild samples west of the Euphrates) is found with the *Bp2A* (EF108668) allele in six of 12 domesticates, and with the *PsyA* (EF115015) allele in five of these 12, but these combinations are not present in any of the wild accessions. Similarly, one allele of *ZdsB* (EF115121; found in domesticates and wild emmers from Israel) combines with the *MdhA* (EF109064), *Mp7A* (EF109521), *GdhA* (EF108895) and *PsyA* (EF115015) alleles in four, three, three and three *dicoccum* samples, respectively, but none of these combinations were seen in wild emmer.

Only three loci (*AapA*, *Bp5A* and *MdhA*) were monomorphic in domesticated emmer (Table D in File S1). For two of these (*AapA*, *MdhA*) the apparent monomorphism was associated with low overall genetic diversity. Each of the remaining 18 loci displayed two or more sequence types in the domesticates.

The MP analysis of the concatenated data matrix utilized 218 parsimony-informative characters and identified a single minimal tree with *durum* and *dicoccum* accessions forming a monophyletic group (Figure B in File S1), similar to that previously reported [38]. However, the score of the most parsimonious tree (652) was 1.73 × higher than the sum of the MP trees computed for each locus individually (377), which is symptomatic of gene-tree conflicts. Subsequent inspection of the 15 rooted gene-trees with the autumn algorithm revealed that 84 out of the 105 possible tree-pairs contained one or more conflicts. On average 3.1 hybridization events per conflicting tree-pair are necessary to explain the observed data. The bootstrap analysis of the concatenated data did not provide any statistical support for monophyletic domestication (<50%; Figure B in File S1). As the portion of the phylogenetically informative characters would seems to be sufficient for resolution of the major clades, and sequence homoplasy is unlikely to play a significant role within the studied evolutionary time-span, we conclude that the absence

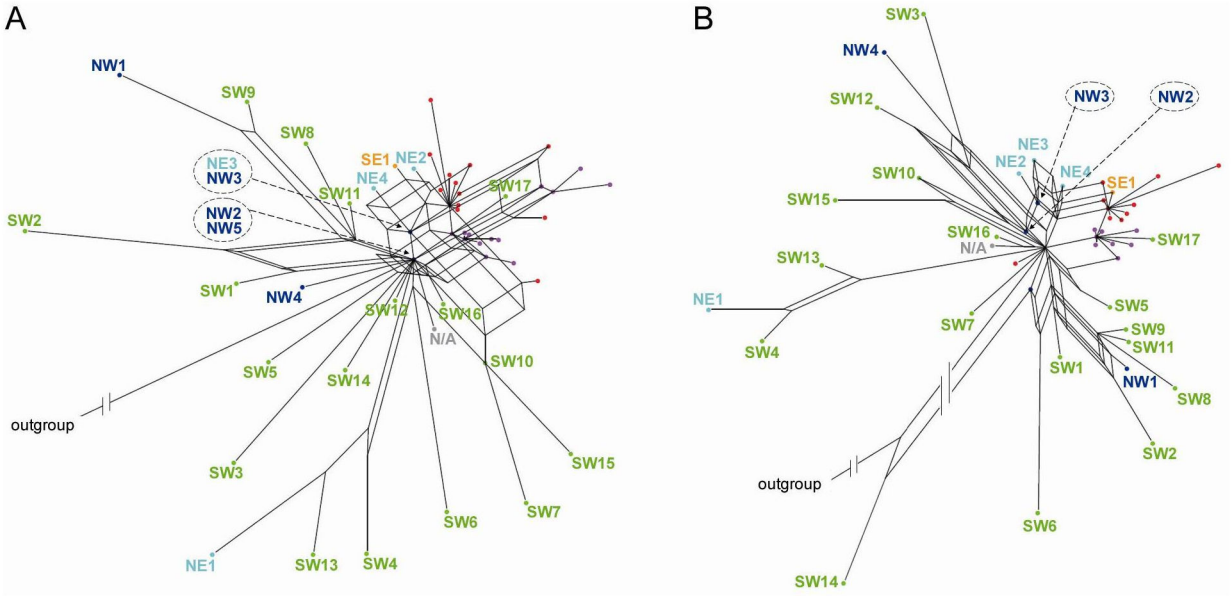


Figure 5. Filtered supernetworks constructed from multiple nuclear gene sequences. [38]. (A) Full dataset composed of 21 partial trees, and (B) a reduced dataset containing only the 15 rooted partial trees, for sequences from wild emmer (28 accessions, dots color coded according to region, *N/A*, location not available), domesticated emmer (12 accessions, red dots) and *durum* wheat (20 accessions, purple dots). Individual nodes may contain multiple samples. The geographic locations of the wild emmer accessions are indicated in accordance with Figure 4. The outgroup comprises the *T. timopheevii* accessions.

doi: 10.1371/journal.pone.0081955.g005

of statistical support is caused by genetic reticulation of emmer lineages.

Filtered supernetworks constructed from these data (Figure 5) confirmed the reticulated origin of *dicoccum* and *durum* wheat. In both supernetworks, *T. turgidum* subsp. *dicoccum* and *durum* accessions are concentrated in two, partially overlapping clusters that mostly consisted of terminal nodes radiating out of reticulated tangles. Similar to the retrotransposon supernetwork (Figure 4), the free-threshing wheats cluster separately from the main *dicoccum* group. The nodes basal to these tangles were typically represented by northern wild emmers and by the central node and other empty nodes related to some of the Levantine wild emmers. The Iranian wild emmer accession (SE1) and three Karaca Dağ samples (NE2, NE3, NE4) also appear to have a highly reticulated origin and occupy terminal rather than basal positions of the tangle associated with the domesticates. This network topology suggests that the eastern wild samples (SE and NE) and domesticated emmer have partially shared hybridization histories.

Discussion

Reticulated Ancestry of Domesticated Emmer

In this study, we identified both sequence and positional polymorphisms for LTR retrotransposon insertions in the 5S and 5.8S arrays of a large set of tetraploid wheat accessions. Size homoplasy of the detected insertion markers is expected to be minimal because it is unlikely that two retrotransposons would insert at exactly the same position in different gene-spacer units. Since these marker loci have virtually no selection value their evolutionary age is likely to be short and deep coalescence of the marker variants will therefore be limited. We believe that these factors, when combined with the known genomic location and relatively clear evolutionary dynamics, provide this typing system with advantages compared to anonymous multilocus markers such as AFLPs, RFLPs and microsatellites. Because of the high sequence conservation of the rRNA genes and the ubiquity of LTR retrotransposons in eukaryotic genomes, the method is likely to be applicable to many species and evolutionary questions.

Our analysis of retrotransposon polymorphisms in tetraploid wheats showed clear indications of a complex, non-linear relationship between wild and domesticated emmers. These indications included: the presence of four *Jeli*-5.8S allele groups in domesticated emmer, each group also present in wild accessions (Figure 2), a pattern that is inconsistent with a linear monophyletic relationship between the domesticated and wild populations; sharing of individual *Angela*- and *Jeli*-5S insertions, but not combinations of insertions, between wild and domesticated emmers (Figure 3), which is a clear sign of a reticulated relationship between the two populations; considerable incongruence of phylogenetic signals within the supernetwork constructed from the combined retrotransposon data (Figure 4); and the network summary of incongruent phylogenetic signals within the combined retrotransposon dataset, which show that hybridizations are not restricted to

solitary samples but instead involve all the domesticates and a large proportion of the wild emmers (Figure 4).

These results suggest that the origin of domesticated emmer is neither mono- nor polyphyletic, but reticulate, this being the principal reason why the phylogenetic trees previously built from different sets of markers and samples conclude different origins for this crop [20–25]. When considering the evolutionary history of domesticated emmer, it is therefore more meaningful to focus on the geographic distribution of individual phylogenetic signals (e.g. insertion markers and sequence types) rather than the distribution of detected genotypes. Within the domesticated accessions, we have identified phylogenetic signals originating from all the previously reported geographic regions – Gaziantep region, Zagros Mountains, southern Levant and southeastern Turkey. However, the majority of these markers/sequence types are dispersed across broad areas of the Fertile Crescent, and none of them is exclusive to any region east of the Euphrates.

A reticulate origin for domesticated emmer is supported by our reanalysis of published nuclear gene sequences. Filtered supernetworks constructed from these sequences displayed considerable incongruence in the region containing the domesticated accessions, most likely due to hybridization rather than homoplasy or deep coalescence (Figure 5). The networks show an affiliation between domesticated emmer and northern wild emmers, including some from southeastern Turkey and the Zagros region, as reported previously [20,22,23]. However, the topologies do not indicate that *dicoccum* and *durum* wheats descended from these northern genotypes. Instead they suggest that the domesticates and the northern emmers share a related reticulated ancestry, the northern wild emmers forming a genetic mosaic derived from the same ancestral populations that gave rise to the domesticated cluster. We therefore conclude that domesticated emmer and the northern wild populations have common ancestors west of the Euphrates.

Origins of Emmer Cultivation and Possible Impacts on Diversity and Reticulation Patterns

Previous phylogeographic studies searching for the place(s) of emmer domestication employ an implicit assumption that the distribution of wild emmer populations has not changed substantially since the beginning of the domestication process. Although the possibility that macro- and microclimatic variations might have altered the wild emmer distribution has long been recognized [51], the limited ways of investigating such past changes have meant that this problem has received only marginal attention. However, for the correct interpretation of emmer phylogeny, not only the post-domestication but also the pre-domestication distribution changes may be critical. Our data suggest that the distribution of wild emmer was originally confined to that of the 'southern race' [51] in the area of present-day Lebanon, northern Israel and southern Syria, centered around the upper Jordan valley. This is indicated by the topology of the MJ network of *Jeli*-5.8S sequences (Figure 2), in which most of the 'early-diverging' wild emmers, which did not contribute to the domesticated gene pool, as well as the wild forms appearing basal to all the cultivated BA^u wheats and their immediate ancestors, were collected from this area. A very similar picture, where the basal and early-diverging wild emmers originate from the southern Levant while Turkish and Iranian accessions appear only among the phylogenetically recent nodes, is also suggested by network analysis of 64 published wild emmer *Pm3* gene sequences [52] (Figure C in File S1). These observations suggest that BA^u wheat evolved in the southern Levant, as previously suggested [53], or irrespective of origin was restricted by the glaciations to a southern Levant refuge. In either case, the implication is that wild emmer spread to the northern and eastern Fertile Crescent relatively recently.

For a self-pollinating annual that grows in dense stands, long-distance dispersal does not improve survival [54]. Thus, wild wheat seeds lack features that facilitate wind or animal dispersal. However, wheat seed is easily transported by humans, raising the possibility that human communities contributed to the distribution of wild emmer. According to archaeological evidence, wild emmer has been collected by hunter-gatherers since the Upper Palaeolithic, 23,000 BP [24,29]. For millennia, nomadic communities migrated periodically across the Fertile Crescent, hunting gazelles and other ungulates, harvesting cereals from wild stands and, possibly, carrying grain supplies with them into new territories. Wheat grain is also likely to have featured in the Natufian and PPNA trade network, revealed by archaeological findings of obsidian, that extended in two directions – from Kapadokya (~240 km northwest from Gaziantep) to the southern reaches of the Jordan Valley, and from Bingöl (~150 km northeast from Karaca Dağ) to the Zagros Mountains [55,56]. It is therefore plausible that Epipaleolithic and early Neolithic communities contributed to the spread of wild emmer from its origin in the southern Levant. This scenario is consistent with the 5S- and 5.8S-retrotransposon diversity patterns, which show that the genetic variants possessed by the eastern wild samples represent only a small subset of those present in the Levant. The possibility that wild emmer populations from SE Turkey and Iraq/Iran entered the region only with the first cultivators has been proposed before [28,54], and is supported by the absence of wild emmer at the pre-domestication archaeological sites of northern Syria and southeastern Turkey until 10,500 BP, although einkorn, rye and barley are present prior to this period [24].

One possible explanation of the reticulations in the evolutionary history of domesticated emmer is that gene flow has occurred between the wild and domesticated populations, leading to introgression of alleles from dispersed parts of the wild population into different components of the crop. Gene flow between wild and domesticated plants has previously been suggested specifically for emmer [23] and more generally for cereals [57], but the scale of its possible impact has been questioned [58]. A second argument against extensive introgression of wild genes into the crop is the possibility that this will result in replacement of the recessive domestication traits, such as the nonshattering ear, with their wild versions, meaning that the introgressed plants are likely to be lost from the domesticated population.

Our results suggest that the reticulated origins of domesticated emmer are due to gene flow operating within a different scenario. We do not know what husbandry practices were used by the first human communities to cultivate wild cereals [58]. Regardless of whether stands of wild wheat were grown in 'fields', it seems likely that the cultivated material evolved into a blend of different populations, resulting from centuries of collection of wild grain from various sources. This mixture of populations would inevitably lead to genetic reticulation in the wild crop as a whole, the extent of this reticulation depending on the frequency of cross-pollination. The frequency of outcrossing in wild emmer has been estimated at 3%, based on the observed heterozygosity deficit [59], or below 1% based on a one-season field experiment [60]. Cross-pollination is enabled by open flowerage, which occurs in humid conditions, especially when coupled with cold, therefore the upper estimate may be more credible as a long term average. In a simple model where two large, genetically distinct emmer populations are intermixed in equal proportions, assuming absence of selection and 1% frequency of cross-pollination, the individuals gradually become a genetic mosaic of the initial two populations (Figure D in File S1). After 200 generations, the frequency of the original genotypes is below 10%, and the pace of these changes is even faster if three populations are intermixed, or more frequent cross-pollination is assumed. Hence, rare random hybridizations can result in relatively complex reticulation, despite the predominantly self-pollinating nature of wheat. We propose that the gene flow leading to the reticulations that we observe in the evolutionary record of domesticated emmer occurred predominantly during the pre-domestication phase, when mixed populations of wild emmer were being cultivated in the southern Levant.

A Dynamic Model for the Emergence of Agriculture

According to the scenario described above, the wild emmer of SE Turkey, Iran and Iraq, described as "never really abundant … in sporadic, isolated patches and thin scattered stands" and "hardly have been very attractive to the food-collecting cultures of the region" [51], is descended from the wild emmer grain from the Levant, taken to those more northern regions by humans during the Epipaleolithic and early Neolithic periods. This leads to a model for the origins of domesticated emmer that is consistent both with the previous genetic studies [20–25] and the archaeobotanical evidence for a protracted domestication process [13,57]. We hypothesize that during the Epipaleolithic, the hunter-gatherer communities of the southern Levant began to exercise control over their food supplies by managing wild stands of emmer, these activities progressing to the stage where wild grains were collected and cultivated to provide the next season's resource. The mobility of these communities engendered a dynamic situation in which distinct wild emmer populations from a large area west of the Euphrates became intermixed in the fields of these early farmers. Hybridization between different lineages within these fields, possibly over an extended period of time, gave rise to pre-domesticated crops that displayed a reticulated evolutionary relationship with the wild populations from which they were originally derived. Accompanying these events, cultivation of wild emmer spread from the southern Levant in a clock-wise direction to other parts of the Fertile Crescent. Movement away from the range of the wild population established reproductive isolation between wild and cultivated plants, facilitating human selection for the domestication traits [5]. The first domesticated plants therefore appeared when cultivation reached northern Syria, southeast Turkey and northern Iraq. The wild emmers found today in the Karaca Dağ and Sulaymaniyah regions, being the remnants of the cultivated population from which the first domesticates evolved, are therefore identified as genetically proximal to the domesticated genepool when phylogenetic methods that enforce a treelike pattern of evolution are used. The genetically domesticated varieties then spread to other farming regions, such that the proportion of domesticated to wild grain at these sites gradually increased. Eventually the domesticated population spread outside the Fertile Crescent, resulting in independent bottlenecks which gave rise to the geographical variability of the tetraploid landraces observed today.

The protracted model for the origins of agriculture has repeatedly been interpreted as requiring no conscious human involvement in the domestication process [6,58]. Our dynamic model assumes that the transition from hunting-gathering to agriculture was protracted but equally assumes that humans played an active and intuitive role throughout the process. We do not, however, believe that in order to assign humans an active role it is necessary to interpret the transition as a teleological process culminating in conscious human selection of fully domesticated plants. Emergence of the domestication traits might appear, from a retrospective viewpoint, to be the endpoint of a progressive evolutionary process, but the data we present show that, in reality, the genetic changes that underlay 'domestication' were merely part of a lengthy series of events that began in the cultivated wild populations, and continued in the domesticated population for centuries after the 'origin' of agriculture.

Supporting Information

File S1. (DOCX)

Acknowledgements

We thank Heather Robinson and Wanjun Qi for technical contributions. We also thank Martin Jones (University of Cambridge) for discussions that contributed to development of the model described in this paper.

Author Contributions

Conceived and designed the experiments: PC TAB. Performed the experiments: PC ZI. Analyzed the data: PC. Wrote the manuscript: PC TAB.

References

1. Bowles S (2012) Cultivation of cereals by the first farmers was not more productive than foraging. Proc Natl Acad Sci of the USA US. 22: 4760–4765.
2. Bettinger R, Richerson P, Boyd R (2009) Constraints on the development of agriculture. Curr Anthropol 50: 627–631. doi: 10.1086/605359. PubMed: 20642149.
3. Bar-Yosef O (2011) Climatic fluctuations and early farming in West and East Asia. Curr Anthropol 52: S175–S193. doi:10.1086/659784.
4. Rindos D (1984) The origins of agriculture: an evolutionary perspective. New York: Academic Press.
5. Jones MK, Brown TA (2007) Selection, cultivation and reproductive isolation: a reconsideration of the morphological and molecular signals of domestication. In: TP DenhamJ IriarteL Vrydaghs. Rethinking agriculture: archaeological and ethnoarchaeological perspectives. Walnut Creek: Left Coast Press. pp. 36–49.
6. Abbo S, Lev-Yadun S, Gopher A (2011) Origin of Near Eastern plant domestication: homage to Claude Levi-Strauss and ''La Pensée Sauvage''. Genet Resour Crop Evol 58: 175–179. doi:10.1007/s10722-010-9630-0.
7. Cauvin J (2000) The beginnings of agriculture in the Near East: a symbolic interpretation. Cambridge: Cambridge University Press.
8. Brown TA, Jones MK, Powell W, Allaby RG (2009) The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol Evol 24: 103–109. doi:10.1016/j.tree.2008.09.008. PubMed: 19100651.
9. Zohary D (1996) The mode of domestication of the founder crops of Southwest Asian agriculture. In: DR Harris. The origins and spread of agriculture and pastoralism in Eurasia. London: UCL Press. pp. 142–158.
10. Hillman GC, Davies MS (1990) Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. J World Prehist 4: 157–222. doi:10.1007/BF00974763.
11. Diamond J (1997) Location, location, location: the first farmers. Science 278: 1243–1244. doi:10.1126/science.278.5341.1243.
12. Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M et al. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278: 1312–1314. doi:10.1126/science.278.5341.1312.
13. Charmet G (2011) Wheat domestication: lesson for the future. C R Biol 334: 212–220. doi:10.1016/j.crvi.2010.12.013. PubMed: 21377616.
14. Haldersen S, Akan H, Çelik B, Heun M (2011) The climate of the Younger Dryas as a boundary for einkorn domestication. Veget Hist Archaeobot 20: 305–318.
15. Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3: 429–441. PubMed: 12042770.
16. Allaby RG, Brown TA (2003) AFLP data and the origins of domesticated crops. Genome 46: 448–453. doi:10.1139/g03-025. PubMed: 12834061.
17. Allaby RG, Fuller DQ, Brown TA (2008) The genetic expectations of a protracted model for the origins of domesticated crops. Proc Natl Acad Sci U S A 105: 13982–13986. doi:10.1073/pnas.0803780105. PubMed: 18768818.
18. Fuller DQ, Willcox G, Allaby RG (2011) Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. World Archaeol 43: 628-652. doi:10.1080/00438243.2011.624747.
19. Fuller DQ, Asouti E, Purugganan MD (2012) Cultivation as slow evolutionary entanglement: comparative data on rate and sequence of domestication. Veget Hist Archaeobot 21: 131–145. doi:10.1007/s00334-011-0329-8.
20. Ozkan H, Brandolini A, Schäfer-Pregl R, Salamini F (2002) AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 19: 1797–1801. doi:10.1093/oxfordjournals.molbev.a004002. PubMed: 12270906.
21. Mori N, Ishi T, Ishido T, Hirosawa S, Watatani H, et al. (2003) Origins of domesticated emmer and common wheat inferred from chloroplast DNA fingerprinting. In: NE PognaM RomanoEA PognaG Galterio. Proceedings of the 10th international wheat genet symposium, Paestum, Italy. Rome: Instituto Sperimentale per la Cerealicoltura. pp. 25–28.
22. Ozkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F (2005) A reconsideration of the domestication geography of tetraploid wheats. Theor Appl Genet 110: 1052–1060. doi:10.1007/s00122-005-1925-8. PubMed: 15714326.
23. Luo M-C, Yang ZL, You FM, Kawahara T, Waines JG, Dvorak J (2007) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet 114: 947–959. doi:10.1007/s00122-006-0474-0. PubMed: 17318496.
24. Özkan H, Willcox G, Graner A, Salamini F, Kilian B (2011) Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). Genet Resour Crop Evol 58: 11–53. doi:10.1007/s10722-010-9581-5.
25. Brown TA, Lindsay S, Allaby RG (2006) Using modern landraces of wheat to study the origins of European agriculture. In: T MotleyN ZegeraH Cross. Darwin's harvest: new approaches to the origins, evolution, and conservation of crops. New York: Columbia University Press. pp. 197–212.
26. Fuller DQ (2007) Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old. World - Ann Bot 100: 903–909. doi:10.1093/aob/mcm048.
27. Tanno K, Willcox G (2006) How fast was wild wheat domesticated? Science 311: 1886. doi:10.1126/science.1124635. PubMed: 16574859.
28. Feldman M, Kislev ME (2007) Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. Isr J Plant Sci 55: 207–221. doi:10.1560/IJPS.55.3-4.207.
29. Weiss E, Kislev ME, Hartmann A (2006) Autonomous cultivation before domestication. Science 312: 1608–1610. doi:10.1126/science.1127235. PubMed: 16778044.
30. Tanno K-I, Willcox G (2011) Distinguishing wild and domestic wheat and barley spikelets from early Holocene sites in the Near East. Veget Hist Archaeobot 27: 107–115.
31. Baum DA (2007) Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56: 417–426.
32. Huson DH, Dezulian T, Klöpper T, Steel MA (2004) Phylogenetic super-networks from partial trees. IEEE-ACM Trans Comput Biol Bioinform 1: 151–158
33. Huson DH, Steel MA, Whitfield J (2006) Reducing distortion in phylogenetic networks. In: P BücherBME Moret. WABI 2006. Lecture notes in computer science 4175. New York: Springer. pp. 150–161.
34. Mukai Y, Endo TR, Gill BS (1990) Physical mapping of the 5S rRNA multigene family in common wheat. J Hered 81: 290–295.
35. Röder MS, Sorrells ME, Tanksley SD (1992) 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232: 215–220. PubMed: 1557028.
36. Mukai Y, Endo TR, Gill BS (1991) Physical mapping of the 18S.26S rRNA multigene family in common wheat: Identification of a new locus. Chromosoma 100: 71–78. doi:10.1007/BF00418239.
37. Allaby RG, Brown TA (2001) Network analysis provides insights into the evolution of 5S rDNA arrays in *Triticum* and *Aegilops*. Genetics 157: 1331–1341. PubMed: 11238418.
38. Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D et al. (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24: 1506–1517. doi:10.1093/molbev/msm077. PubMed: 17443011.
39. Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H et al. (2008) *Cassandra* retrotransposons carry independently transcribed 5S. RNA - Proc Natl Acad Sci U_S_A 105: 5833- 5838. doi:10.1073/pnas.0709698105.
40. Wicker T, Matthews DE, Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends Plant Sci 7: 561–562. doi:10.1016/S1360-1385(02)02372-5.
41. Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. Genes Genomes Genomics 3: 1–14.
42. Charles M, Belcram H, Just J, Huneau C, Viollet A et al. (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics 180: 1071–1086. doi:10.1534/genetics.108.092304. PubMed: 18780739.
43. Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M et al. (2010) Geneious v5.0. Available: http://www.geneious.com.
44. Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37–48. doi: 10.1093/oxfordjournals.molbev.a026036. PubMed: 10331250.
45. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267. PubMed: 16221896.
46. Maddison WP (1997) Gene trees in species trees. Syst Biol 46: 523–536. doi:10.1093/sysbio/46.3.523.
47. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Seattle (WA): Department of Genome Sciences, University of Washington. Available: http://evolution.genetics.washington.edu/phylip.html.

48. Huson DH, Scornavacca C (in press) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol.
49. Huang S, Sirikhachornkit A, Su X, Faris J, Gill B et al. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci U S A 99: 8133–8138. doi:10.1073/pnas.072223799. PubMed: 12060759.
50. Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B et al. (2008) *Acc* homoeoloci and the evolution of wheat genomes. Proc Natl Acad Sci U S A 105: 9691–9696. doi:10.1073/pnas.0803981105. PubMed: 18599450.
51. Harlan JR, Zohary D (1966) Distribution of wild wheats and barley. Science 153: 1074–1080. doi:10.1126/science.153.3740.1074. PubMed: 17737582.
52. Yahiaoui N, Kaur N, Keller B (2009) Independent evolution of functional *Pm3* resistance genes in wild tetraploid wheat and domesticated bread wheat. Plant J 57: 846–856. doi:10.1111/j.1365-313X.2008.03731.x. PubMed: 18980638.
53. Ozbek O, Millet E, Anikster Y, Arslan O, Feldman M (2007) Comparison of the genetic structure of populations of wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides*, from Israel and Turkey revealed by AFLP analysis. Genet Resour Crop Evol 54: 1587–1598. doi:10.1007/s10722-006-9171-8.
54. Elbaum R, Zaltzman L, Burgert I, Fratzl P (2007) The role of wheat awns in the seed dispersal unit. Science 316: 884–886. doi:10.1126/science.1140097. PubMed: 17495170.
55. Mithen SJ (2003) After the ice: a global human history, 20,000–5000 BC. London: Weidenfeld & Nicolson Ltd.
56. Sherratt A (2005) Obsidian trade in the Near East, 14,000 to 6,500 BC. ArchAtlas, February 2010, Edition 4.. Available: http://www.archatlas.org/ObsidianRoutes/ObsidianRoutes.php.
57. Purugganan MD, Fuller DQ (2011) Archaeological data reveal slow rates of evolution during plant domestication. Evolution 65: 171–183. doi:10.1111/j.1558-5646.2010.01093.x. PubMed: 20666839.
58. Heun M, Abbo S, Lev-Yadun S, Gopher A (2012) A critical review of the protracted domestication model for Near-Eastern founder crops: linear regression, long-distance gene flow, archaeological, and archaeobotanical evidence. J Exp Bot 63: 4333–4341. doi:10.1093/jxb/ers162. PubMed: 22717409.
59. Nevo E, Golenberg E, Beiles A, Brown AHD, Zohary D (1982) Genetic diversity and environmental associations of wild wheat, *Triticum dicoccoides*, in Israel. Theor Appl Genet 62: 241–254.
60. Golenberg EM (1988) Outcrossing rates and their relationship to phenology in *Triticum dicoccoides*. Theor Appl Genet 75: 937–944.

APPENDIX V

Taxonomical classification and origin of Kamut®

Michalcová V., Dušinský R., Sabo M., Al Beyroutiová M., Hauptvogel P., Ivaničová Z. and Švec M.

Plant Syst Evol, 300(7), 1749-1757, 2014

Plant Syst Evol
DOI 10.1007/s00606-014-1001-4

ORIGINAL ARTICLE

Taxonomical classification and origin of Kamut® wheat

Veronika Michalcová · Roman Dušinský · Miroslav Sabo · Maja Al Beyroutiová · Pavol Hauptvogel · Zuzana Ivaničová · Miroslav Švec

Received: 16 September 2013 / Accepted: 25 January 2014

Abstract Bioagriculture and healthy lifestyle are trends of the twenty-first century. Bioagriculture involves the breeding of crops without using modern synthetic substances. Kamut brand wheat is one of the popular biocereals grown mainly in the USA and Europe. This cereal has the status of ancient wheat, not only because it has been grown since the era of the ancient Egyptian civilization, but also for its properties favorable for modern breeding programs and modern food marketing. In spite of Kamut's® interesting history and stable place in the market, it is not a common subject of genetic studies. It is also interesting that it has not been successfully taxonomically classified yet. There are a few studies which classify this tetraploid wheat as *Triticum polonicum* L., *T. turanicum* Jakubz., *T. turgidum* L. and *T. durum* Desf. These studies are based on cytological and comparative methods. We chose molecular (transposable element resistance gene analog polymorphism, diversity arrays technology, sequencing of genes *SBEIIa,* and *ψLpx-A1_like*) and statistical methods to classify Kamut® wheat. According to our experiments we suggest that Kamut brand wheat originated as a natural hybrid between *Triticum dicoccon* conv. *dicoccon* and *T. polonicum* and is not original ancient Egyptian wheat. We suggest that Etruscan wheat has the same parents as Kamut®.

Keywords Kamut® · Molecular taxonomy · Origin · TERGAP · DArT

V. Michalcová (✉) · R. Dušinský · M. Al Beyroutiová · M. Švec
Department of Genetics, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynska dolina, 842 15 Bratislava, Slovakia
e-mail: michalcova.veronika3@gmail.com

M. Sabo
Department of Mathematics and Constructive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinskeho 11, 813 68 Bratislava, Slovakia

P. Hauptvogel
Research Institute of Plant Production, Bratislavska cesta 122, 921 68 Piešťany, Slovakia

Z. Ivaničová
Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany AS CR, v. v. i., Šlechtitelů 31, 783 71 Olomouc, Holice, Czech Republic

Introduction

The Kamut brand wheat (Kamut®) is a tetraploid wheat commonly known as King Tut's or Khorasan wheat (also nomenclature for *T. turanicum*). The nomenclature "kamut" means "wheat" in ancient Egyptian language. Egyptologists claim the root meaning of this word is "soul of the Earth". Since 1990, a registered trademark (Kamut®) has been used in marketing products of the protected cultivated *T. turanicum* variety "QK-77" (Quinn 1999).

The origin of this wheat is intriguing. During World War II an American airman claimed to have taken a few grains of some cereal from an ancient Egyptian tomb near Dashare. The story tells that he gave these grains to a farmer, who grew plants from them and resurrected a long-forgotten cereal. This story is just a modern legend and as a fact it has to be rejected, for most scientists believe that it probably survived the years as an obscure grain kept alive by the diversity of crops common to peasant farmers, perhaps in Egypt or Asia Minor. It is thought to have evolved contemporarily with the free-threshing tetraploid wheat.

Published online: 14 February 2014

Besides its unusual history, this crop is interesting for its properties that are due to isolation from modern breeding. Kamut® has not been in contact with synthetic substances commonly used in modern breeding programs (Hammer et al. 2000). Kamut brand wheat is of ≈127 cm height and has two to three times larger grains than other wheat cultivars. The grains are narrow, vitreous and flinty with a characteristic hump (Vavilov 1951). The grain contains 20–30 % more proteins, higher levels of eight out of nine minerals, more lipids and up to 65 % more amino acids than other wheat cultivars. Alleles for prolamin, related to good pasta quality, were identified (Rodríguez-Quijano et al. 2010). Since lipids present more energy than carbohydrates, Kamut® is characterized as high-energy wheat. Kamut® products are marketed mainly through health food outlets. Due to its sweet taste, it plays a special role in bakeries as there is no need to add any sugar to pastries produced from Kamut® flour (Quinn 1999).

In the past it was believed that Kamut® wheat did not induce as strong allergy as other wheat in patients suffering from gluten intolerance (Quinn 1999). This information was refuted by the discovery that Kamut brand wheat causes the same allergic reactions as *T. durum* (Simonato et al. 2002). That means that, in spite of many marketing affirmations, products made of Kamut® wheat are not suitable for celiac disease patients. The most recent study on Kamut brand wheat showed that this grain protects organisms from oxidative stress better than *T. durum* (Benedetti et al. 2012).

The growing of this special wheat is exclusively managed by license agreements and requires organic certification of the crop. It is grown mainly in the USA and in a limited area of Austria (Grausgruber et al. 2004).

There are still disagreements about the nearest relatives of Tut's wheat. First, it was classified as *T. polonicum* and it was also believed that Kamut® was a natural hybrid between *T. durum* and *T. polonicum* (Kuckuck 1959). Some groups classify this wheat as *T. turanicum* (Percival 1921) or *T. turanicum* var. *notabile* (Perciv.) Gökg. (Gowayed 2009) and others as *T. turgidum* or *T. durum* (Brouwer 1972).

Materials and methods

Samples of tetraploid wheat

The analyzed samples belonged to the group of species with GGAA genomes: *T. timopheevii* (Zhuk.) Zhuk. (TIM), *T. araraticum* Jakubz. (ARM); and species with BBAA genomes: *T. dicoccoides* (Körn. ex Asch. et Graebn.) Schweinf. (DCS), *T. dicoccon* Schrank (DIM, for European samples *T. dicoccon* convar. *dicoccon* EUR abbreviation was used), *T. turgidum* L. (TRG), *T. carthlicum* Nevski (CAR), *T. durum* Desf. (DUR), *T. turanicum* Jakubz. (TRN), *T. polonicum* L. (PLN), *T. ispahanicum* Heslot (ISP), one sample of Kamut® wheat, and one sample of Etruscan wheat (BVAL 212017). We used one sample of *Aegilops speltoides* Tausch (SPE, SS genome) and one sample of *T. urartu* Thum. ex Gandil. (URA, AA genome) as outgroups. Wheat nomenclature in this article is according to Dorofeev et al. (1979) for naked wheat and Szabó and Hammer (1996) and Hammer et al. (2011) for hulled wheat.

Isolation of DNA samples

In our experiments, we analyzed 57 tetraploid wheat samples. DNA samples were isolated using CTAB extraction buffer [200 mmol l^{-1} Tris–HCl (pH 8.0), 1.4 mol l^{-1} NaCl, 20 mmol l^{-1} EDTA, 2 % (w/v) CTAB, 0.2 % (v/v) 2-mercaptoethanol] and High Pure PCR Product Purification Kit (Roche) from wheat grains.

Transposable element resistance gene analog polymorphism

Isolated DNA was analyzed using the TERGAP method. To avoid non-specific products, PCR was performed using hot start polymerase TrueStart *Taq* (Fermentas). The reaction mixture was composed of: 1× PCR buffer mixture 1:1 (GoTaq, Promega; TrueStart, Fermentas); 1 U *Taq* polymerase (TrueStart, Fermentas); 2.0 mM $MgCl_2$; 0.25 mM dNTP; 0.4 μM primers and 30 ng of template DNA in total volume 12.5 μl. We designed 17 combinations of primers. Primer sequences and combinations are summarized in Tables 1 and 2. We optimized PCR cycle: starting denaturation 95 °C (2 min); 34 cycles with temperatures 94 °C (45 s), 54 °C (45 s), 72 °C (2 min); final polymerization 72 °C (7 min). Amplification results were evaluated by 1.5 % agarose electrophoresis gel; fragments were separated over 2.5 h in constant voltage 100 V (3.5 V/cm of electrode distance) and stained with ethidium bromide.

Diversity arrays technology polymorphism (DArT)

Isolated DNA samples were sent to Triticarte™ whole-genome genotyping service for wheat DArT analysis, which is based on hybridization of DNA with special probes. These probes are designed strictly for wheat and carry up to 2,500 polymorphic markers. Binary data are output of this analysis.

Sequencing of genes *SBEIIa* and *ψLpx-A1_like*

Some of isolated DNA samples were used for sequencing of genes *SBEIIa* and *ψLpx-A1_like*. Each PCR reaction

Table 1 Sequences of primers used in TERGAP reactions

Primer sequences	
Primer name	Sequence (5′–3′)
M1fwd1	CAG GCT CCA GAA ATG CTT CA (Civáň unpublished)
M1rev1	GCT TGC AAA TGA AGT GCA GA (Civáň unpublished)
Jeli	CCT AGG AAC ATA GCT TCA TC (Civáň et al. in press)
Xa1LRF	CTC ACT CTC CTG AGA AAA TTA C (Civáň unpublished)
Ptokin1	GCA TTG GAA CAA GGT GAA (Civáň unpublished)
NLRRfor	TAG GGC CTC TTG CAT CGT (Civáň unpublished)
CLRRfor	TTT TCG TGT TCA ACG ACG (Civáň unpublished)
NLRR-INV1	TGC TAC GTT CTC CGG G (Civáň unpublished)
RLRRfor	CGC AAC CAC TAG AGT AAC (Civáň unpublished)
As1-INV	CCT AAC GGT GAT CGC AAC (Civáň unpublished)
Xa1NBS-R	CTC TGT ATA CGA GTT GTC (Civáň unpublished)

Table 2 Review of primer combinations used in TERGAP reactions

Primer combinations		
Claudia retrotransposon + RGA primer	Jeli retrotransposon + RGA primer	Single RGA primer reaction
M1fwd1 + Xa1LRf	Jeli + Xa1LRf	–
M1rev1 + Ptokin1	Jeli + Ptokin1	–
M1fwd1 + NLRRfor	Jeli + NLRRfor	–
M1fwd1 + CLRRfor	Jeli + CLRRfor	–
M1fwd1 + NLRR-INV1	M1fwd1 + NLRR-INV1	NLRR-INV1
M1fwd1 + RLRRfor	Jeli + RLRRfor	–
M1fwd1 + As1-INV	Jeli + As1-INV	–
M1fwd1 + Xa1NBS-R	Jeli + Xa1NBS-R	–

mixture was composed of 1× PCR buffer (HotFirePol), 1 U *Taq* polymerase (HotFirePol, Solis BioDyne), 1.6 mM $MgCl_2$, 0.2 mM dNTP, 0.2 µM primers and 40 ng of template DNA in a total volume 50 µl. Primer sequences are summarized in Table 3. We optimized the PCR cycle as: starting denaturation 95 °C (12 min); 35 cycles with temperatures 95 °C (45 s), 59 °C (45 s), 72 °C (45 s); final polymerization 72 °C (7 min). The PCR products were purified using High Pure PCR Product Purification Kit (Roche). The products were sequenced using ABI 3730xl sequencing machine (Applied Biosystems). The sequences were aligned by software Mega5 (Tamura et al. 2011).

Table 3 Sequences of primers used for sequencing of genes *SBEIIa* and *ψLpx-A1_like*

Primer sequences	
Primer name	Sequence (5′–3′)
SBEII_fwd	CCTGTTTCTGGTCTGATGGTC
SBEII_rev	ATGGGAGATCCCTACAATGC
ψLpx-A1_like_fwd	CCAACGACGTGAGTGATCCTTTTGC
ψLpx-A1_like_rev	AGCGCGAACCGTCATCTCGAA

Construction of phylogenetic trees

DNA banding patterns were visualized using Vilber Lourmat Gel Documentation system, and detected bands were sorted according to their size in bp with Bio-1D software (Vilber Lourmat, France). A discrete-state data matrix in binary format was constructed manually from the Microsoft Excel output of Bio-1D software, assigning value 1 to band presence and value 0 to band absence for each observed band size value. The phylogenetic analysis was performed using the programs included in the PHYLIP package, version 3.69 (Felsenstein 2005). A total of 500 bootstrap replicates of binary data matrix generated by the SEQBOOT program were analyzed by the PARS program with Wagner parsimony method using randomized input order of taxons with 100 times to jumble. An extended majority rule consensus tree was constructed using the CONSENSE program. The resulting phylogenetic tree was visualized with the Mega5 program with ClustalW algorithm (Tamura et al. 2011).

Statistical evaluation of data

For estimating population average pairwise differences, we used Arlequin software (version 3.5) in which we chose distance method pairwise difference. We computed Nei's average number of differences between populations (Nei and Li 1979).

For visualization purposes, we used multidimensional scaling technique (MDS) that is one of the dimension reduction methods (techniques that can reduce any multi-dimensional space to lower dimensional space). The main advantage is that it tries to preserve the original distances between pairs of objects (but similar to other dimension reduction techniques, there is also some distortion in the final projection). It takes dissimilarity matrix as the input, so any kind of data where one can compute dissimilarities between objects can be used. In our case we first used two dimensions to make the final 2D MDS projection (Fig. 1).

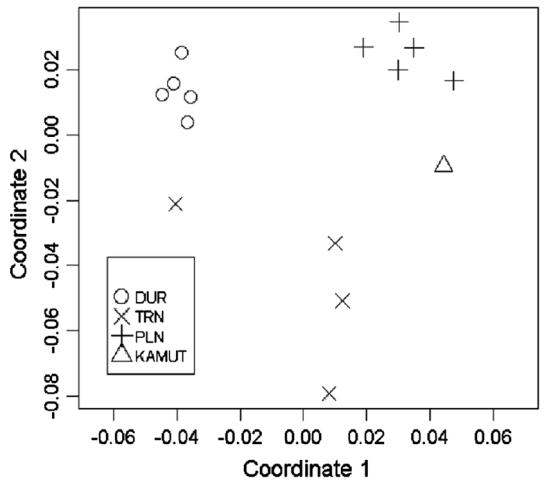


Fig. 1 MDS distance comparison of Kamut brand wheat and three genetically close species

The hybrid status of samples was evaluated by means of STEM-hy program (Kubatko et al. 2009). This is a program to perform maximum-likelihood analysis for estimation of the species tree from multilocus data under the coalescent process. Coalescent trees for each of the sequenced genes were constructed by the Mesquite program (Maddison and Maddison 2011). The most probable gene tree from 10,000 trees was chosen by the COAL program (Degnan and Salter 2005).

Parental populations of Kamut® variety were estimated from the binary data obtained from DNA polymorphism of DArT type by means of the HIest program (Fitzpatrick 2012). This program uses likelihood to estimate ancestry and heterozygosity. The *threeway* module of this program enables the maximum-likelihood estimates of ancestry for a sample of hybrid in a three-way hybrid zone, i.e., from three potential parental lineages. The output parameters s1, s2 and s3 of this program mean the proportion of alleles derived from parental lineages P1, P2 and P3.

Results

TERGAP analysis

TERGAP reactions were analyzed by agarose gel electrophoresis. During gel analyzing, several specific polymorphisms were identified. These genome-specific and species-specific polymorphisms were crucial for phylogenetic tree construction. Nine B genome-specific and three G genome-specific polymorphisms were identified. These polymorphic markers included three length polymorphisms and seven band presence/absence-based polymorphisms. Species-specific polymorphisms were identified for *T. polonicum*, *T. turgidum*, *T. dicoccoides* and *T. timopheevii* species. We also found polymorphism specific for *T. polonicum* in Kamut's® genotype. Two polymorphisms were specific for multiple taxons: one was common for *T. dicoccon* and *T. ispahanicum*, and the other one was common for *T. durum, T. turanicum* and *T. polonicum* species.

Phylogenetic tree

In the phylogenetic tree constructed by the PHYLIP programs, *GGAA* and *BBAA* lines were divided into two major clusters. Hulled and naked wheat was also distinguished (Fig. 2). Most samples from the same species were grouped into separate clusters. GGAA group was clearly divided into *T. timopheevii* and *T. araraticum* species. In contrast, the BBAA group was not differentiated as clearly as the GGAA group. There were no individual clusters for *T. dicoccon*, *T. dicoccoides* and *T. ispahanicum* (which formed a common cluster) and for *T. turanicum* and *T. polonicum* samples, which also clustered together. The sample of Kamut® wheat was integrated into one cluster with *T. turanicum* and *T. polonicum* samples.

Population average pairwise differences

Statistical analysis of species sample groups revealed the most significant resemblance of Kamut® wheat and *T. polonicum* samples, which was demonstrated by a value of 8.5. This number shows the lowest value of Kamut's® distance from each analyzed species (Table 4). In Table 4 it is necessary to pay attention to the diagonal line. Values above the diagonal reflect the average number of pairwise differences between populations (sample groups in this case), diagonal elements mirror the average number of pairwise differences inside populations and values below the diagonal show the corrected average pairwise differences. The highest diversity presented by the highest value of average number of pairwise diversity was detected inside the *T. turanicum* sample group. In contrast, the *T. araraticum* sample group seemed the most compact. The highest diversity between groups of samples was detected between *T. araraticum* and *T. turanicum* groups. While evaluating Table 4, it is not possible to estimate the columns "SPE, BVAL, KAMUT, CYL, URA" as groups/populations, as we used only one sample for each genotype.

Multidimensional scaling

In this method, we compared Kamut® wheat with its three genetically close species (*T. polonicum*, *T. turanicum* and *T. durum*). The status of Kamut® on the multidimensional scaling graph (Fig. 1) proves our theory that Kamut® is the most related to *T. polonicum*, which also proves the results obtained by the pairwise difference method (Table 4).

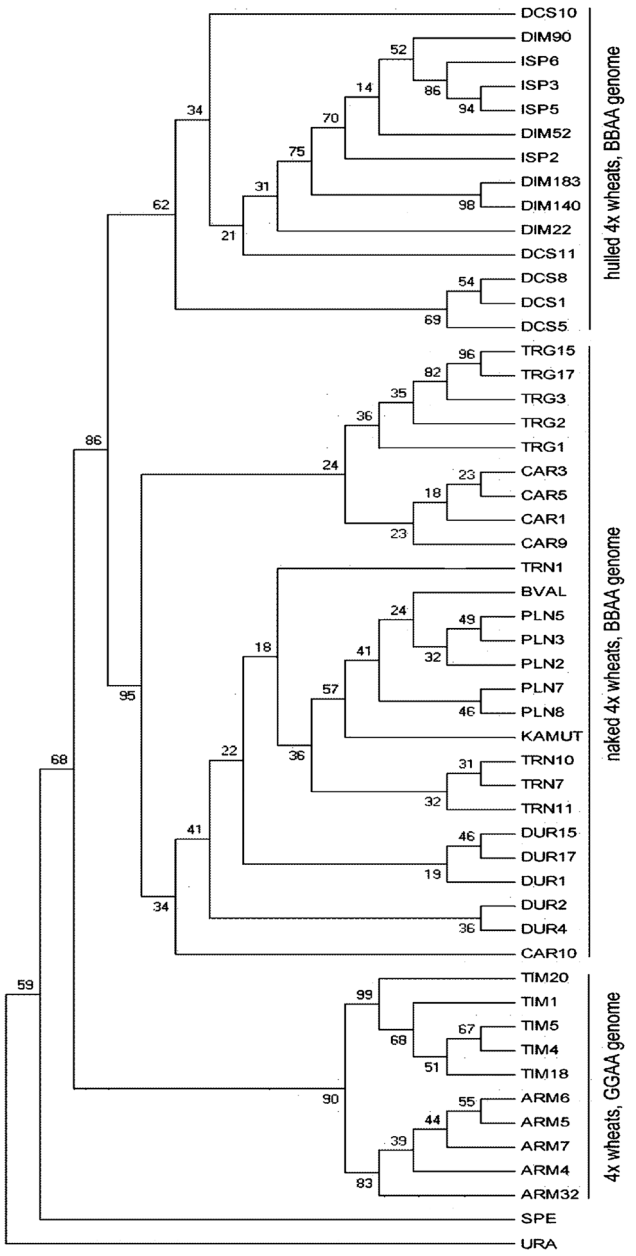


Fig. 2 Phylogenetic tree constructed by PHYLIP 3.69; sample URA was used as an outgroup. Samples are divided into two main clusters, one with BBAA genomes and the other with GGAA genomes. We could also distinguish hulled and naked wheat. Kamut® wheat was integrated right on the edge between *Triticum turanicum* and *T. polonicum* species

Haplotypes *ψLpx-A1_like*

Sequencing of *ψLpx-A1_like* gene divided our samples into four haplotypes (Fig. 3). According to these results, the Kamut® sample carries the same haplotype as some European samples of *T. dicoccon* convar. *dicoccon*, which indicates that Kamut® is most probably a hybrid wheat, because it reflects similarity both with naked (*T. polonicum*) and hulled taxons (*T. dicoccon*).

STEM-hy analyses

The results of sequencing of *ψLpx-A1_like* and *SBEIIa* have pushed us forward into proving that our hypothesis about the hybrid origin of Kamut® is right. According to the four haplotypes obtained by sequencing of *ψLpx-A1_like*, we chose five samples for STEM-hy analyses (Kubatko et al. 2009), which was designed to estimate hybrid samples (Table 5). These five samples (DCS 11, Kamut, PLN 1, EUR 42, EUR 132) are typical representations of four haplotypes. The results for sample DCS 11 are not given in Table 5 because this sample is ancestral and thus we do not expect its hybrid origin. In this table, Kamut® sample has the highest values of likelihood, which suggests that this sample is a hybrid. Another output of STEM-hy software is a hybrid tree (Fig. 4) that shows which one of the tested samples has the largest amount of genetic information common with Kamut®. This is presented by branch localization; the branch nearest to the tested sample is genetically most similar to this sample and it may also indicate the sample's parental ancestors in some cases. In hybrid tree, Kamut® is localized on the same branch as the reticulated sample EUR 42, indicating that these two samples are probably the progeny of the same crossing. This result signifies that one of Kamut's® parents is the European genotype of *T. dicoccon* convar. dicoccon, which is very similar to the EUR 42 sample.

HIest analysis

Phylogenetic tree and MDS diagram constructed by using TERGAP binary data have indicated *T. polonicum* to be one of Kamut's® parental genotypes. HIest software was used to confirm this hypothesis using three potential parental populations: P1 (DUR 16, DUR 21, DUR 24, DUR 30, DUR 38), P2 (PLN 1, PLN 2, PLN 3, PLN 5, PLN 11, PLN 14) and P3 (EUR 132, EUR 387, EUR 189). This software operated with DArT binary data and calculated ancestry indices s1, s2 and s3 from 125 diagnostic markers which are specific for each parental population. Each ancestry index (s) shows the proportion of alleles derived from a particular parental lineage (P). The highest values of ancestry index were detected in population P2 (Table 6), which supports our hypothesis that Kamut® also has a naked ancestor within *T. polonicum*. As shown in Table 6, this model is more probable when a hybrid sample of Kamut® was included in the calculation of parental frequencies of *T. polonicum* for each of the 125 diagnostic markers. This is proved by a higher value of likelihood (−54.861).

Table 4 Average numbers of pairwise differences by distance method pairwise difference

	ARM	TIM	DCS	DIM	ISP	TRG	CAR	DUR	TRN	PLN	SPE	BVAL	KAMUT	URA
ARM	*5.8*	24	43.9	51.1	59.3	50.6	53.9	51.6	58.1	57.5	63.8	59.4	57.8	84.8
TIM	17.9	*6.4*	55.2	60.5	66.6	60.9	62.6	60.4	66.5	63.2	64	65.2	65.2	89.4
DCS	33.3	44.3	*15.4*	23.9	30.3	42	41	36.9	42.7	41.8	72.6	47.4	38.2	86.8
DIM	40.3	49.4	8.3	*15.8*	19.9	43.4	39	40.2	42.1	46.6	64.4	52.4	39.6	89.8
ISP	50.8	57.8	16.9	6.4	*11.2*	51.2	46.2	45.4	48.3	45.1	72.3	52.8	44.3	96.3
TRG	42.4	52.4	29	30.2	40.3	*10.6*	15.6	20.4	26.4	28.3	75.8	27.8	29.8	81.2
CAR	46.2	54.6	28.5	26.3	35.8	5.5	*9.6*	17.2	23.4	26.7	69.4	28.2	27	80.8
DUR	45	53.6	25.6	28.7	36.2	11.5	8.8	*7.2*	19.2	17.3	77.6	18	19.6	83.4
TRN	47	55.1	26.8	26	34.5	12.9	10.4	7.4	*16.3*	20.3	72.5	23.5	18	88.5
PLN	50.5	55.9	29.9	34.6	35.4	18.9	17.8	9.6	8	*8.2*	82.6	12.2	12.6	94.4
SPE	60.9	60.8	64.9	56.5	66.7	70.5	64.6	74	64.3	78.5	*0*	86	76	91
BVAL	56.5	62	39.7	44.5	47.2	22.5	23.4	14.4	15.3	8.1	86	*0*	18	91
KAMUT	54.9	62	30.5	31.7	38.7	24.5	22.2	16	9.8	*8.5*	76	18	*0*	93
URA	81.9	86.2	79.1	81.9	90.7	75.9	76	79.8	80.3	90.3	91	91	93	*0*

Values above diagonal average number of pairwise differences between populations. *Diagonal elements* average number of pairwise differences inside populations. *Values below diagonal* corrected average pairwise differences. The highest values show the highest distances inside or between populations

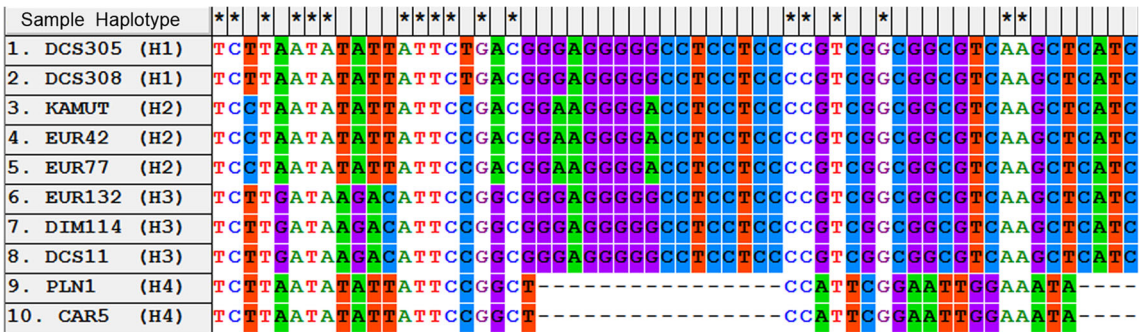


Fig. 3 Variable position of *ψLpx-A1_like* gene for selected tetraploid wheat samples, which represent four haplotypes (*H1*, *H2*, *H3*, *H4*). The majority of conservative positions were deleted

Table 5 Values of likelihood and AIC of hybrid trees calculated by STEM-hy for postulated hybrids

Postulated hybrid	Likelihood	AIC
KAMUT	−5,791.83	11,593.663
PLN 1	−5,900.83	11,811.662
EUR 42	−5,791.83	11,593.663
EUR 132	−5,853.13	11,716.276

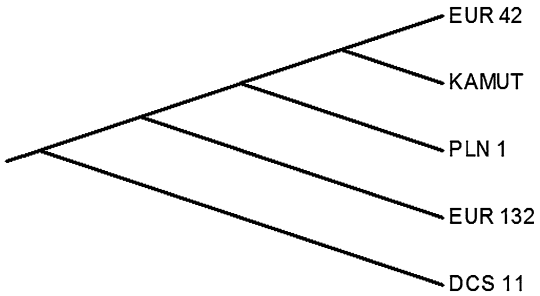


Fig. 4 Cladogram produced by STEM-hy software

Similarity analysis of DArT polymorphism

Similarity analysis was chosen to estimate the most probable *T. polonicum* parental genotype of Kamut®. This polymorphism included binary data for 2,571 markers. The similarity of sample pairs was estimated by similarity coefficient *S* (McGraw-Hill Science & Technology Dictionary), computed by pattern $S = n_s/(n_s + n_d)$, where n_s represents the number of positive features shared by two samples, and n_d represents the number of features positive for one sample and negative for the other. From the total of

Table 6 Analysis of ancestry by HIest program

s1	s2	s3	Likelihood	Hybrid sample KAMUT
0.33	0.48	0.19	−62.842	Not included in the calculation of parental frequencies of *T. polonicum*
0.17	0.67	0.15	−54.861	Included in the calculation of parental frequencies of *T. polonicum*

P1, P2, P3 possible parental populations (samples from *T. durum*—P1, *T. polonicum*—P2 and *T. dicoccon* convar. *dicoccon*—P3); *s1, s2* and *s3* ancestry indices calculated for KAMUT from 125 diagnostic markers

20 analyzed samples of *T. polonicum*, the highest value of similarity coefficient *S* was detected between Kamut® and PLN 1 sample (0.872), meaning that the most probable naked parent of Kamut® was PLN 1-like sample.

Discussion

TERGAP technology seems to be appropriate for evolution studies. Electrophoretic analysis of PCR fragments revealed polymorphic markers specific for both B and G genomes, and also for some analyzed species. We could identify nine polymorphisms specific for the B genome and three specific for the G genome. Three of these were length polymorphisms, and seven were polymorphisms based on the presence or absence of specific bands. These polymorphic markers caused an almost exact clustering of the same species samples.

Phylogenetic tree constructed on the basis of binary data obtained from electrophoretic analyses was logically divided into two major clusters. These clusters represented the GGAA line and BBAA line taxons and were statistically cogently divided. Samples from the GGAA line were strictly clustered into *T. timopheevii* and *T. araraticum* species.

BBAA line taxons were not clustered as strictly as GGAA taxons, but naked and hulled wheat was clearly distinguished. *T. durum*, *T. turgidum* and *T. carthlicum* samples were divided into separate clusters, except the CAR10 sample, which is regarded to be a hybrid genotype. Hulled wheat samples (*T. dicoccoides, T. dicoccon* and *T. ispahanicum*) clustered together with no further division. *T. dicoccon* and *T. dicoccoides* wheats are very closely related, which is probably the reason for their common grouping in the tree. *T. dicoccon* samples were mixed with *T. ispahanicum* samples, which might be due to a close geographic location during their evolution. The close relationship of these two taxa has been confirmed by other experiments (unpublished data). In naked wheat clusters, *T. turgidum* and *T. carthlicum* samples grouped into common subcluster and *T. turanicum, T. polonicum* and *T. durum* samples formed another subcluster. *T. durum* species was sorted into a separate group in the mentioned subcluster. Samples of *T. turanicum* and *T. polonicum* were mixed in a common subcluster as they are closely related, which is seconded by their very similar morphology. Kamut® wheat sample was included in this group, right on the edge between *T. turanicum* and *T. polonicum* species. According to our dendrogram, it is probable that sample TRN 1 is phylogenetically the oldest one of the *T. turanicum* and *T. polonicum* samples, and during evolution other *T. turanicum* samples used in this study differentiated followed by Kamut® wheat and *T. polonicum* genotypes.

Our multiple statistical analyses classified naked Kamut® wheat to be most closely related to naked *T. polonicum* (Tables 3, 6; Fig. 1), which favors our molecular analyses conclusions and also the earlier hypothesis of Kuckuck (1959). Our results are in contrast to two other studies which had classified Kamut® as the *turanicum* genotype (Kokindova and Kraic 2003; Grausgruber et al. 2004).

Population average pairwise difference analysis suggested the highest interspecific distances between *T. araraticum* and *T. turanicum* sample groups. The highest intraspecific distances were detected inside the *T. turanicum* group and by contrast we identified the *T. araraticum* group to be the most compact and least diversified (Table 3). The highest variability of genetic distances inside the *T. turanicum* group is probably caused by the original classification of Kamut® and BVAL samples into this species. If these two samples are hybrids, it is highly probable that this may induce increased variability in the *T. turanicum* sample group.

These results led us to another experiment to clarify our theory about the hybrid origin of Kamut®. Detection of four haplotypes in *ψLpx-A1_like* gene (Fig. 3) illustrated that the Kamut® sample was most similar to the European *T. dicoccon* convar. *dicoccon* (EUR) samples. This result is interesting, because Kamut® is naked wheat and *T. dicoccon* species has hard glumes, so we suggest it has to be a hybrid sample. Using STEM-hy software we tested the hypothesis that one of four samples listed in Table 5 is a hybrid. Likelihood values indicate that the samples of Kamut® and EUR 42 are most probably hybrid samples, because likelihood values are highest for these two samples' hybrid trees. The relevancy of this model is supported by the AIC criterion (Akaike Information Criterion—a measure of the relative quality of a statistical model for a given set of data), whose values for these samples' hybrid trees are lowest in comparison with alternative hypothesis that PLN 1 and EUR 132 samples are hybrids. This analysis results in the conclusion that Kamut® and EUR 42

(European material from *T. dicoccon* convar. *dicoccon*) resulted from the same crossing.

Using the DArT binary data and HIest software, we proved our previous hypothesis that *T. polonicum* is one of Kamut's® parents. HIest software compared ancestry indices for 125 diagnostic markers of Kamut® with three potential parental lineages P1 (*T. durum*), P2 (*T. polonicum*) and P3 (European *T. dicoccon* convar. *dicoccon*). The values of ancestry indices (s2—Table 6) were highest for Kamut® in the parental lineage of *T. polonicum* (P2). Computing of similarity coefficient *S* revealed that Kamut® was most similar to PLN 1 sample, that is why we suggest that the parents in hybrid crossing were *T. polonicum*, most probably PLN 1-like sample with origin in western Turkey (near Balikesir city) and some genotype of *T. dicoccon* convar. *dicoccon* samples (EUR 42-like). According to microsatellite analysis, Khlestkina et al. (2006) also suggested that *T. polonicum* from the Fertile Crescent could be one of the Kamut's® parents.

Although our results indicate that *T. polonicum* is one of Kamut's® parents, Kamut® is morphologically very similar to *T. turanicum* var. *notabile*. This discrepancy between botanical and genetical approaches to classification of Kamut® may be caused by the hybrid origin of *T. turanicum* species as it was hypothesized by Percival (1921) and Kuckuck (1970). They supposed that *T. turanicum* originated from the hybridization between *T. polonicum* and *T. durum*. According to Dorofeev et al. (1979), *T. turanicum* var. *notabile* is geographically distributed in the countries of Central Asia, Fertile Crescent and in Egypt, Libya and Sudan. It is probable that Kamut® was introduced from the country of its origin to Egypt together with samples of *T. turanicum*.

It is interesting that the sample named BVAL (Etruscan wheat, BVAL 212017) is almost identical to Kamut®. This sample had the same haplotype in the *ψLpx-A1_like* sequence as Kamut®. Within DArT polymorphism, Kamut® and BVAL differed only in the values of 8 markers from the 2,571 analyzed (similarity coefficient 0.9968, unpublished data). We achieved sample BVAL 212017 from the National Inventory of Plant Genetic Resources in Austria (Linz) and its origin is unknown. We could only hypothesise whether this sample was brought to Europe by Etruscans. According to our results we suggest that Etruscan wheat has the same parents as Kamut®. In our statements we allege that these parents were PLN 1-like and EUR 42-like genotypes. Samples PLN 1 and EUR 42 have origin in Turkey, so we might suggest that Kamut® and Etruscan wheat also originate from Turkey. If Etruscan wheat really originates in western Turkey, it could indicate that Etruscans also originates from this area as Greek philosopher Herodotos claimed.

Acknowledgments This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0661-10 and No. APVV-0197-10.

References

Arlequin ver 3.5 http://cmpg.unibe.ch/software/arlequin3. Accessed 10 September 2013

Benedetti S, Primiterra M, Tagliamonte MC, Carnevali A, Gianotti A, Bordoni A, Canestari F (2012) Counteraction of oxidative damage in the rat liver by an ancient grain. Nutrition 28(4):436–441

Brouwer W (1972) Handbuch des Speziellen Pflanzenbaues, Band 1. Weizen—Roggen—Gerste—Hafer—Mais. Paul Parey, Berlin

Degnan JH, Salter LA (2005) Gene tree distributions under the coalescent process. Evolution 59(1):24–37

Dorofeev VF, Filatenko AA, Migushova EF, Udaczin RA, Jakubziner MM (1979) Wheat. In: Flora of Cultivated Plants, vol 1. Kolos, p 346

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Fitzpatrick BM (2012) Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evolut Biol 12:131. doi:10.1186/1471-2148-12-131

Gowayed S (2009) Egyptian wheat. Dissertation, University of Kassel

Grausgruber H, Sailer C, Ruckenbauer P (2004) Khorasan wheat, Kamut® and 'Pharaonenkorn': origin, characteristics and potential. Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 55:75–80

Hammer K, Filatenko AA, Korzun V (2000) Microsatellite markers—a new tool for distinguishing diploid wheat species. Genet Resour Crop Evol 47:497–505

Hammer K, Filatenko AA, Pistrick K (2011) Taxonomic remarks on *Triticum* L. and *xTriticosecale* Wittm. Genet Resour Crop Evol 58:3–10

Khlestkina EK, Röder MS, Grausgruber H, Börner A (2006) A DNA fingerprinting-based taxonomic allocation of Kamut wheat. Plant Genet Resour Charact Util 4(3):172–180

Kokindova M, Kraic J (2003) DNA polymorfizmus pšeníc druhu *Triticum turgidum* L. (article in slovak). Nova Biotechnolica 3(1):52–62

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25(7):971–973

Kuckuck H (1959) Neuere Arbeiten zur Entstehung der hexaploiden Kulturweizen. Z Pflanzenzüchtg 41:205–226

Kuckuck H (1970) Primitive wheats. In: Frankel OH, Bennet E (eds) Genetic resources in plants, their exploration and conservation. IBP Handbooks no. 11, pp 249–266

Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Percival J (1921) The wheat plant. A monograph. Duckworth & Co., London

Quinn RM (1999) Kamut®: ancient grain, New Cereal. ASHS Press, Alexandria, pp 182–183

Rodríguez-Quijano M, Lucas R, Ruiz M, Giraldo P, Espí A, Carrillo JM (2010) Allelic variation and geographical patterns of prolamins in the USDA-ARS Khorasan wheat germplasm collection. Crop Sci 50:2383–2391

Simonato B, Pasini G, Giannattasio M, Curioni A (2002) Allergenic potential of Kamut® wheat. Allergy 57:653–654

Szabó AT, Hammer K (1996) Notes on the taxonomy of farro: *Triticum monococcum, T. dicoccon* and *T. spelta*. In: Padulosi S, Hammer K, Heller J (eds) Hulled wheats, Proceedings of the First International Workshop on Hulled Wheats, Castelvecchio Pascoli, Tuscany, Italy

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

Triticarte™ whole-genome genotyping service for wheat and barley http://www.triticarte.com.au/. Accessed 20 November 2013

Vavilov NI (1951) The origin, variation, imunity and breeding of cultivated plants. Chronica Botanica 13. Chronica Botanica Co., Waltham

APPENDIX VI

Different daylength – different effect

Ivaničová Z., Milec Z., Trávníčková M., Valárik M., Prášil I.T., Pánková K., Snape J.W. and Šafář J.

In: Abstracts of the "Olomouc Biotech 2013. Plant Biotechnology: Green for Good II". Olomouc, Czech Republic, 2013

Different daylength – different effect

Z. Ivaničová[1], Z. Milec [1], M. Trávníčková[2], M. Valárik [1], I.T. Prášil [2], K. Pánková[2], J.W. Snape [3] and J. Šafář [1]

[1] Centre of the Region Hana for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc, ČR
[2] Crop Research Institute, Drnovská 507/73, Prague, CR
[3] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK
ivanicova@ueb.cas.cz

Introduction

The determination of the flowering time (FT) is one of the most important decision in plant life. Plant have to process plenty of external and internal signals to avoid unfavorable conditions (cold, hot and dry period) and to ensure seed propagation.

Major gene classes that control flowering process in wheat include photoperiod (*Ppd*), vernalisation (*Vrn*), and earliness *per se* (*Eps*).

In 1987 Košner identified **delaying** of FT under short days (SD) conditions in Sandra CP3B* (S3B) substitution line. On the contrary, the **acceleration** was reported under long days (LD) conditions (Košner and Pánková, 2002; our observations).
A subtle effect (2-7 days) we examined suggests an *Eps* nature of the novel gene located on chromosome 3B of **Česká přesívka**.

The **bipolar effect** of analyzed gene could be useful in future breeding programs to create the most suitable varieties for local environments.

The precise identification of the gene, its function analysis are the challenges we will be facing in near future.

** 3B chromosome originated from Česká přesívka*

Mapping population

To localize the gene of interest a nearly isogenic lines (**NILs**) mapping population has been created (*Fig. 1*).

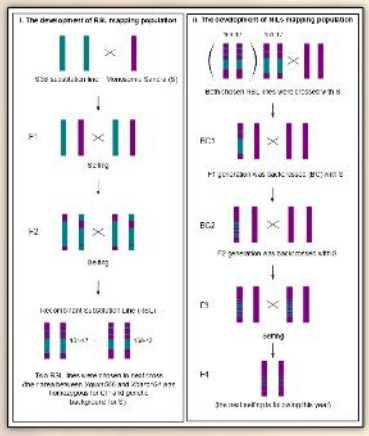


Fig. 1: Scheme of generation of NILs mapping population.

i. genotyping

We used two methods for identification of polymorphic markers: i. polyacrylamide gel electrophoresis **(PAGE)**(*Fig. 2*), ii. **fragmentation analysis** (*Fig. 3*). Besides mapping of known markers, Illumina sequencing data provided new sources of polymorphism which could be analyzed by **KASPar** genotyping system (*Fig. 4*).

Fig. 2: The analysis of length polymorphism by PAGE.
S – Sandra, CP – Česká přesívka, H – heterozygot

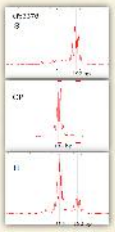


Fig. 3: The analysis of length polymorphism by fragmentation analysis. The amplicons were labeled with PET.
S – Sandra, CP – Česká přesívka, H – heterozygot

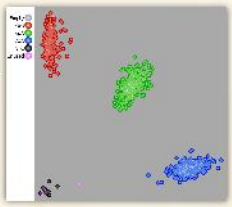


Fig. 4: Graphical view of KASPar genotyping data. (www.lgcgenomics.com.uk)

ii. phenotyping

Accurate phenotype identification is an important step in fine mapping of the gene. We have analyzed the different stage of **apex development** (*Fig. 5*) and the **heading time** (*Fig. 6*) in controlled conditions of **fytotron** chambers as well as in the field experiment.

Fig. 5. Apex development stage: in S3B and S.
S3B – terminal spikelet stage, S – double ridge stage

Fig. 6: Heading time in S and S3B.

Fig. 7: The genetic map of 3B chromosome. The 30 cM region is spanning with 30 new markers.

iii. integration

The combination of data set gained from genotyping (*Fig. 8*) and phenotyping of mapping population was processed by JoinMap® 4 software. The resulted **genetic map** (*Fig. 7*) is showing the order and the relative distance between analyzed markers.

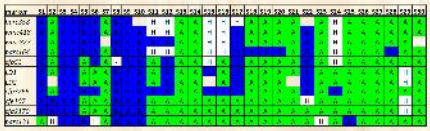


Fig. 8. Genotyping of 30 (S1-S30) individuals of NILs mapping population.
A – Sandra, B – Česká přesívka, H – heterozygot

Conclusions

i. The bipolar effect of flowering time gene was examined:
   - delaying – LD (7 days difference)
   - accelerating – SD (4 days difference)

ii. The locus was localized into 30 cM region spanning *Xbarc164* and *Xcfa2170* markers and it has been saturated by 30 new polymorphic markers.

iii. The Illumina sequencing of 3B chromosomes from CP and S provided 147,852 SNPs, which will be use to delineate the region of interest *QFt.cri-3B.1*. However, efficiency of SNP transformation to PCR based markers is 30 % in allotetraploids (Byers et al. 2012), in allohexaploid wheat is possibly even lower.

iv. Presumably, the FT changes are caused by *Eps* gene. Our observations indicate interaction of analysed gene with photoperiod. Association of *Eps* gene with photoperiod gene (*Ppd-D1*) or temperature is proved in some other studies (Valárik et al., 2006; Kamran et al., 2013).

Literature cited

Appendino, M.L., Slafer, G.A. (2003) Earliness per se and its dependence upon temperature in diploid wheat lines differing in the major gene Eps A-1 alleles. J Agri Scien. 141, 149-154.

Byers, R.L., Harker, D.B., Yourstone, S.M., Maughan, P.J., Udall, J.A.: (2012) Development and mapping of SNP assays in allotetraploid cotton. Theor Appl Genet 124, 1201-1214.

Kamran, A., Randhawa, H., Iqbal, M., Navabi, A., Pozniak, C., Spaner, D. (2013) Earliness per se QTLs and their interaction with the photoperiod insensitive allele Ppd-D1a in the Cutler 3 AC

Barrie spring wheat population. Theor Appl Genet.

Košner, J.: (1987) Scientia Agriculturae Bohemoslovaca 19: 33 - 45.

Košner, J., Pánková, K.: (2002) The effect of chromosome 3B gene/s of Česká přesívka on vernalisation response, photoperiod sensitivity and earliness of wheat. Czech J. Genet. Plant Breed. 38, 41-49.

Van Ooijen, J.W., (2006) JoinMap® 4. Software for the calculation of genetic linkage maps in experimental populations. KyazmaB.V., Wageningen, Netherlands.

Acknowledgments

This work has been supported by the Czech Science Foundation (P501/10/1778) and MSMT CR, EU Operational Programme Research and Development for Innovations (ED0007/01/01) and Internal Grant Agency (PrF-2012-003).

APPENDIX VII

Chasing a new flowering time gene

Ivaničová Z., Šafář J., Trávničková M., Pánková K., Bartoš J., Milec Z. and Doležel, J.

In: Book of abstracts EUCARPIA Cereal Section ITMI Conference in Weringerode, Lohwasser U. and Börner, A. (eds.), 156, 2014

Chasing a new flowering time gene

Zuzana Ivaničová[1], Jan Šafář[1], Martina Trávníčková[2], Kateřina Pánková[2], Jan Bartoš[1], Zbyněk Milec[1] and Jaroslav Doležel[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc, Czech Republic

[2] Crop Research Institute, Drnovská 507/73, Prague, Czech Republic

Introduction

Bread wheat (*Triticum aestivum* L.) is a staple food for 40 % of the world's population. Population growth and extreme weather fluctuations call for development of new cultivars resistant to environmental stress and with increased and more stable yield. The yield can be modified by a number of ways. One of them being optimization of flowering time to local climatic conditions.

Apart from major flowering time genes, which have already been determined, there are minor genes, some of which have been identified as QTLs. However, most of them have not been discovered yet and although our current knowledge of pathways regulating flowering time is improving, it is far from completion.

Clearly, there is a need to identify alleles and genes associated with flowering, which may be useful in wheat improvement.

The main objective of our work is to localize a new flowering time gene. We have detected 12-day difference in heading time between Paragon (P) and Kaerntner Frueher (KF) varieties of bread wheat. There were no differences in the main flowering time genes (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *vrn-B3* and *Ppd-D1b*) between these two varieties. Thus, we hypothesize that there may be a novel gene/allele involved.

F_2 mapping population

1.

KF and P were crossed in order to develop F_2 mapping population.

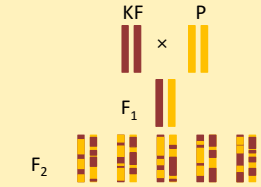


Figure 1: **Schematic illustration of F_2 population development.**

DArT analysis

3.

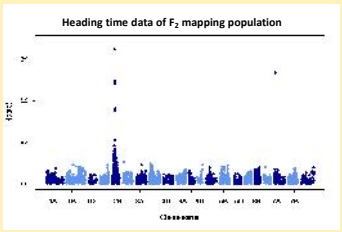


Figure 3: **Manhattan plot of F_2 mapping population heading time data.**
The y-axis plots -log10 (P-values) and the x-axis plots the physical position of over 13,000 DArT markers. The most significant loci were located on chromosome 2B.

4.

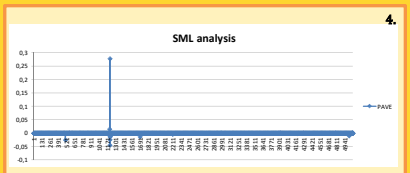


Figure 4: **Results of SML (Statistical machine learning) analysis.**
The PAVE value determines contribution of each marker to phenotype. Heading time difference was associated with DArT marker 985149 and was located within 0.75 cM fraction from the putative gene. (Bedo *et al.* 2008)

Heading time difference

a) 12-day difference in heading time between P and KF varieties,
b) temperature-dependent effect of the hypothetic gene in field and in controlled conditions (16/8 h, 20/16 °C) - lower temperature prolonged heading time difference up to 28 - 30 days in field and growth room, respectively.

2.

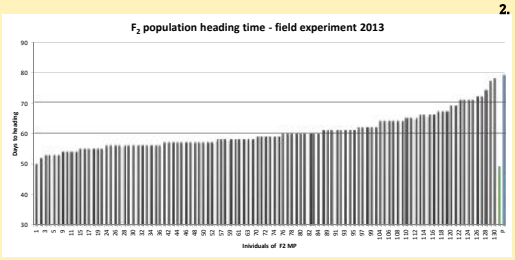


Figure 2: **Heading time difference within individuals of F_2 mapping population.**
Comparison with mean heading time of 5 replications of KF and P.

Identification of source of variation

The most possible candidate gene was *Ppd-B1*. To confirm/disprove the results of QTL analysis, KASP genotyping and whole gene sequencing were performed.

5.

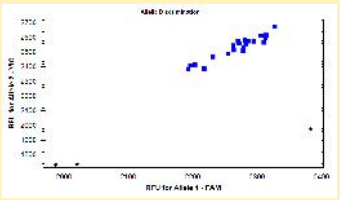


Figure 5: **The results of KASP analysis of KF and P with allele specific primers.**
Dot plot shows presence of the same *Ppd-B1* allele in both lines.
(L. Dixon, pers. communication)

6.

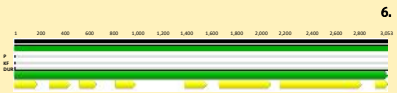


Figure 6: **Schematic comparison of P, KF and DUR (*Triticum durum*) - *Ppd-B1* sequence.**
Sequences of Paragon (DQ885762.2), Kaerntner Frueher and *T. durum* (AB692906) was aligned and analyzed for the presence of any kind of DNA polymorphisms. We did not detected any variation in sequence of Kaerntner Frueher in both, the gene and the promoter region. In the picture is illustrated identity of all three sequences and CDS of the *Ppd-B1* gene.

Conclusions

In order to identify the source of heading time variation between Paragon and Kaerntner Frueher varieties:
a) F_2 mapping population was created
b) QTL analysis from the DArT data was performed
c) the gene closely linked to DArT markers was identified

To confirm/reject *Ppd-B1* gene as a possible source of heading time variation:
a) 985149 DArT marker sequence was BLASTed against 2BS survey sequence
b) *Ppd-B1* gene of cv. Kaerntner Frueher in all length was sequenced
c) KASP analysis was performed to identify the presence of the same/different alleles of *Ppd-B1* gene

Our analyses did not prove that *Ppd-B1* was the gene influencing the trait, and the BLAST against 2BS survey sequence did not find any significant hit.
A possible explanation of the heading time difference could be copy number variation (CNV) of *Ppd-B1* gene. Díaz *et al.* (2012) identified early and day neutral heading time in different commercial varieties of wheat (Paragon, Sonora64, Récital, Timstein, C591) that was caused by increased CNV of *Ppd-B1* allele. Their conclusions are on line with our identification of Paragon as a late flowering variety and the ability of Kaerntner Frueher to flower in short days. Thus our next step is to determine differences in CNV between Paragon and Kaerntner Frueher by quantitative PCR.

References

Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012). PLoS ONE 7(3): e33234. doi: 10.1371/journal.pone.0033234.

Bedo J, Wenzl P, Kowalczyk A., Kilian A (2008). doi:10.1186/1471-2156-9-35.

Acknowledgements

This work has been supported by Internal Grant Agency (PrF-2012-003) and Ministry of Agriculture of the Czech Republic (award no MZE002700604). ZM was supported by the Operational Program Education for Competitiveness — European Social Fund (project CZ.1.07/2.3.00/20.0165).

APPENDIX VIII

High-throughput gene resources and phenotyping - the key factors in fine mapping of new Eps gene

Ivaničová Z., Milec Z., Trávníčková M., Abrouk M., Valárik M., Prášil I.T., Pánková K., Snape J.W. and Šafář, J.

In: Sborník abstrakt, Bulletin České společnosti experimentální biologie rostlin, "6. Metodické dny". Seč, Česká republika, 2014

High-throughput gene resources and phenotyping - the key factors in fine mapping of new *Eps* gene

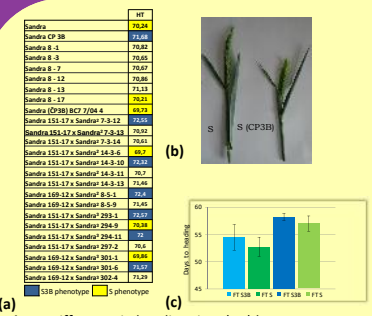
Zuzana Ivaničová[1], Zbyněk Milec[1], Martina Trávníčková[2], Michael Abrouk[1], Miroslav Valárik[1], Ilja T. Prášil [2], Kateřina Pánková[2], John W. Snape [3], Jan Šafář[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc, CR
[2] Crop Research Institute, Drnovská 507/73, Prague, CR
[3] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK

Introduction

One-day acceleration and two-day delay (Fig. 1) of heading time was identified in Sandra (CP3B) substitution line under both short and long day conditions, respectively. To identify locus conferring this phenotype, the 3B chromosome from Czech wheat landrace Česká přesívka (CP) was substituted to Sandra (S) variety. Afterwards, RSLs (Recombination Substitution Lines) mapping population from cross of Sandra (CP3B) and Sandra was developed. The *Eps* locus was localized within 30 cM region (*barc164 - cfa2170)* of long arm of chromosome 3B and designated as *QFt.cri-3B.1* (Fig. 2, 3) (Pánková et al. 2008).

To precisely determine the gene of interest, mapping population with higher mapping power (e.g. Nearly Isogenic Lines-NILs) and development of new markers were introduced. Important source of markers are (i) publicly available SSR and STS markers, (ii) DArT markers, (iii) Single Nucleotide Polymorphisms (SNPs) markers. Combination of next generation sequencing and flow cytometry approach have proved to be credible source of SNPs markers (Paux 2008, Vrána et al. 2000). Moreover, SNPs can be easily transformed into sequence based or KBioscience competitive allele-specific PCR (KASP) markers.



	HT
Sandra	70,24
Sandra CP 3B	71,68
Sandra 8 -1	70,82
Sandra 8 -3	70,65
Sandra 8 - 7	70,67
Sandra 8 - 12	70,86
Sandra 8 - 13	71,13
Sandra 8 - 17	70,21
Sandra (CP3B) BC7 7/04 4	69,73
Sandra 151-17 x Sandra² 7-3-12	72,55
Sandra 151-17 x Sandra² 7-3-13	70,92
Sandra 151-17 x Sandra² 7-3-14	70,61
Sandra 151-17 x Sandra² 14-3-6	69,7
Sandra 151-17 x Sandra² 14-3-10	72,32
Sandra 151-17 x Sandra² 14-3-11	70,7
Sandra 151-17 x Sandra² 14-3-13	71,46
Sandra 169-12 x Sandra² 8-5-1	72,4
Sandra 169-12 x Sandra² 8-5-9	71,45
Sandra 151-17 x Sandra² 293-1	72,57
Sandra 151-17 x Sandra² 294-9	70,38
Sandra 151-17 x Sandra² 294-11	72
Sandra 151-17 x Sandra² 297-2	70,6
Sandra 169-12 x Sandra² 301-1	69,86
Sandra 169-12 x Sandra² 301-6	73,57
Sandra 169-12 x Sandra² 302-4	71,29

Fig. 1: Difference in heading time (HT) between recurrent parent -Sandra and substitution line-Sandra(CP3B). (a) days to heading in field experiment in 2013 (b) difference in heading time - Sandra is earlier than Sandra (CP3B) in LD condition (b) results of experiment in controlled condition (24˚C/20˚C day/night, long day – 16 hours).

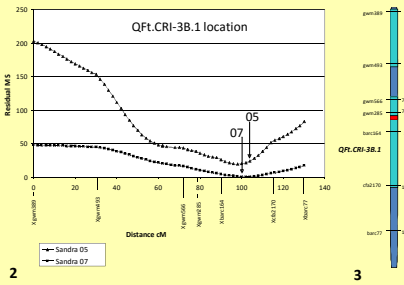


Fig. 2: QTL profiles from marker regression analysis. Analysis of Sandra//Sandra 3B/CP 3B RILs (QTL café) from two experiments: (i) growth room - short days, winter 2005/06 (05), (ii) glasshouse summer – long days 2007 (07) (Pánková et al. 2008).

Fig. 3: Genetic map and *QFt.CRI-3B.1* location. (Pánková et al. 2008).

Results

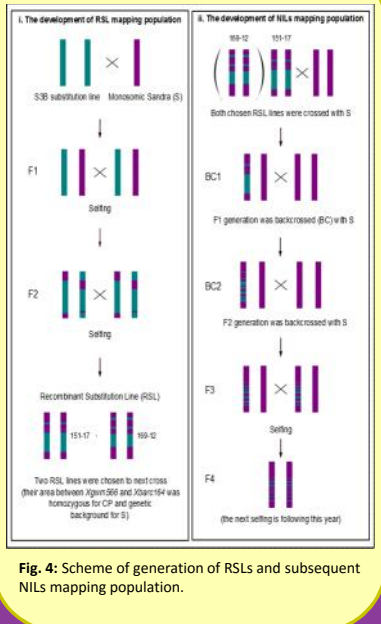


Fig. 4: Scheme of generation of RSLs and subsequent NILs mapping population.

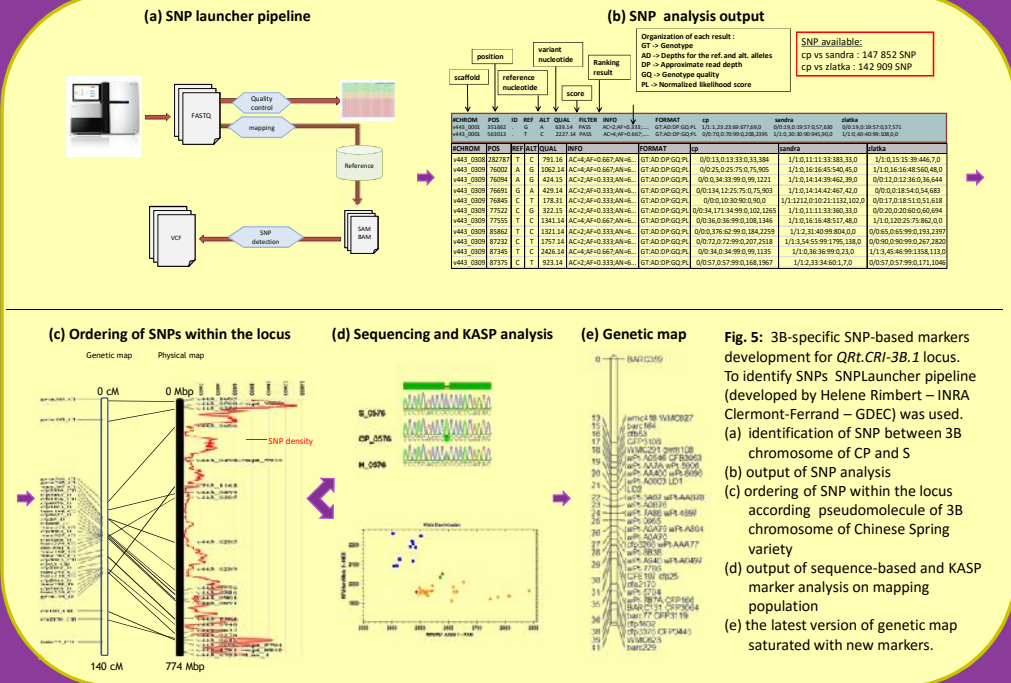


Fig. 5: 3B-specific SNP-based markers development for *QRt.CRI-3B.1* locus. To identify SNPs SNPLauncher pipeline (developed by Helene Rimbert – INRA Clermont-Ferrand – GDEC) was used.
(a) identification of SNP between 3B chromosome of CP and S
(b) output of SNP analysis
(c) ordering of SNP within the locus according pseudomolecule of 3B chromosome of Chinese Spring variety
(d) output of sequence-based and KASP marker analysis on mapping population
(e) the latest version of genetic map saturated with new markers.

Conclusions

For precise mapping of *Eps* locus (i) the NILs mapping population is being developed (Fig. 4) and (ii) polymorphic markers in *QRt.CRI-3B.1* locus are being determined. To date, 274 polymorphic DArT, SSR and STS markers were identified. Later on, sequencing of 3B chromosome from Česká přesívka (CP) and Sandra (S) lead to identification of 27,388 SNPs within the *QFt.cri-3B.1* region. Identified SNPs have been used for sequence-based and KASP markers development and subsequently to saturate the region of interest (Fig. 5). The efficiency of SNP markers conversion was nearly 80 % for KASP markers and 30 % for markers derived from Sanger sequencing. In total, 55 % of converted markers was possible to use in genetic mapping. So far, 30 markers were used to saturate the region between *barc164* and *cfa2170* (Fig. 5(e)).

Both F_5 NILs mapping population as well as obtained SNP genotyping sources will be used for precise fine mapping of the gene which will be eventually followed by positional cloning. Candidate gene/genes analysis and its detailed characterization will provide novel information which could lead to deeper understanding of fine tuning flowering time in wheat. The bipolar effect of analyzed gene could be useful in future breeding programs to create the most suitable varieties for local environments.

Pánková, K., Milec, Z., Simmonds, J., Leverington-Waite, M., Fish, L., Snape, J.W.: (2008). Euphytica. 164, 779-787.
Paux, E. Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A, et al.: (2008). Science. 322, 101-104.
Vrána, J., Kubaláková, M., Šimková, H., Čihalíková, J., Lysák, M.A., Doležel, J.: (2000). Genetics. 156, 2033-2041.

Acknowledgements
This work has been supported by the Czech Science Foundation (P501/10/1778) and MSMT CR and EU (Operational Programme Research and Development for Innovations No. ED0007/01/01).

APPENDIX IX

Identification and characterization of a new *Vrn-A1f-like* allele responsible for flowering lateness in wheat

Ivaničová Z., Jakobson I., Reis D., , Milec Z., Abrouk M., Doležel J., Järve K. Šafář J. and Valárik M.

In: Proceedings 13th International Wheat Genetics Symposium, p. 173, 2017

Identification and characterization of a new *Vrn-A1f-like* allele responsible for flowering lateness in wheat

Zuzana Ivaničová[1], Marie Šoborová[1], Irena Jakobson[2], Diana Reis[2], Zbyněk Milec[1], Michael Abrouk[1], Jaroslav Doležel[1], Kadri Järve[2], Jan Šafář[1] and Miroslav Valárik[1]

[1]Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-783 71 Olomouc - Holice, Czech Republic,
[2]Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

Achievements

Flowering is a crucial process in plant life, which impacts crop yield. Here we identify tetraploid wheat *Triticum militinae* (2n=4x=28, AAGG) as a resource for mining new genes/alleles influencing flowering time. We observed flowering time variation in a mapping population of doubled haploid (DH) lines. It was developed from a cross between introgressive line 8.1 and elite bread wheat cv. Tähti. Line 8.1 carries introgressions from *T. militinae*, in Tähti background (**Fig. 2**). The analysis of DH mapping population revealed presence of seven loci influencing flowering time. The most significant QTL for the flowering time variation was identified within the introgressed region on chromosome 5A and its largest effect was associated with the *VRN-A1* locus (**Fig. 1**), covering up to 67% of phenotypic variation. The analysis of F_2 mapping population developed from cross DH81 × cv. Mooni confirmed the effect of this locus on flowering time. DNA sequence analysis revealed the origin from *T. militinae* and the allele was designated as *VRN-A1f-like* (KT696537). The allele incurred a delay of 1.9 - 18.6 days in flowering in different growing conditions (**Tab. 1**). Comparison *VRN-A1f-like* allele to *VRN-A1a* sequences from the common wheat parental lines of the mapping populations revealed major mutations in the promoter region as well as in the first intron, including a MITE insertion and a large deletion (**Fig. 4**). Moreover, the allele was identified to be responsible for spring habit emergence in 80% of the analyzed tetraploid wheat varieties with AAGG genome (*T. timopheevii* sp.). Finally, analysis of difference in the relative expression level between parental lines of two mapping populations (DH81, cv. Tähti and cv. Mooni) was carried out. The identification and quantification of the effect of the *VRN-A1f-like* allele from *T. militinae* provides a valuable source of new alleles suitable for wheat improvement as well as for studying fine regulation of flowering pathways in wheat.

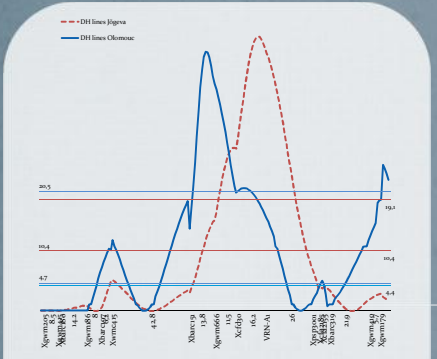


Figure 1: Location of the flowering time QTL on chromosome 5A.
Simple interval mapping analyses of the DH mapping population based on DH mapping population flowering time data from Olomouc (solid line) and Jõgeva (dashed line) localised the flowering time QTL in the region between markers *Xbarc151* and *Xcfa2155*, with the peak at the *Xgwm666*/*VRN-A1* locus. The suggestive threshold line (likelihood ratio statistic LRS: 4.7/ 4.4), the significant threshold line (LRS: 10.4/10.4) and highly significant threshold line (LRS: 20.5/19.1) are displayed for Olomouc/ Jõgeva, respectively. The threshold lines of respective QTL lines are marked with the same colour. The distance in cM between markers is shown on x axis.

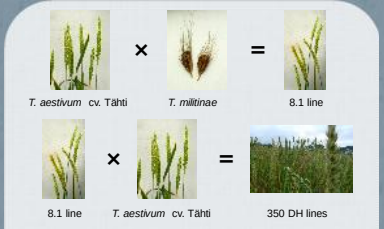


Figure 2.: Schematic illustration of DH mapping population development.

Experiment	Parental lines	Background	Location	Greenhouse/field	FT (days)	MIN (days)	MAX (days)	SD	VRN-A1f-like contribution	FT difference between parents
1a	8.1/Täh6	Tähti	O	f	90	81	98	3	+1.9	+3.5
1b	8.1/Täh6	Tähti	J	g	76	65	87	5	+5.8	+5. 5
2	DH97/DH81	Tähti	O	f	86	82	94	2	+1.7	+6.5
3a	DH81/Mooni	Mooni	O	f	84	78	90	2	+3.3	+9.8
3b	DH81/Mooni	Mooni	T	g	66	52	89	9	+18.6	+29.9

Table 1: Statistical analysis of phenotyping data.
Phenotyping experiments were performed under controlled environmental conditions (greenhouse; 1b, 3b), or in field conditions (1a, 2, 3a). The experiment 1 was conducted on the DH mapping population; the experiment 2 was performed on population F_2 (DH397 x DH81) and experiment 3 was performed on population F3(Mooni x DH81).
g-greenhouse; f – field; O – Olomouc; T – Tallinn; J – Jõgeva; FT – flowering time

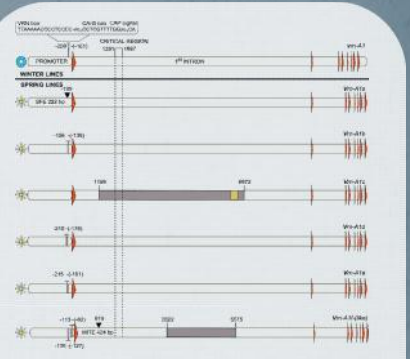


Figure 5: The *VRN-A1* alelles.
The schematic representation of sequence changes in ***dominant*** *Vrn-A1* alleles present in ***spring wheat*** compared with **recessive** *vrn-A1* (**winter wheat**). Orange arrows – exons, white rectangles – introns, rectangle before first exon – promoter, insertion - black triangle, deletion - grey rectangle, yellow rectangle - duplication. Their position is indicated by number of base pairs from the START codon in comparison to *vrn-A1* allele. The NCBI database sequences used for schematic comparision: *vrn-A1* (AY616455.1), *Vrn-A1a* (AY616458.1), *Vrn-A1b* (AY616461.1), *Vrn-A1c* (AY747599.1), *Vrn-A1d* (AY616462), *Vrn-A1e* (KT361213), *Vrn-A1f* (GQ451750.1) and *Vrn-A1f-like* (KT696537.1).
SFE – Spring Foldback Element, MITE – Miniature Inverted-repeat Transposable Element.

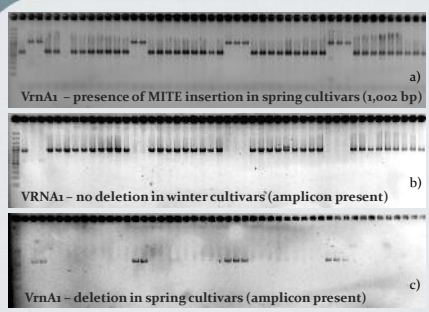


Figure 4.: *VRN-A1f-like* allele is present in spring but not in winter *T. timopheevii* sp. wheat.
All investigated spring lines has MITE insertion (424 bp) and 2,753 deletion in first intron.
a) 1,002 bp amplicon – MITE insertion, 569 bp amplicon – intact sequence of first intron
b) presence of amplicon in cultivars with intact first intron (no deletion present)
c) presence of amplicon in cultivars with 2,753 deletion in first intron

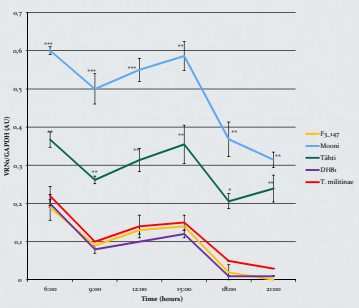


Figure 5.: *VRN-A1f-like* gene expression analysis confirmed FT differences among analysed lines.
Parental lines of mapping populations - cv. Tähti, cv. Mooni, DH81 line, F_3 (Mooni × DH81) 147 line and *T. militinae* (source of *Vrn-A1f-like* allele) was used for expression analysis. Samples were collected from the dawn (6:00) up to the night (21:00), every three hours, from the younges fully unfolded leaf of 20 days old plants. The asterix shows significancy level from one way ANOVA (* < 0,05; ** < 0,01,; *** <0,001) in comparison to *T. militinae* expression levels.

Conclusions

❖Lateness of DH81 line is caused by ***Vrn-A1f-like*** allele.

❖*Vrn-A1f-like* allele has changes in promotor and first introne.

❖Analysis of tetraploid AAGG wheat cultivars (*T.t.* subsp. *timopheevii*, *T.t.* subsp. *araraticum*) showed, that deletions in promoter are present in both – spring and winter cultivars (**Fig. 4**).

❖Presence of MITE insertion and/or deletion in the first introne of *Vrn-A1f-like* allele is crucial for emergence of spring habit (**Fig. 4**).

❖*Vrn-A1f-like* allele expression analysis (**Fig. 5**) correlates with flowering time delay.

This work has been supported by the Operational Program Education for Competitiveness — European Social Fund (project CZ.1.07/2.3.00/20.0165), the award LO1204 from the National Program of Sustainability I, by the Czech Science Foundation (grant award 14-07164S), by the Czech Ministry of Education, Youth and Sports (OP VK CZ.1.07/2.3.00/20.0165) and by institutional research funding IUT 193 of the Estonian Ministry of Education and Research.

Palacký University Olomouc

Faculty of Science

Department of Botany

and

Centre of the Region Haná for Biotechnological and Agricultural Research

Institute of Experimental Botany

Olomouc

Zuzana Ivaničová

Mapping and identification of flowering time genes in bread wheat (*Triticum aestivum* L.)

P1527 Biology – Botany

Summary of Ph.D. Thesis

Olomouc

2017

Ph.D. thesis was carried out at the Department of Botany, Faculty of Science, Palacký University Olomouc, between the years 2012-2017.

Ph.D. candidate: **Mgr. Zuzana Ivaničová**

Supervisor: **RNDr. Jan Šafář, Ph.D.**
Institute of Experimental Botany, Olomouc

Reviewers: **Assoc. Prof. Ing. Petr Smýkal, Ph.D.**
Palacký University, Olomouc

Assoc. Prof. RNDr. Eduard Kejnovský, CSc.
Institute of Biophysics, Brno

Prof. Ing. Vladislav Čurn, Ph.D.
University of South Bohemia, České Budějovice

The evaluation of this Ph.D. thesis was written by Prof. Ing. Aleš Lebeda, DrSc., Department of Botany, Faculty of Science, Palacký University Olomouc.

The summary of the Ph.D. thesis was sent for distribution on ……

The oral defense will take place on …… in front of the Commission for the Ph.D. thesis of the Study Program Botany in a conference room of the Institute of Experimental Botany, Šlechtitelů 31, Olomouc-Holice.

The Ph.D. thesis is available in the Library of the Biological Departments of the Faculty of Science at Palacký University Olomouc, Šlechtitelů 11, Olomouc-Holice.

Prof. Ing. Aleš Lebeda, DrSc.
Chairman of the Commission for the Ph.D.
Thesis of the Study Program Botany
Department of Botany, Faculty of Science
Palacký University Olomouc

CONTENT

1 INTRODUCTION

The bread wheat (*Triticum aestivum* L.) is one of the most important crops for humankind. It is a staple food for 30 % of population. Wheat fields cover the largest area of all crops (http://faostat3.fao.org/). Currently, the wheat yield is threatened by various abiotic and biotic stresses. The abiotic stresses are dangerous especially for reproductive phase of wheat development. The low temperature in the early spring can lead to damage of floret parts and hot and dry conditions during following period can influence grain filling. The flowering in optimal time enables using available sources and avoiding unsuitable conditions in certain microenvironment. The study of flowering time determination is important for an understanding of this complex process.

The pioneering flowering pathway studies started in the model organism of *Arabidopsis thaliana*. This organism has many advantages (small genome size, short life cycle, available mutant phenotypes and comprehensive sequencing data) which made it the ideal object for determination of flowering process studies [1]. The knowledge gained by these studies showed complexity of flowering pathway. These studies emphasize an importance of external conditions and preparedness of the plant itself for flowering before the initiation of flower development [2]. Usage of these data in analyses of agriculturally important plants (wheat, barley and rice) proved conservation of some but not all biochemical pathways (for review see: [3]). This just underlined the importance of study of these objects.

Determination of genes influencing flowering time in wheat is hampered by its huge 17-Gbp genome, its hexaploid nature and presence > 90 % repetitive elements in the genome. In addition, the influence of loci with just a minor effect on flowering time is shielded by manifestation of major loci. That is why the genes conferring photoperiod sensitivity and vernalization requirement, their allelic variations and mutual interactions are the most studied ones [4–12]. Recent progress in genotyping platforms, phenotyping methods and genetic mapping techniques also enabled identification of several minor loci which belong into the group of *Earliness per se* genes [13–15]. However, the absolute knowledge of determination of flowering time is still far away.

Landrace varieties and both wild and domesticated bread wheat progenitors are important source of new alleles controlling flowering time. The study of these forgotten varieties, determination of their evolutionary relationship and characterization at the genetic level was

the main aim of another study presented in doctoral thesis. These pieces of information are important for finding cultivars which can be used in the breeding process.

In the presented study, we aimed to identify new genes/alleles influencing flowering time in *T. millitinae*. The doubled haploid mapping population was used for primary analysis. It consists of individuals with introgressions from *T. millitinae* in *T. aestivum* cv. Tähti genetic background. Overall, seven loci influencing flowering time have been identified. In the following study, we focused on the most significant locus located on chromosome 5A. This locus was responsible for up to 70 % phenotypic variation and located in the region of *VRN-A1* gene. The sequence analysis revealed unique variation in the first intron and led to identification of new *Vrn-A1f-like* (KT696537) allele.

In the second part of the work we determined copy number variation of *Ppd-B1* gene as cause of variation between spring cultivars of *T. aestivum* – Kaerntner Frueher (KF) (three copies of *Ppd-B1*) and Paragon (P) (one copy of *Ppd-B1*). However, we discovered significant variation within lines with three copies of *Ppd-B1* gene. This group can be divided to early (KF – like) and late (P – like) lines. In addition, this phenomenon was detectable throughout subsequent generations (F_2- F_7). We analyzed *Ppd-B1a* expression and methylation pattern in chosen individuals of F_7 generation in order to reveal the nature of this observation.

2 AIMS OF THE THESIS

The main aim of the thesis was identification and characterization of new genes/alleles responsible for determination of flowering time. This aim was accomplished by two different topics which are represented by two separated parts.

I. Identification and characterization of new *Vrn-A1f-like* allele responsible for delayed flowering.

The aim of the first part of the thesis was determination and following characterization of the gene responsible for the difference in flowering time within doubled haploid mapping population. This consists of individuals which contain different combination of introgressions from *T. millitinae* in elite bread wheat cultivar Tähti.

II. Analysis of expression profile and methylation pattern of lines with three copy of *Ppd-B1* gene as a possible cause of intragroup heading time variation.

The second part of the thesis dealt with identification of gene responsible for difference in flowering time between two bread wheat cultivars – Paragon and Kaerntner Frueher. The aim was further complemented by analysis of lines which have a common character – three copies of *Ppd-B1* gene, although they significantly differed in the flowering time.

3 MATERIALS AND METHODS

3.1 I. Identification and characterization of new *Vrn-A1f-like* allele responsible for delayed flowering.

Plant material

We used 314 lines of doubled haploid mapping population for identification of genes/alleles responsible for difference in flowering time. These lines contained different combination of introgressions from tetraploid wheat with AAGG genome *T. militinae* in elite bread wheat cultivar Tähti (located on chromosomes 1AS, 2AS, 2AL, 4AL, 5AL, 7AS, 1BS, and 5BS). Furthermore, two verification mapping population (94 lines F_2 individuals coming from cross cv. Mooni×DH81 and 98 F_2 individual coming from cross DH397×DH81 and 96 F_2 individuals from reciprocal cross DH81×DH397) were used for determination of the effect of identified loci in different genetics backgrounds. DH397 line was selected from DH mapping population and contained introgressions on chromosomes 1A, 2AL and 7A.

Phenotyping of mapping populations

The phenotyping experiments were performed during three different seasons under both field and controlled conditions. Prior to planting in the field, plants were synchronized by soaking the seeds in tap water for 24 hours at room temperature, followed by incubation of the seeds on wet filter paper for two days at 4°C and one day at 25°C in plastic Petri dishes.

Preliminary experiment with 113 DH lines (six replications) was performed in Jõgeva, Estonia (winter 2009/2010) in greenhouse (16/8 hours day/night, temperature 20–23°C). The whole DH mapping population (314 lines – four replications) were further analyzed in the field conditions in Olomouc during season 2013.

94 lines of F_3 Mooni×DH81 mapping population (four replicates) were analyzed under controlled conditions in Tallinn, Estonia in 2013 (16/8 hours day/night, 20–23 °C). 80 lines selected from this mapping population were phenotyped in field conditions in Olomouc during season 2014.

F_2 plants originating from both reciprocal crosses (DH397×DH81 and DH81×DH397) were studied in field conditions in Olomouc (in 2014).

Genotyping

The presence of introgressions in all lines was determined by SSR markers identified in the previous study. These markers were shown to be polymorphic between cv. Tähti and DH81. Overall 57 markers were mapped in DH mapping population. For genotyping of verification mapping population we used at least one marker per introgression. Mapping of flowering time

genes (*Ppd-B1*, *VRN2* and *VRN3*) presented in introgression were done using sequence based markers.

QTL mapping and statistical analysis

The QTL analysis was done using statistic software MapManager QTX Version b16 [16]. Genetic distances were established using the Kosambi mapping function. The threshold value for Likelihood ratio statistic (LRS) for suggestive ($P < 0.05$), significant ($P < 0.001$) and highly significant ($P < 0.0001$) loci was calculated for each experiment using permutation test at 1000 iterations. Single marker regression was used to associate phenotype with marker in each experiment. The position of detected QTLs was determined using simple interval mapping. The confidence interval was established by bootstrap analysis with 1000 iterations using the same software. Statistical significance was determined using single factor analysis of variance (ANOVA).

3.2 II. Analysis of expression profile and methylation pattern of lines with three copy of *Ppd-B1a* gene as a possible cause of intragroup heading time variation.

Plant material

We used 92 individuals of F_2 mapping population which was created from cross between two bread wheat cultivars Kaerntner Frueher (KF) and Paragon (P) for the flowering time determination. Both parental lines did not differ in allelic compositions of main flowering time genes (*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*, *vrn-B3*, *Ppd-D1b* and *Ppd-A1b*). Further, the next generations until F_7 generation of mapping population was developed by single seed descent method.

QTL analysis

QTL analysis using DArT markers was performed to indentify location of loci responsible for variation in flowering time. It was carried out by Diversity Arrays Technology (DArT, Canberra, Australia) using 92 individuals of F_2 mapping population.

Copy number determination

The number of *Ppd-B1* copies was determined differently for F_2 individuals and for individuals of F_6 and F_7 generation. Number of copies for F_2 population and both parents was determined at iDNA Genetics (Norwich, UK) by Taqman® assay as described in Díaz et al. [17]. The number of *Ppd-B1*copies for F_6 and F_7 generation was determined by sequence-based marker. The two SNPs localized in the sequenced area of 5´UTR region was used for distinguishing of different copy number. Primer pair was designed manually to overlap SNPs located at the positions +10,427bp (G-A) and +10,656bp (A-G) from START codon (forward primer GACTCCTGCCATGAGTTTTGATG and reverse primer ACCGCAGTGTG ACTTCGATTATC). Amplicons were purified and sequenced as described in Ivanicova et al.

[18]. The sequences were trimmed and assembled using Geneious 5.6.4 (http://www.geneious.com).

RNA extraction and expression analysis

RNA was extracted from leaves of 20-day-old seedlings using commercial kit RNasy Plant Mini Kit (Qiagen); leaves were collected from all lines in three biological replicates three hours after dawn when the *Ppd-B1* expression reaches the maximum level. Digestion of DNA during RNA purification was performed using the RNase-Free DNase Set (Qiagen). cDNA was synthesized with Transcription High Fidelity cDNA Synthesis Kit (Roche) following manufacturer´s protocol with 2 µg of total RNA and anchored-oligo $(dT)_{18}$ primers.

The expression level was determined using reverse transcription-qPCR (RT-qPCR). RT-qPCR was performed using qPCR 2x SYBR Mt. Mix (Top-Bio) on CFX96TM Real-Time PCR Detection System (Bio-Rad, USA). The expression level of *Ppd-B1* was standardized against reference gene *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) and data analyzed with software CFX Manager 3.0 (BioRad) using $2^{-\Delta\Delta Cq}$ method. Three replicate PCR amplifications were performed for each sample. The transcript level of target gene *Ppd-B1* in control plants was designated as 1.0. Primers were designed with Primer Express® Software v3.0.1 (ThermoFisher Scientific, USA): primers for *Ppd-B1* were Ppd-B1_101F (CGCCACTGCATGTACCAAGTTA) and Ppd-B1_101R (CTGTCAGAACAAGGTCGATG TTG); primer efficiency E=102.8 % and correlation coefficient R^2 = 0.991. Primers for reference gene *GAPDH* were TTAGACTTGCGAAGCCAGCA and TTAGACTTGCGAAG CCAGCA as described in Sun et al. [11].

Methylation status determination

Genomic DNA was extracted from young plant leaves of F_7 individuals and parental varieties (all in three biological replicates) using commercial kit Invisorb Spin Mini Plant Kit (Stratec Molecular) following manufacturer´s instruction. To determine methylation status, bisulfite conversion (BS) was performed using EZ DNA Methylation-Gold™ Kit (Zymo Research). We focused on the promoter region described in Sun et al. (2014) as "Region II" for it is associated with significant changes in methylation level between one and three copies of *Ppd-B1*.

a) genomic DNA amplification

Unconverted DNA was amplified with newly designed forward primer PromF2 (GCCTTACGCACATCATCAGC) and reverse primer PromR2 (GGTGACGTGGACGAAA TGGA).

b) bisulfite converted DNA amplification

Amplification of converted DNA was carried out with primers for Region II described in Sun et al. (2014). PCR amplicons from both amplifications were purified using Agenocourt AMPure XP (Beckman Coulter, USA) and cloned into pCR™4-TOPO® TA vector (TOPO® TA Cloning® Kit for Sequencing, with One Shot® TOP10 Electrocomp™ E. coli; Invitrogen, USA) according to manufacturer instructions. Clones from all samples were amplified in total volume of 20 µl. PCR products were purified and sequenced as described in Ivanicova et al.

[18]. The resultant sequences were trimmed and assembled using Geneious 5.6.4 (http://www.geneious.com). Methylation status was determined with Kismeth online tool available at http://katahdin.mssm.edu/kismeth /revpage.pl.

4 SUMMARY OF RESULTS

I. Identification and characterization of new *Vrn-A1f-like* allele responsible for delayed/late flowering.

The new *Vrn-A1f-like* allele was discovered during analysis of doubled haploid mapping population. This population was developed as described in Jacobson et al. [19,20]. The parental lines of this mapping population were elite bread wheat cultivar Tähti and line 8.1. Line 8.1 consists of introgressions of *T. militinae* in cv. Tähti background. The QTL analysis of doubled haploid mapping population showed presence of seven loci which influenced flowering time. The locus located in the position of *VRN-A1* gene on 5A chromosome had the most significant effect on flowering time difference. Its effect was later confirmed by analysis of two other verification mapping populations (F_2 Mooni×DH81 and F_2 DH397×DH81). Newly identified allele of *VRN-A1* gene was responsible for up to 70 % of phenotypic variation. The sequence analysis of this allele showed presence of two deletions in promoter region which was previously identified in *Vrn-A1f* allele from *T. timopheevii* [21] and two unique structural changes in the first intron (424 bp insertion of miniature inverted-repeats transposable element and 2,753 deletion with unique flanking sites). This allele was responsible for flowering delay and was designated as *Vrn-A1f-like*.

II. Analysis of expression profile and methylation pattern of lines with three copy of *Ppd-B1a* gene as a possible cause of intragroup heading time variation.

The aim of this study was identification of gene responsible for heading time variation within individuals of F_2 mapping population developed from cross between two bread wheat cultivars Kaerntner Frueher (KF) and Paragon (P). Identification of higher number of copies of *Ppd-B1* gene in KF corresponded with its earliness and photoperiodic insensitivity. The analysis of F_2 mapping population showed/revealed significant variation within groups with the same number of *Ppd-B1* copies. This variation was detectable up to F_7 generation. In this state, the genetic information of individual lines is considered homogenous enough to decrease the effect of genetic background. But still, we were able to identify significant variation within individuals with three copies of *Ppd-B1*. Selected individuals of F_7 generation with flowering time similar either to P or KF (but with three copies of *Ppd-B1*) were further analyzed. The final goal was identification of expression pattern and methylation status of

Ppd-B1a allele of selected individuals. The results showed that expression pattern correlates with differences in the flowering time – thus KF-like varieties had higher *Ppd-B* expression level than P-like one. The methylation status results were inconclusive when we were not able to obtain statistical significant difference between all P- and KF-like lines. This interesting phenomenon needs to be part of further studies.

5 CONCLUSIONS

Plants are sessile organisms and they are not able to escape when they are threatened by unsuitable conditions. They must evolve mechanism which enables them to grow, flower and develop seeds under the optimal conditions to ensure creation of new generation. The question how they evaluate suitability of external condition is indeed intriguing. For timing of every single life event to the optimal conditions, they developed mechanism which use the most stable pattern – the earth rotation (responsible for day/night cycle) and the tilt of an earth axis (responsible for season alternation and changes in day length during seasons).

By monitoring changes in day length and temperature conditions they are able to determine their position and to time their life cycle accordingly. Not surprisingly, the gene network responsible for perceiving and evaluation of these inputs is complex and it consists of many genes as well as of many interactions between them and proteins. These pathways can be revealed by identification of genes responsible for tuning of flowering time to fit the certain environment and by determination of interaction between these genes.

The identification of novel genes/alleles in wheat was the main aim of the thesis. We succeeded in identification of new allelic variation of *VRN-A1* gene designated *Vrn-A1f-like* allele. This also revealed that different alleles conferring spring habit in wheat affects flowering time in different way.

In the second study we were able to identify intragroup variation within lines with the same number of copies *Ppd-B1* allele. We tried to correlate this variation to changes in expression pattern and methylation status of *Ppd-B1* gene as the effect of epigenetic changes in flowering time determination.

Identification of new genes/alleles and elucidation of flowering time process is important also for breeding new varieties which will be adapted to certain microenvironmental conditions.

6 REFERENCES

[1] Koornneef M, Meinke D. The development of Arabidopsis as a model plant. Plant J 2010;61:909–21. doi:10.1111/j.1365-313X.2009.04086.x.

[2] Andrés F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 2012;13:627–39. doi:10.1038/nrg3291.

[3] Milec Z, Valárik M, Bartoš J, Šafář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv 2014;32:200–14. doi:10.1016/j.biotechadv.2013.09.008.

[4] Yan L, Loukoianov a, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci U S A 2003;100:6263–8. doi:10.1073/pnas.0937399100.

[5] Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J. Allelic variation at the VRN-1 promoter region in polyploid wheat. Theor Appl Genet 2004;109:1677–86. doi:10.1007/s00122-004-1796-4.

[6] Yan L, Fu D, Li C, Blechl a, Tranquilli G, Bonafede M, et al. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci U S A 2006;103:19581–6. doi:10.1073/pnas.0607142103.

[7] Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 2004;303:1640–4. doi:10.1126/science.1094305.

[8] Fu D, Szucs P, Yan L, Helguera M, Skinner JS, Von Zitzewitz J, et al. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Mol Genet Genomics 2005;273:54–65. doi:10.1007/s00438-004-1095-4.

[9] Beales J, Turner A, Griffiths S, Snape JW, Laurie D a. A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theor Appl Genet 2007;115:721–33. doi:10.1007/s00122-007-0603-4.

[10] Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 2012;7. doi:10.1371/journal.pone.0033234.

[11] Sun H, Guo Z, Gao L, Zhao G, Zhang W, Zhou R, et al. DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (Triticum aestivum). New Phytol 2014;204:682–92. doi:10.1111/nph.12948.

[12] Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, et al. Structural variation in the 5´ upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (Triticum aestivum L.), and their effect on heading time. Mol Breed 2013;31:27–37. doi:10.1007/s11032-012-9765-0.

[13] Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, et al. Natural variation

in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 2012;44:1388–92. doi:10.1038/ng.2447.

[14] Gawroński P, Ariyadasa R, Himmelbach A, Poursarebani N, Kilian B, Stein N, et al. A distorted circadian clock causes early flowering and temperature-dependent variation in spike development in the Eps-3Am mutant of einkorn wheat. Genetics 2014;196:1253–61. doi:10.1534/genetics.113.158444.

[15] Laurie DA, Pratchett N, Bezant JH, Snape JW. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (Hordeum vulgare L.) cross. Genome 1995;38:575–85.

[16] Manly KF, Cudmore RH, Meer JM. Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome 2001;12:930–2. doi:10.1007/s00335-001-1016-3.

[17] Díaz A, Zikhali M, Turner AS, Isaac P, Laurie D a. Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 2012;7. doi:10.1371/journal.pone.0033234.

[18] Ivaničová Z, Jakobson I, Reis D, Šafář J, Milec Z, Abrouk M, et al. Characterization of new allele influencing flowering time in bread wheat introgressed from Triticum militinae. N Biotechnol 2016;33. doi:10.1016/j.nbt.2016.01.008.

[19] Jakobson I, Peusha H, Timofejeva L, Järve K. Adult plant and seedling resistance to powdery mildew in a Triticum aestivum x Triticum militinae hybrid line. Theor Appl Genet 2006;112:760–9. doi:10.1007/s00122-005-0181-2.

[20] Jakobson I, Reis D, Tiidema A, Peusha H, Timofejeva L, Valárik M, et al. Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat-Triticum militinae introgression line. Theor Appl Genet 2012;125:609–23. doi:10.1007/s00122-012-1856-0.

[21] Golovnina K a, Kondratenko EY, Blinov AG, Goncharov NP. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. BMC Plant Biol 2010;10:168. doi:10.1186/1471-2229-10-168.

7 LIST OF AUTHOR´S PUBLICATIONS

7.1 Original papers

Ivaničová Z., Jakobson I., Reis D., Šafář J., Milec Z., Abrouk M., Doležel J., Järve K., Valárik M. (2016): Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*. New Biotechnology 33 (5B), 718-727. doi:10.1016/j.nbt.2016.01.008.

Ivaničová Z., Valárik M., Pánková K., Trávníčková M., Doležel J., Šafář J. and Milec, Z. Heritable heading time variation in wheat lines with the same number of Ppd-B1 gene copies- PLoS ONE, PLoS ONE; E 12(8): e0183745, 2017, doi: 10.1371/journal.pone. 0183745.

Ivaničová Z. (2016): Kvitnúť či nekvitnúť. Svet prírody, 4, 19-21, 2016 (in Slovak).

Civáň P., **Ivaničová Z.** and Brown T.A. (2013): Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent. PLoS ONE 8(11), e81955. doi:10.1371/journal.pone.0081955.

Michalcová, V., Dušinský, R., Sabo, M., Al Beyroutiová, M., Hauptvogel, P., **Ivaničová Z.** and Švec, M. (2014). Taxonomical classification and origin of Kamut® wheat. Plant Syst Evol. 300 (7), 1749-1757, doi: 10.1007/s00606-014-1001-4.

7.2 Published abstracts – poster presentations

Ivaničová Z., Milec Z., Trávníčková M., Valárik M., Prášil I.T., Pánková K., Snape J.W., Šafář J.: Different daylenght – different effect. In: Abstract of the International conference "Olomouc Biotech 2013 Plant Biotechnology: Green for Good II". Olomouc, 2013.

Ivaničová Z., Šafář J., Trávníčková M., Pánková K., Bartoš J., Milec Z. and Doležel J. Chasing a new flowering time gene. In: Book of abstracts EUCARPIA Cereal SectionITMI Conference in Weringerode, Lohwasser U. and Börner, A. (eds.), 156, 2014.

Ivaničová Z., Milec Z., Trávníčková M., Abrouk M., Valárik M., Prášil I., Pánková K., Snape J.W., Šafář J. High-throughput gene resources and phenotyping - the key factors in fine mapping of new *Eps* gene. 6. Metodické dny. Bulletin České společnosti

experimentální biologie rostlin a Fyziologické sekce Slovenské botanické společnsti. (2014). No 2. ISSN 1213-6670.

Ivaničová Z., Jakobson I., Reis D., , Milec Z., Abrouk M., Doležel J., Järve K., Šafář J. and Valárik M. Identification and characterization of a new *Vrn-A1f-like* allele responsible for flowering lateness in wheat. IWGS 2017.

7.3 Published abstracts – oral presentations

Ivaničová Z., Jakobson I, Šafář J, Milec Z, Järve K, Valárik M, Doležel J. Characterization of new alleles influencing flowering time introgressed to bread wheat from *T. militinae*. In: Abstracts of the "Olomouc Biotech 2015. Plant Biotechnology: Green for Good III". Olomouc, Czech Republic, 2015.

8 SUMMARY (in Slovak)

Identifikácia a mapovanie génov kvitnutia u pšenice letnej (*Triticum aestivum* L.)

Kvitnutie je pre rastlinu esenciálny proces. Vedie k tvorbe semien a teda zabezpečuje jej prežitie do ďalšej generácie. Optimálny čas kvitnutia umožňuje, aby sa rastlina vyhla nepriaznivým podmienkam prostredia. Zároveň však vedie k využitiu dostatočného množstva živín, ktoré sa často vyskytujú len v časovo limitovanom období. Vhodné podmienky prostredia sú preto dôležité nielen pre tvorbu kvetných častí, ale aj vývin semien a vplývajú na ich množstvo a taktiež ich kvalitu. Práve tieto dve charakteristiky ovplyvňujú výnos a následné spracovanie semien. Kvitnutie vo vhodnom čase je teda dôležité aj z pohľadu človeka.

Ako však rastlina dokáže identifikovať čas a miesto v ktorom sa nachádza, a teda určiť najvhodnejší čas na iniciáciu tohto procesu? Umožňuje jej to zmena teplôt a striedanie dňa a noci. Rastlina totiž dokáže detegovať pravidelné sa striedanie chladného a teplého obdobia a zmeny dĺžky dňa s tým súvisiace umožnené rotáciou Zeme a naklonením zemskej osi.

Okrem vyhodnotenia externých podmienok, musí rastlina vyhodnotiť aj jej endogénne procesy, a teda jej celkovú pripravenosť na proces reprodukcie. Biochemické dráhy, ktoré umožňujú získanie týchto informácií, ich spracovanie a následné kvitnutie sú veľmi zložité a komplexné. Sú kódované veľkým množstvom génov, a práve identifikácia týchto génov a ich interakcií môže prispieť k poznaniu tohto procesu. Ďalším krokom je identifikácia nových alelických variácií génov ovplyvňujúcich kvitnutie. Zdrojom nových alel sú variety, ktoré neboli poľnohospodársky využívané, a preto si zachovali veľkú časť pôvodnej genetickej variability.

Identifikácia nových génov/alel ovplyvňujúcich čas kvitnutia bola aj predmetom našej štúdie. V našej štúdii sme identifikovali novú alelu *VRN-A1* génu, ktorú sme nazvali *Vrn-A1f-like*. *T. militinae*, ktorá je pôvodným zdrojom tejto alely je taktiež druh, ktorý nie je využívaný moderným poľnohospodárstvom. Táto alela bola zodpovedná až za 70 % identifikovanej variability vo fenotype a viedla k oneskoreniu iniciácie kvitnutia. Týmto spôsobom sme

identifikovali rozdiel medzi dvoma alelami (*Vrn-A1a* a *Vrn-A1f-like*) v ovplyvnení procesu kvitnutia.

Ďalší projekt sa zaoberal vysvetlením signifikantnej variability v čase kvitnutia v líniách s troma kópiami génu *Ppd-B1a*. Táto variabilita bola identifikovaná v mapovacej populácii F_2 generácie a neskôr bola zistená aj v ďalších generáciách. Línie z generácie F_7, ktoré reprezentovali obe hraničné varianty, a teda skoro a neskoro kvitnúce línie boli analyzované a bola u nich determinovaná miera expresie a metylácie génu *Ppd-B1a*.

Pochopenie determinácie kvitnutia a identifikácia nových génov/alel, ku ktorej prispela aj naša práca sú dôležité pre šľachtenie nových kultivarov, ktoré budú lepšie adaptované práve na určité podmienky prostredia.