Plant Mol Biol (2015) 88:149–164
DOI 10.1007/s11103-015-0313-2

Discovery, evaluation and distribution of haplotypes and new alleles of the *Photoperiod-A1* gene in wheat

Alexandr Muterko[1] · Ruslan Kalendar[2] · James Cockram[3] · Irina Balashova[1]

Received: 18 October 2014 / Accepted: 29 March 2015 / Published online: 8 April 2015

Abstract Photoperiod response in wheat is determined to a large extent by the homoeologous series of *Photoperiod 1* (*Ppd1*) genes. In this study, *Ppd-A1* genomic sequences from the 5′ UTR and promoter region were analysed in 104 accessions of six tetraploid wheat species (*Triticum dicoccoides*, *T. dicoccum*, *T. turgidum*, *T. polonicum*, *T. carthlicum*, *T. durum*) and 102 accessions of six hexaploid wheat species (*T. aestivum*, *T. compactum*, *T. sphaerococcum*, *T. spelta*, *T. macha*, *T. vavilovii*). This data was supplemented with in silico analysis of publicly available sequences from 46 to 193 accessions of diploid and tetraploid wheat, respectively. Analysis of a region of the *Ppd-A1* promoter identified thirteen haplotypes, which were divided in two haplogroups. Distribution of the *Ppd-A1* haplogroups and haplotypes in wheat species, and their geographical distributions were analysed. Polymerase chain reaction combined with a heteroduplex mobility assay was subsequently used to efficiently discriminate between *Ppd-A1* alleles, allowing identification of the *Ppd-A1b* haplotypes and haplogroups. The causes of anomalous migration of *Ppd-A1* heteroduplexes in gels were found to be the localization of mismatches relative to the center of fragment, the cumulative effect of neighbouring polymorphic sites, and the location of mismatches within A/T-tracts. Analysis of the *Ppd-A1* 5′ UTR in hexaploid wheat revealed a novel mutation within the "photoperiod critical" region in a subset of *T. compactum* accessions. This putative photoperiod insensitive allele (designated *Ppd-A1a.4*) includes a 684 bp deletion which spans region in common with deletions previously identified in other photoperiod insensitive *Ppd1* alleles.

Keywords Cereal · Flowering time · Genetic variation · Heteroduplex mobility assay · Molecular evolution · Photoperiod response · *Ppd-A1* haplotype · Wheat

Introduction

In polyploid wheat (*Triticum* spp.), genetic control of sensitivity to long day (LD) photoperiods (~16 h daylengths) is mainly determined by homoeologous *Photoperiod 1* (*Ppd1*) genes (reviewed by Bentley et al. 2013). Recessive wild-type *ppd1* alleles confer sensitivity to day length, with flowering delayed under short days (SDs, <10 h light), while flowering is promoted under LDs. Dominant *Ppd1* mutants confer photoperiod insensitivity, and result in rapid flowering under both SD and LD photoperiods. Wild type *ppd1* alleles are circadianly expressed, with very low transcript levels at dawn, followed by a peak 3–6 h after dawn and a subsequent drop to very low levels in the dark. In contrast, dominant *Ppd1* alleles are characterized by high expression levels throughout the 24 h period, particularly at dawn and during the dark period (Beales et al. 2007;

Electronic supplementary material The online version of this article (doi:10.1007/s11103-015-0313-2) contains supplementary material, which is available to authorized users.

✉ Alexandr Muterko
muterko@gmail.com

[1] Department of Genomics and Biotechnology, Plant Breeding and Genetics Institute – National Center of Seed and Cultivar Investigation, Ovidiopolskaya Road 3, Odessa 65036, Ukraine

[2] Laboratory of Plant Genomics and Bioinformatics, RSE "National Center for Biotechnology", Sh. Valikhanov 13/1, Astana, Kazakhstan 010000

[3] National Institute of Agricultural Botany (NIAB), Huntington Road, Cambridge CB3 0LE, UK

Wilhelm et al. 2009; Shaw et al. 2012). *Ppd1* genes are important regulators of flowering in cereals, and the use of photoperiod insensitive (PI) alleles has become widespread in wheat varieties following the "green revolution" (Beales et al. 2007; Turner et al. 2005; Wilhelm et al. 2009; Shaw et al. 2012; Nishida et al. 2013). Such early flowering PI varieties represent an important regional adaptation to avoid the effects of high summer temperatures in wheat (Law and Worland 1997; Kato and Yokoyama 1992; Worland et al. 1998) and other cereal species such as barley (Jones et al. 2008, 2011).

The wheat homoeologous *Ppd1* genes are located on the short arms of the group 2 chromosomes (Scarth and Law 1984; Beales et al. 2007; Wilhelm et al. 2009) and belong to the pseudo-response regulator (*PRR*) family (Beales et al. 2007; Turner et al. 2005; Cockram et al. 2012). PRR proteins are characterized by a pseudo-receiver (REC) domain near the N-terminus and a CCT [*CONSTANS* (*CO*), *CO*-like, *TIMING OF CAB EXPRESSION 1* (*TOC1*)] domain towards the C-terminus (Mizuno and Nakamichi 2005). Wheat *Ppd1* genes are homologous to a family of genes that operate in the Arabidopsis circadian clock (*PRR1*, *PRR3*, *PRR5*, *PRR7*, *PRR9*; Beales et al. 2007; Turner et al. 2005). While *Ppd1* shows highest homology to Arabidopsis *PRR7*, the effects on flowering appear to be modulated via an alternative route, most likely by up-regulating *FT1* with a feedback effect that represses *CO* expression (Shaw et al. 2012). In Arabidopsis, *PRR7* is involved in the circadian clock and plays an important role in flowering under LDs. Within the circadian clock, *PRR7* down-regulates *CDF1*, a repressor of *CO* (Imaizumi et al. 2005; Nakamichi et al. 2007). However in wheat, overexpression of *Ppd1* does not affect expression of the circadian clock genes *CCA1*, *TOC1*, *GI*, *PRR73,* or for *CDF1* which acts downstream of the circadian clock, indicating *Ppd1* is not important for wheat circadian clock feedback loops (Shaw et al. 2012). A direct correlation between the expression of dominant *Ppd1* alleles, *FT1* and flowering has been shown (Beales et al. 2007; Nishida et al. 2013; Wilhelm et al. 2009; Shaw et al. 2012), including a strong positive correlation between the number of PI alleles present from different genomes and *FT* expression (Shaw et al. 2012). In Arabidopsis, CO protein is stabilized by light, and activation of its downstream targets (in particular, *FT*) occurs under LD photoperiods (Valverde et al. 2004; Suárez-López et al. 2001). However in wheat, *FT1* can be regulated by *Ppd1* independently of *CO* through an alternative pathway, which is not affected by the circadian clock or genes upstream of *CO* (Campoli et al. 2012; Shaw et al. 2012).

In all cases, mutations of wheat *Ppd1* alleles increases basal transcription levels and determine changes in their diurnal expression (Beales et al. 2007; Wilhelm et al. 2009; Diaz et al. 2012; Shaw et al. 2012). These mutations are commonly caused by large deletions or transposon insertion within the promoter region (Beales et al. 2007; Wilhelm et al. 2009; Nishida et al. 2013), or by copy number variants (CNVs; Diaz et al. 2012). Large deletions in mutant *Ppd-A1a.2*, *Ppd-A1a.3* and *Ppd-D1a* alleles upstream of the coding region are correlated with increased basal transcription level of *PRR-2A* and *PRR-2D*, respectively (Beales et al. 2007; Wilhelm et al. 2009; Shaw et al. 2012). This results in the expression peak shifting from the light to the dark period, and the activation of *FT1* expression in the morning. CNVs present at the *Ppd-B1a*, *Ppd-B1c* and *Ppd-B1d* alleles (which possess 4, 3 and 2 *PRR-2B* copies, respectively) cause increased basal *PRR-2B* transcription levels, and up-regulation during the evening with peak expression at dawn (Diaz et al. 2012; Shaw et al. 2012). For *Ppd-D1*, coding region missense mutations or decreased expression have not been found to influence photoperiod response or result in later flowering under LDs (Beales et al. 2007; Guo et al. 2010). Dominant *Ppd1* alleles on the A, B and D genomes are characterized by different degrees of PI and flowering time. The dominant *Ppd-D1a* allele confers the strongest insensitivity to photoperiod, resulting in the earliest flowering. This is followed by *Ppd-A1a* and *Ppd-B1a* (Bentley et al. 2011). Furthermore, the large *Ppd1* promoter region deletions have a greater effect on decreasing photoperiod sensitivity than CNVs.

Defining *Ppd1* allelic variation in hexaploid and tetraploid wheat species is of importance, as they are used by plant breeders as a source of agronomicly valuable traits. Recent studies have identified a range of PI *Ppd-A1* alleles with promoter deletions, including *Ppd-A1a.1* (1085 bp deletion) identified in hexaploid *T. aestivum* (Nishida et al. 2013), as well as *Ppd-A1a.2* (1027 bp deletion) and *Ppd-A1a.3* (1117 bp deletion) in tetraploid *T. durum* (Wilhelm et al. 2009). These all carry promoter deletions that are associated with overexpression of *FT1* in SDs and early flowering. However the influence of these deletions on flowering time is different. Accordingly, investigation and characterisation of new sources of *Ppd-A1* promoter region polymorphism among natural mutants in different wheat species has the potential to provide new alleles for deployment in breeding programs. Previously, molecular studies have investigated *Ppd-A1* sequence data from 5′/3′ UTR, intronic, and coding regions in some diploid and tetraploid wheat species (Takenaka and Kawahara 2012; Takenaka and Kawahara 2013). However, there are no reports investigating *Ppd-A1* in hexaploid wheat species other than *T. aestivum*. In this study, we investigate *Ppd-A1* genetic variation in a large collection of tetraploid and hexaploid wheat species and accessions. We identify haplotypes of *Ppd-A1b* and investigated the use of the simple and inexpensive PCR-based heteroduplex mobility assay (HMA) for detection of haplotypes and haplogroups. We analyse

Table 1 *Ppd-A1* PCR primer details

Primers	Primer sequence (5′-3′)	Annealing temp (°C)	Expected PCR product size (bp)	Allele	References
durum_Ag5del_F2	tgtcacccatgcactctgtt	58	452	*Ppd-A1b*	(Wilhelm et al. 2009)
durum_Ag5del_R2	ctggctccaagaggaaacac		456[a]	*Ppd-A1b*	
Ppd-A1proF	gtgtcgcacggattttgctc	61	405	*Ppd-A1a.1*	This study
durum_Ag5del_R2	ctggctccaagaggaaacac		463	*Ppd-A1a.2*	(Wilhelm et al. 2009)
			372	*Ppd-A1a.3*	
			806	*Ppd-A1a.4*	
Ppd-A1proF1	ttaccaacaaactaaggaaac	58	507	*Ppd-A1a.1*	This study
durum_Ag5del_R2	ctggctccaagaggaaacac		565	*Ppd-A1a.2*	(Wilhelm et al. 2009)
			474	*Ppd-A1a.3*	
			908	*Ppd-A1a.4*	
Ppd-A1proF	gtgtcgcacggattttgctc	61	669	*Ppd-A1a.1*	This study
TaPpd-B1int1R1	ccgagccagtgcaaattaac[b]		727	*Ppd-A1a.2*	(Nishida et al. 2013)
			636	*Ppd-A1a.3*	
			1070	*Ppd-A1a.4*	

[a] for *T. monococcum* (This study)

[b] in 8-site "A" can be replaced on "T"

the likely causes of anomalous migration of DNA heteroduplexes in PCR–HMA, and formulate hypothesis which explain the observed results. We develop a novel algorithm to determine phylogenetic relationships between haplotypes. As will be shown, this method has very good performance in practice and was integrated as a part of efficient approach for investigations of *Triticum* evolution. Lastly, a novel putative PI allele, *Ppd-A1a.4*, was identified in the hexaploid wheat species *T. compactum*, as well as putative novel mutant alleles in *T. spelta*.

Materials and methods

Plant material

A total of 206 accessions representing six hexaploid and six tetraploid wheat species from 57 countries and different eco-geographic areas were investigated (Supplementary Table 1). Hexaploid species (genome BBA^uA^uDD) include domesticated hulled wheat (spelt wheat: *T. spelta* L., *T. macha* Dekap, *T. vavilovii* Jakubz) and free-threshing wheat [*T. aestivum* L., bread wheat, *T. compactum* Host (club wheat) and *T. sphaerococcum* Percival (Indian dwarf wheat)]. Tetraploid species (genome BBA^uA^u) include domesticated free-threshing wheat [*T. durum* Desf., *T. turgidum* L., *T. polonicum* L. (Polish wheat) and *T. carthlicum* Nevski (Persian wheat)], domesticated hulled wheat (*T. dicoccum* Schrank, emmer) and wild hulled emmer wheat (*T. dicoccoides* Körn). Germplasm was obtained from the National Plant Germplasm System (NPGS, USA) and National Center of Plant Genetic Resources (Ukraine).

DNA extraction, PCR amplification and heteroduplex mobility assay

Total DNA from 4-day-old wheat seedlings was extracted following a modified CTAB-method (Doyle and Doyle 1987). PCR reactions consisted of: DNA (~40 ng), 20 mM Tris–HCl (pH 8.8), 10 mM $(NH_2)_2SO_4$, 2.4 mM $MgCl_2$, 1 mM KCl, 0.1 % Triton X-100, 250 µM dNTPs, 1.5 % DMSO, 0.5 µM each primer and 1.2 U *Taq*-polymerase. The promoter region (5′ UTR) of *Ppd-A1b* was amplified using previously published primers: durum_Ag5del_F2/durum_Ag5del_R2 (Wilhelm et al. 2009). Additionally, primers Ppd-A1proF, Ppd-A1proF1 (this study) and TaPpd-B1int1R1 (Nishida et al. 2013) were used to further study dominant *Ppd-A1a* alleles (Table 1).

PCR was performed using the following program: denaturation at 94 °C (2 min); 30 cycles of amplification: 94 °C (10 s), annealing (15 s), 74 °C (50 s) per cycle, 3 cycles: annealing (15 s), 74 °C (50 s) and a final elongation step of 72 °C for 3 min. Further details of all primers, including annealing temperatures, are listed in Table 1. Amplification products were separated on 6.5 % nondenaturing polyacrylamide (PAA) gels (mono/bis acrylamide ratio 29:1) in TBE buffer, at temperatures <32 °C, and run at 5 V/cm until the bands had migrated ~80 % of the length of the gel. Visualization of amplification products was conducted using a modified silver staining protocol (Budowle et al. 1991). Agarose gel electrophoresis was performed using 1.5 % agarose.

HMA protocols are previously described (Delwart et al. 1993). Here, heteroduplexes were formed by mixing 1 µl of each PCR product and 0.2 µl of 10× annealing buffer (1 M

NaCl, 20 mM EDTA). Mineral oil was overlaid on top of the reaction mixture to prevent evaporation. Samples were centrifuged at 14,000 rpm for 1 min, denatured at 97 °C for 3 min and rapidly cooled to 10 °C for 5 min. Heteroduplexes were resolved on a 6.5 % polyacrylamide gel using a TBE buffer, and run at 5 V/cm.

Cloning and sequencing of PCR fragments

PCR amplicons were separated on 1.5 % agarose gels, gel-extracted using a GeneJET Gel Extraction Kit (Thermo Scientific), and DNA fragments ligated into pGEM-T (Promega), following the manufacturer's instructions. Plasmid DNA was transformed into JM109 *Escherichia coli* competent cells (L2001, Promega). Transformed cells containing the plasmid carrying an insertion of foreign DNA fragment were detected using white-blue selection on growth medium containing ampicillin, X-Gal, and IPTG. Positive colonies were tested for the presence of cloned PCR products by PCR with universal pUC primers (forward and reverse M13 primers), followed by separation and visualization of PCR products on ethidium bromide stained agarose gels. Plasmid DNA was extracted using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), and sequencing PCRs performed using BigDye Terminator Kit v3 (Applied Biosystems), following the manufacturer's instructions. Fluorescently labeled extension products were precipitated and resuspended in HiDi (Applied Biosystems) and detected using an ABI3700 PRISM DNA Analyser (Applied Biosystems).

Sequence data from this article can be found in the GenBank data libraries under the following accession numbers: Hap-4 (KF758437), Hap-1 (KF834261, KF834265, KF834266), Hap-3 (KF834262), Hap-12 (KF834264), Hap13 (KF834263), *Ppd-A1a.2* (KJ767781), *Ppd-A1a.4* (KJ767779, KJ767780).

Reconstruction of the DNA path

The calculation of reconstructed DNA trajectory was performed based on average values of the local helical parameters deduced from the published B-DNA solution structures (PDB ID: 1RVI, 1RVH, 1FZX, 1G14, 1AXP, 2L8Q, 2M2C, 2MCI, 1DK9, 2K0V, 1NEV, 1LAI) by using Curves+ (Lavery et al. 2009). Only the central base pairs were used that largely eliminates errors due to end effects. Parameters of dinucleotide steps were applied in tetranucleotide context, where possible. In all other cases, averaged dinucleotide parameters deduced from all DNA structures were applied. The DNA curvature of ith position was evaluated as a value which is the inverse of the radius of a circumscribed circle of a triangle with vertices on helix axis coordinates at i-window, i and i+ window, where window size was 10 bp. Curvature was measured in DNA curvature units where one curvature unit is defined as the average curvature of DNA in the nucleosome core particle, 1/42.8 Å. Local bend angles at each ith positions were measured as the angles between the tangent vectors in the direction along the contour DNA from i-window to i and from i to i+ window, where window size was 2 bp. Structures were visualized using UCSF Chimera (Pettersen et al. 2004).

Phylogenetic analysis

To investigate phylogenetic relationships between *Ppd-A1b* haplotypes and to define haplogroups, a binary adjacency matrix (which reflects the dependence between alleles) was determined based on the principal according to which a given allele is associated with another allele if during haplotype sampling it's presented only in combination with the second allele [this approach is in accordance with the principles of character-based evolutionary reconstruction (Gusfield 1997)]. The more detailed description of this algorithm including the calculation of interallelic associations and phylogenetic relationships between haplotypes, the reconstruction of the hierarchical clustering of interallelic associations, definition of putative root, haplotypes of primary divergence and haplogroups can be found in the Supplementary Materials. Phylogenetic analysis was also conducted using the Median-Joining algorithm (MJ) (Bandelt et al. 1999), optimized using the maximum parsimony (MP) method (Polzin and Daneschmand 2003). Undirected graphs were visualized using Dendroscope 2.5 (Huson et al. 2007). Phylogenetic network reconstruction was performed in Network 4.6.1.1. (http://fluxus-engineering.com). Multiple sequence alignments were generated using Clustal W (Thompson et al. 1994).

Results

Ppd-A1 sequence polymorphism

Primer pair durum_Ag5del_F2/durum_Ag5del_R2 was used to amplify a region of the *Ppd-A1* promoter (−19 to −470 bp) in 102 accessions of six hexaploid wheat species and 104 accessions of six tetraploid species. Separation of PCR amplicons by PAA gel electrophoresis revealed differences in the migration rate for the 452 bp fragments (Fig. 1a) that was not observed using standard agarose gel separation: slow migrating amplicons (designated 452s) were detected in 66 % of accessions, whereas fast migrating amplicons (452f) were observed in the remaining 34 %. Four 452f amplicons and one 452s amplicon from hexaploid and tetraploid wheat were cloned and sequenced (GenBank accessions: KF758437, KF834261, KF834262,

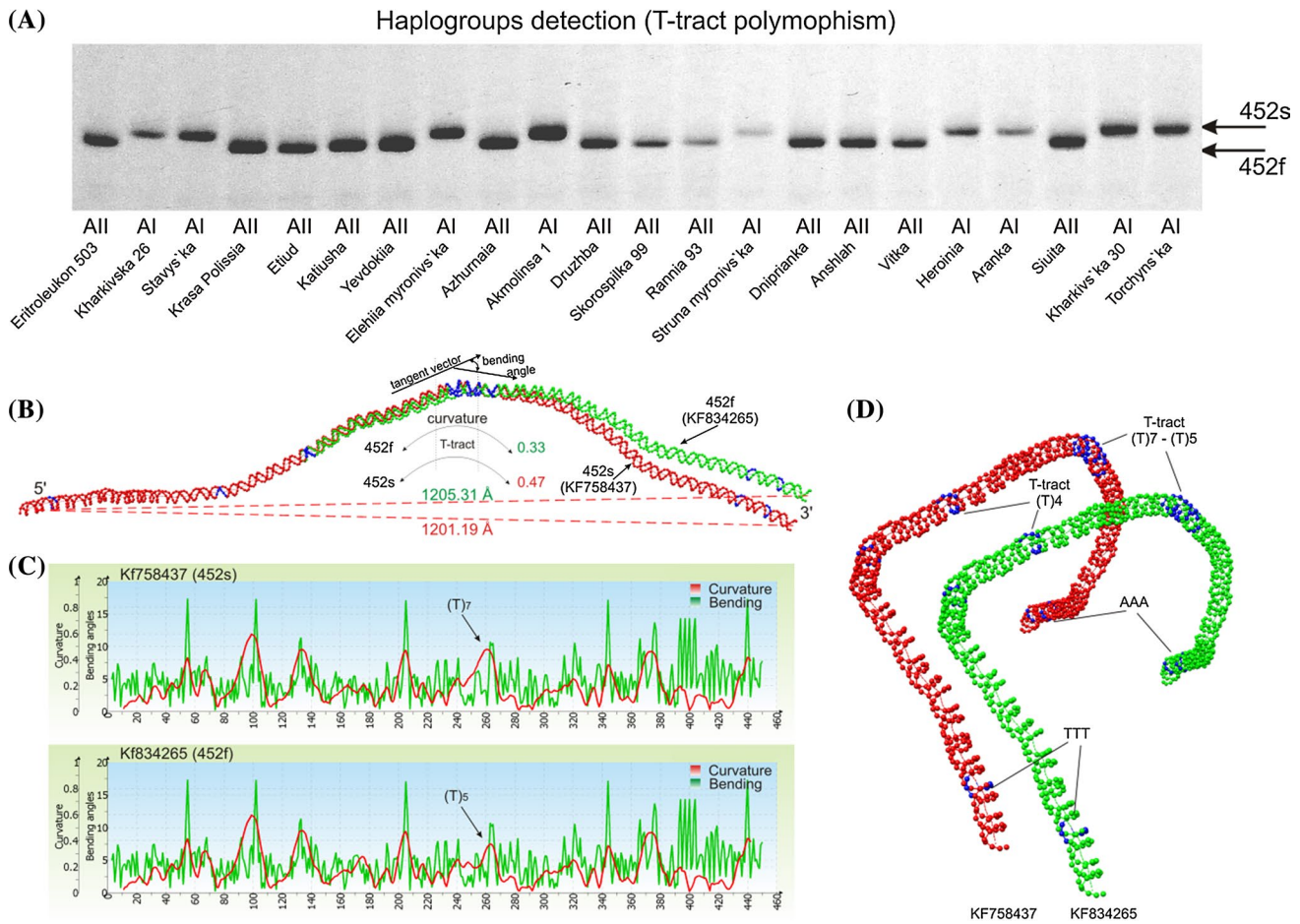

Fig. 1 Detection and analysis of *Ppd-A1b* polymorphism **a** Polymorphism of *Ppd-A1b* promoter region and detection of haplogroup in *T. aestivum* varieties. Selected soft wheat cultivars are shown, highlighting the diversity of *Ppd-A1b* haplogroups among genetic materials, commonly used in wheat breeding. Differential mobility is explained by the presence of T-tract polymorphism and indicates the AI (slow-migrating 452 bp fragments, 452s) and AII (fast-migrating 452 bp fragments, 452f) haplogroups. **b** Supperposition of DNA paths of the polymorphic fragments reconstructed according to dinucleotide parameters deduced from published NMR solution structure of B-DNA. DNA molecule straightening due to a break in the T-tract is accompanied by increasing end-to-end distance. The curvature at position 260 bp is indicated for two fragments (measured in curvature units). **c** Distribution of curvature and bend angles. The curvature reduction at position 260 bp, at the center of T_7-tract, is indicated. **d** Localization of A/T-tracts in PCR-fragments

KF834265, KF834266). DNA sequence alignment did not identify insertion/deletion (InDel) mutations between the two polymorphic bands which might explain their differential migration in PAA gels. However, two SNPs were identified. It is known that DNA molecules containing adenine- or thymine-tracts migrate anomalously slowly in PAA gels (Marini et al. 1982; Koo et al. 1986; Diekmann 1989), but not in agarose gels, most likely because the agarose gel fibers are oriented by the electric field creating transient pores in the gel matrix (Stellwagen 2009). To investigate the possible effect of the SNPs identified in the amplicons on PAA electrophoretic mobility, global 3D structures of the DNA molecules were calculated (Fig. 1b, d). Evaluation of DNA curvature and local bending angle distributions found all of the 452f PCR fragments to be characterized by a predicted reduction in curvature, due to a T → C SNP within a thymine-tract [T-tract = $(T)_7$], located close to the center of the amplicon (Fig. 1c). This SNP breaks the T-tact, resulting in a shortened stretch of five thymines [T-tract = $(T)_5$], a straightening of the DNA molecule, and thus increased migration in PAA gels.

To further investigate this region of *Ppd-A1*, in silico PCR using accessions identified in GenBank, EMBL and DDBJ databases was conducted. Analysis of in silico amplicons from seven tetraploid wheat species with genome BBA^uA^u, *T. timopheevii* (Zhuk.) (GGA^uA^u), hexaploid wheat *T. aestivum* (BBA^uA^uDD) and two diploid wheats [*T. urartu* Tumanian ex Gandilyan (A^uA^u); *T. monococcum* L. (A^mA^m)] (Supplementary Materials 2) identified amplicons of 452 bp (230 accessions, including seven accessions sequenced here), 453 bp (nine sequenced accessions) and 456 bp (nine sequenced accessions). The

Table 2 *Ppd-A1b* haplotypes, based on partial promoter sequences of 452 bp amplicons identified in the different wheat species

Haplo-types	Positions within amplicon (bp)												
	21	54	60	130	149	172	180	254	258	315	317	329	341
Hap-1	T	C	T	C	C	C	G	G	C	G	C	C	C
Hap-2	T	C	T	T	C	C	G	G	T	G	C	C	C
Hap-3	T	C	T	C	C	C	G	C	C	G	T	C	C
Hap-4	T	C	T	C	C	G	G	G	T	G	C	C	C
Hap-5	T	C	T	C	C	C	G	G	T	C	C	C	C
Hap-6	T	C	T	C	C	C	G	G	C	G	T	C	C
Hap-7	T	C	T	C	C	C	G	G	T	G	C	C	C
Hap-8	T	C	C	C	C	C	G	G	T	G	C	C	C
Hap-9	T	C	C	C	T	C	G	G	T	G	C	C	C
Hap-10	T	T	C	C	C	C	G	G	T	G	C	C	C
Hap-11	C	C	T	T	C	C	G	G	T	G	C	C	C
Hap-12	T	C	T	C	C	C	C	G	C	G	C	C	C
Hap-13	T	C	T	C	C	C	G	G	C	G	C	A	T

456 bp amplicon was detected only in *T. monococcum*, while 453 bp fragments were present pred ominantly in *T. timopheevii*, as well as in one *T. dicoccoides* accession. Multiple sequence alignment of the 452 bp promoter fragment for 240 accessions (including sequences identified in this study; Supplementary Materials 2) identified 13 SNPs, forming 13 distinct combinations (Table 2). Here, these combinations are considered as individual haplotypes, and each SNP as an individual allele.

Phylogeny of haplotypes and haplogroups discovery

To investigate phylogenetic relationships between *Ppd-A1b* haplotypes and to define haplogroups, three approaches were taken. (1) A binary adjacent matrix was formed based on the association between alleles, and visualized as a connected acyclic graph (tree) with root (Fig. 2a). (2) A haplotype-based tree (Fig. 2b), allowing haplogroups to be determined based on the presence of common mutations. (3) A non-rooted quasi-median MJ network, optimized by MP (Fig. 2c). The haplotype-based tree inferred from the allele-based tree, as emergence of a new mutation is accompanied by the formation of new combinations of mutations, and thus, new haplotypes. Defined in this way, phylogenetic relationships between haplotypes are consistent with the haplotype network calculated by MJ method (Fig. 2b, c). However, as the tree is un-rooted, it reflects only the phylogenetic relationships between haplotypes and their geneal-ogical structure.

According to phylogenetic analyses based on the binary adjacent matrix approach, Hap-7 and Hap-1 represent haplotypes of primary divergence (ancestral haplotypes) from which all haplotypes of the AI and AII haplogroups derive, respectively (Fig. 2b). Hap-7 and Hap-1 differ by one SNP at position 258 bp (258T/258C). It is assumed that Hap-7 is older than Hap-1 and represents the ancestral haplotype (root) as (1) Hap-7 has more derivative haplotypes than Hap-1, and (2) the haplotypes of AI haplogroup are found in more wheat species. Moreover, all einkorn wheat haplotypes belong to haplogroup AI, which also predominates in tetraploid species. Further indication that Hap-7 represents a putative ancestral haplotype was determined using the MJ method, where it forms the basal haplotype within the haplotype network (Fig. 2c). It is also likely that Hap-7 is ancestral for polyploid wheat, as Hap-8, Hap-9 and Hap-10 (found in *T. urartu*) carry a specific allele (60C) and form a distinct cluster (Fig. 2b, c).

Haplotype detection using the heteroduplex mobility assay

In order to provide a rapid and simple method to differentiate between *Ppd-A1* haplotypes, an HMA approach was employed. HMA is based on the principle that DNA heteroduplexes formed between related sequences have reduced mobility in polyacrilamide gels proportional to their degree of divergence. Although the AI haplogroup contains more haplotypes than AII, almost all AI haplotypes were predominantly characterized by species-specific distributions. *In silico* PCR showed Hap-8, Hap-9 and Hap-10 are found only in *T. urartu*, while Hap-11 was found only in *T. timopheevii*. Hap-5 and Hap-7 were present at low frequency (1.12 % each), and were detected only in *T. dicoccoides*. Hap-2 was found in *T. turgidum*, *T. timopheevii* and *T. ispahanicum*. Of these, only accessions from *T. turgidum* and *T. dicoccoides* were investigated experimentally in this study (see "Materials and methods" section). Accordingly, we assume that is not sense to analyze of AI haplotypes for

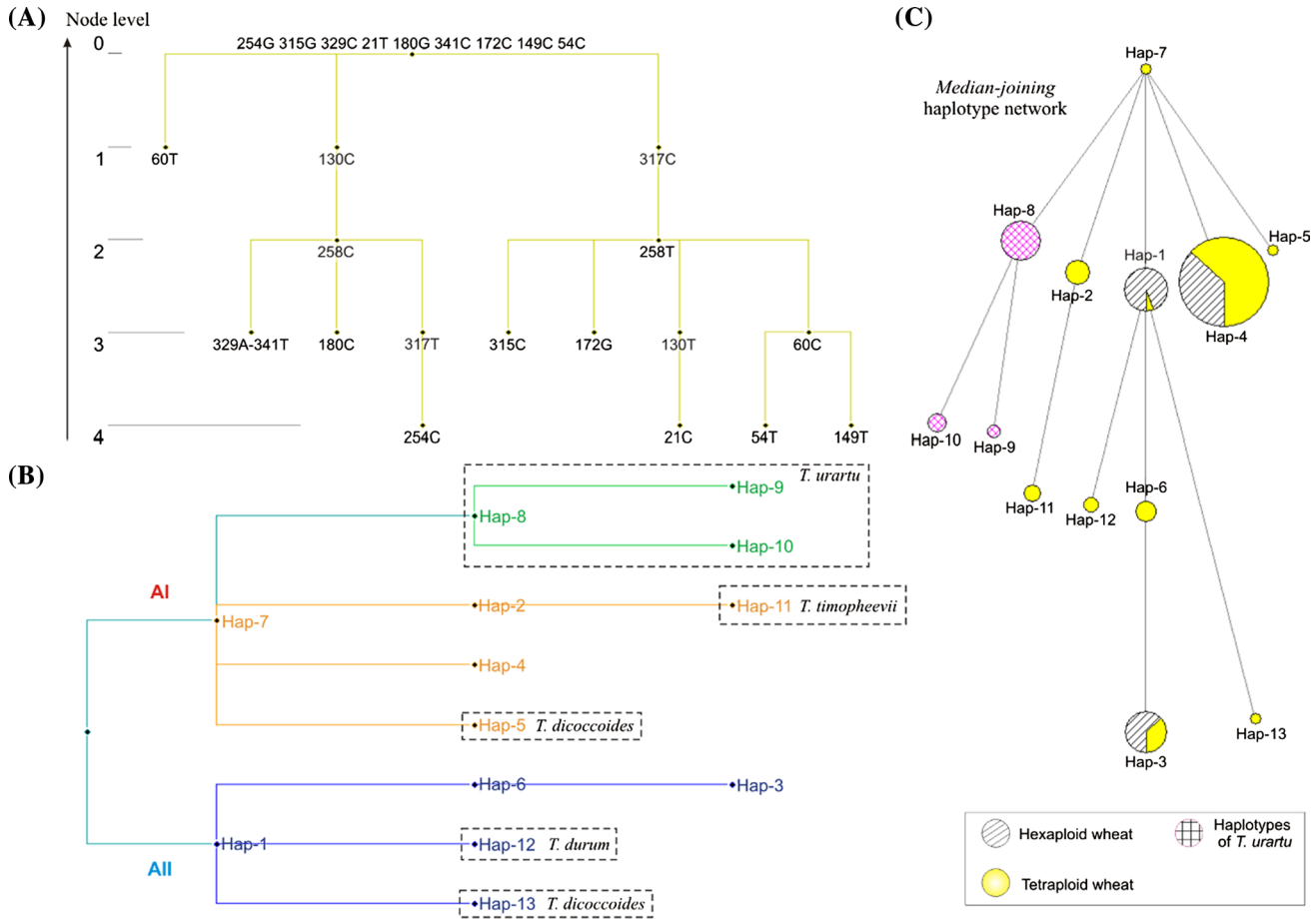

Fig. 2 Interallelic associations and phylogenetic relationships between *Ppd-A1b* haplotypes. **a** Hierarchical clustering of interallelic associations. Node levels are shown. The root was formed through grouping nodes of lowest level of individual connected components. **b** Phylogenetic tree of haplotypes. Species-specific haplotypes are indicated. **c** Median-joining haplotype network. Node sizes are proportional to the number of accessions. The ratio of hexaploid and tetraploid wheat accessions for each haplotypes is shown

hexaploid and for most species of tetraploid wheat. Hence identification of haplotypes was performed for 51 accessions of hexaploid and 15 accessions of tetraploid wheat possessing haplogroup AII (most diverse in terms of haplotypes frequency).

For all HMA analyses, hybridization of amplicons was conducted against three reference sequences: Hap-1 (haplogroup AII, KF834265), Hap-3 (AII, KF834262) and Hap-4 (AI, KF758437; Fig. 3). These haplotypes were used as reference based on their high frequencies within the AI (Hap-4) and AII (Hap-1, Hap-3) haplogroups. Analysis of tetraploid wheat showed heteroduplex mobility polymorphism for one *T. dicoccoides* (PI 466941) and one *T. durum* (PI 94721) accession, identified using heteroduplexes formed by combination with all three reference sequences. PCR fragments amplified from accessions PI 466941 and PI 94721 were cloned and sequenced (GenBank accession KF834263, KF834264), with sequence alignment showing they represent Hap-12 and Hap-13, respectively. Amplicon hybridization of *T. dicoccoides* PI 233288 with the Hap-3 reference resulted in two heteroduplexes, corresponding in migration to Hap-1/Hap-3 and Hap-4/Hap-3 hybrids (Fig. 3). Additionally, one heteroduplex (corresponding in migration to Hap-1/Hap-4) was observed for the following combinations: PI 233288/Hap-1 and PI 233288/Hap-4. This indicates that PI 233288 carries the two haplotypes (from AI and AII haplogroups) simultaneously: Hap-1 (AII) and Hap-4 (AI). Heterogeneousity of PI 233288 was confirmed by autohybridization of PCR products.

Among the heteroduplexes formed by the combinations of accessions from five haplotypes (Hap-1, Hap-3, Hap-4, Hap-12, Hap-13), we observed slowest migration for the Hap-3/Hap-4 heteroduplex (which possessed a four nucleotide mismatch), and fastest migration for Hap-1/Hap-12 and Hap1/Hap13 (one mismatch). Increased migration was observed for heteroduplexes in the order: Hap-3/Hap-13 < Hap-3/Hap-12 < Hap-3/Hap-1 (or Hap-4/Hap-3 < Hap-4/Hap-12 < Hap-4/Hap-1) which contain 4,

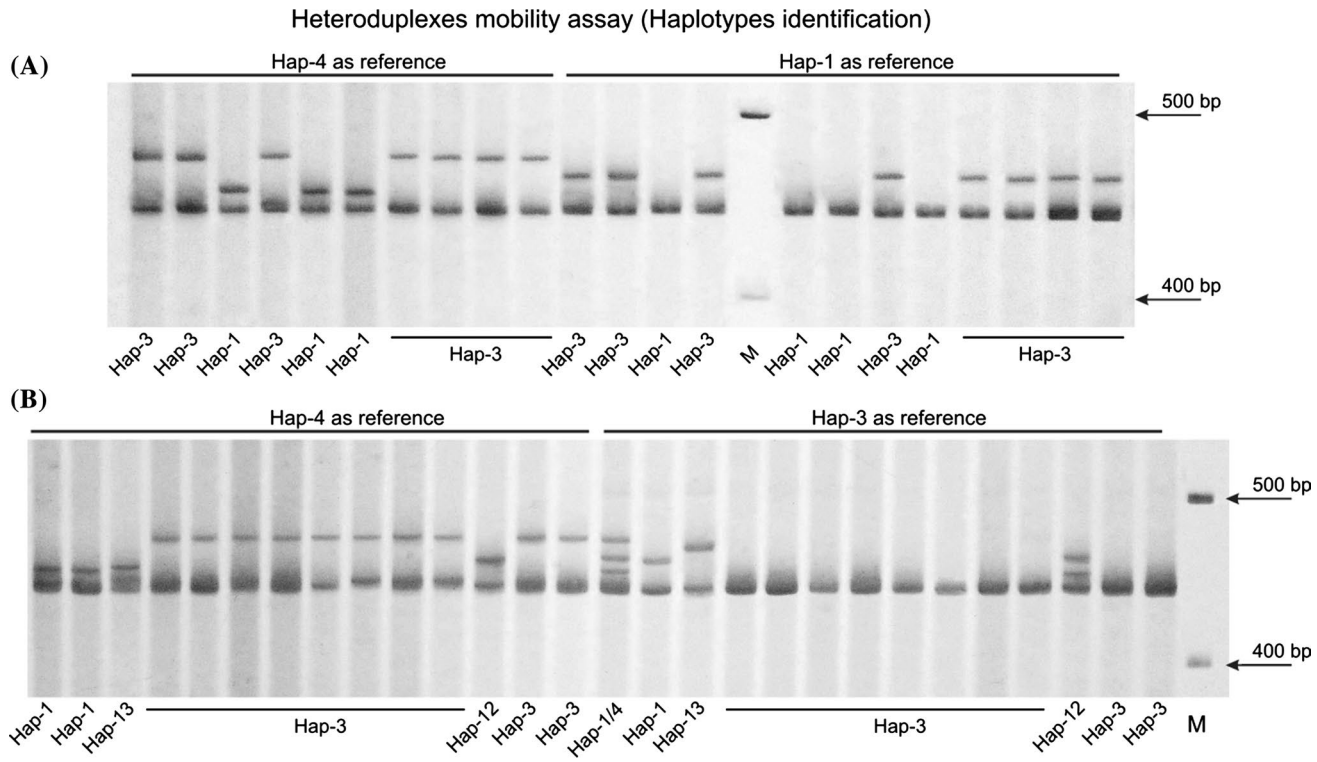

Fig. 3 PCR–HMA test to identification of *Ppd-A1b* haplotypes. Absence of heteroduplexes indicates haplotype analogous to reference. M-DNA molecular size marker. **a** HMA of certain hexaploid wheat accessions with AII haplogroup [PI (168680, 295056, 520066, 159101, 294567, 294892, 352306, 272554, 352466, 355514, 168680, 295056, 520066, 159101, 294567, 294892, 352306, 272554), UA0300245, PI (352466, 355514, 326319, 428343)], where Hap-4 and Hap-1 are used as reference haplotypes for hybridization. **b** Analysis of 14 accessions of tetraploid wheat [PI (233288, 352323, 466941, 190921), UA0300214, CItr (13712, 7688), PI (211705, 220356, 221422, 278596, 94721, 655432, 286070)] using Hap-4 and Hap-3 as reference sequences for formation of heteroduplexes. In the center located heterogeneous accession of *T. dicoccoides* (PI 233288), which provides two heteroduplexes at hybridization with Hap-3, corresponding Hap-4/Hap-3 and Hap1/Hap3

3 and 2 mismatching nucleotides, respectively (Fig. 4). Anomalous migration of heteroduplexes formed between accessions from different haplogroups was also observed. For example, comparison of different hetroduplexes containing equal numbers of nucleotide mismatches found that heteroduplexes between haplogroup AII accessions (Hap-1, Hap-12, Hap-13) and Hap-4 migrate faster than heteroduplexes formed between the same haplotypes with Hap-3 (Fig. 4), indicating the localization of mismatched nucleotides in A/T-tracts increase migration of heteroduplexes due to DNA molecule straightening. However, the difference in migration between heteroduplexes Hap-3/Hap-12 and Hap-4/Hap-12 were not significant. Furthermore, while Hap-3/Hap-4, Hap-4/Hap-13 and Hap-3/Hap-13 all contain the four mismatching nucleotides, Hap-3/Hap-4 migrates slower than Hap-3/Hap-13, which in turn is slower than Hap-4/Hap-13 (Fig. 4). This suggests direct dependence of heteroduplex migration rate upon distance between the polymorphic positions. Similarly, mismatches located closer to the center appear to have a stronger influence on slowing heteroduplexes than distal mismatches. These assumptions explain the anomalously slow migration of Hap-3/Hap4, which contains of more closely located polymorphic positions (4 bp between them) near the center of the fragment, in comparison to other heteroduplexes.

Haplotypes and haplogroups distribution

In the analysed genetic material the AI haplogroup predominates in tetraploid wheat, whereas hexaploid wheat is characterized by the approximately equal distribution of haplogroups AI and AII, although for different species these haplogroups represented in different ratio. Among haplotypes belonging to the AII haplogroup, Hap-1 was identified in 52 % of all accessions whereas Hap-3 was found in 45 %. All *T. sphaerococcum* accessions were characterized by haplotype Hap-1. Four haplotypes (Hap-1, Hap-3, Hap-12 and Hap-13) were found in tetraploid wheat, with Hap-3 predominant (73 %). Only Hap-1 and Hap-3 were present in hexaploid wheat, with Hap-1 predominant (63 %). However, the exclusion of hexaploid species *T. sphaerococcum*, results in a decrease in Hap-1 frequency within

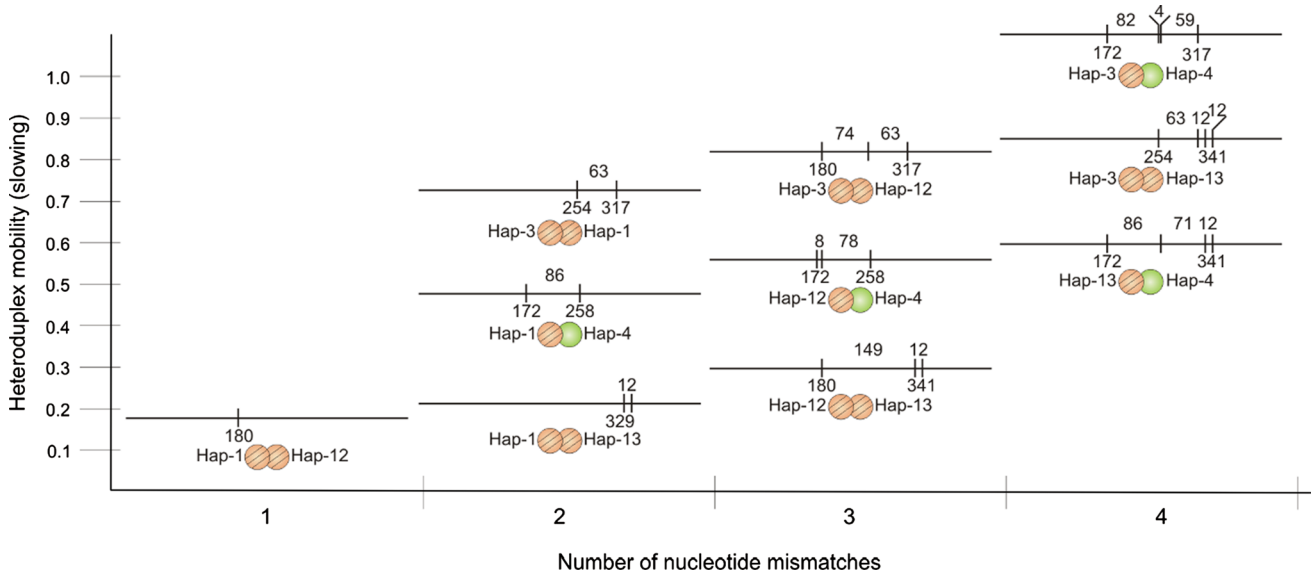

Fig. 4 Heteroduplex mobility. Relative heteroduplex mobilities were calculated as the distance of migration of heteroduplex bands relative to homoduplex bands. Mobility of Hap3/Hap-4 (the slowest) was taken as one. *Open circles* (*green*) indicate the haplotypes of AI haplogroup, *slashed* (*red*) *circles*—haplotypes of AII haplogroup. The distance between nucleotide mismatches and position of mismatches for each heteroduplex are indicated

hexaploid wheat to 42 %. Overall, comparison of hexaploid and tetraploid wheat shows Hap-1 to largely predominate in hexaploid wheat (94 %, or 88 % excluding *T. sphaerococcum*). Hap-12 (*T. durum*) and Hap-13 (*T. dicoccoides*) developed evolutionarily later than Hap-1 and appear to be species-specific.

To further investigate haplotype and haplogroup distributions among wheat species, these results were combined with results of in silico PCR, resulting in the analysis of 426 accessions (Table 3). Within haplogroup AI, phylogenetic analyses indicated the most recent haplotypes were Hap-4, Hap-5, Hap-9, Hap-10, Hap-11. In AII, the most recent were Hap-3, Hap-12 and Hap-13. Hap-4 and Hap-3 were characterized by the widest distribution within haplogroups AI and AII, respectively. Hap-8 is evolutionary older then Hap-9 or Hap-10 and is likely the founder of the *T. urartu* haplotypes analysed here. The highest number of different haplotypes was detected in wild emmer wheat (*T. dicoccoides*, 7 haplotypes). In tetraploid wheat, Hap-4 predominates in *T. durum*, *T. polonicum* and *T. carthlicum*. Integration of haplotype phylogenies with their distribution among different wheat species agreed with the established evolution of *Triticum* species (Fig. 5).

The AI haplogroup was distributed across a wide geographic range (Fig. 6), predominantly in the north and northwest regions: Europe, Russian, Northern Africa, Ethiopia, Near East, Argentina, Canada. In contrast, Haplogroup AII was prevalent in the southeast such as India, Pakistan, Israel, China and largely in USA and partially in Afghanistan, Palestine and Turkey (Supplementary Material 3).

Discovery of the new *Ppd-A1* alleles

Using primers durum_Ag5del_F2/durum_Ag5del_R2, PCR on two *T. compactum* (PI 114638 and PI 164160) and two *T. spelta* (UA0300218, PI 348700) accessions failed to amplify products. This could indicate the presence of large promoter deletions of 1085, 1027 and 1117 bp characteristic of the previously identified *Ppd-A1a.1*, *Ppd-A1a.2* or *Ppd-A1a.3* alleles, respectively. However, *Ppd-A1a.2* and *Ppd-A1a.3* have previously only been identified in tetraploid wheat, and it has been suggested that these alleles have a later origin, predominating in modern *T. durum* wheat varieties (Bentley et al. 2011; Wilhelm et al. 2009). A primer pair for detection of *Ppd-A1a.2* and *Ppd-A1a.3* alleles has previously been published (durum_Ag5del_F1/durum_Ag5del_R2, Wilhelm et al. 2009). However the 1085 bp deletion carried by *Ppd-A1a.1* (identified in hexaploid wheat) includes the annealing site of forward primer durum_Ag5del_F1 (Fig. 7b). To identify the *Ppd-A1a.1* allele, primer pair TaPpd-A1prodelF1/TaPpd-A1prodelR2 has previously been proposed (Nishida et al. 2013). However sequence alignment of the 230 accessions investigated here show that the forward primer annealing site contains several polymorphic positions (SNPs: T6A, C11T, C12T, T20C). Accordingly, here we design a new forward primer Ppd-A1proF (which anneals 47 bp upstream from TaPpd-A1prodelF1) in combination with reverse

Table 3 Distribution of haplotypes in wheat species (summarised data—in this study and in silico PCR accessions)

Species (total accessions)	Haplotypes (numbering) percent
T. urartu (37)	Hap-8 (28) 75.68 %
	Hap-9 (3) 8.11 %
	Hap-10 (6) 16.22 %
T. dicoccoides (43)	Hap-1 (7) 16.28 %
	Hap-5 (2) 4.65 %
	Hap-6 (5) 11.63 %
	Hap-7 (2) 4.65 %
	Hap-3 (7) 16.28 %
	Hap-4 (19) 44.19 %
	Hap-13 (2) 4.65 %
T. dicoccum (12)	Hap-3 (2) 16.67 %
	Hap-4 (10) 83.33 %
T. turgidum (86)	Hap-2 (8) 9.3 %
	Hap-3 (22) 25.58 %
	Hap-4 (52) 60.47 %
	Hap-1 (4) 4.65 %
T. durum (91)	Hap-4 (85) 93.4 %
	Hap-12 (4) 4.4 %
	Hap-3 (2) 2.2 %
T. carthlicum (18)	Hap-3 (1) 5.56 %
	Hap-4 (17) 94.44 %
T. polonicum (22)	Hap-4 (22) 100 %
T. timopheevii (8)	Hap-11 (5) 62.5 %
	Hap-6 (2) 25 %
	Hap-2 (1) 12.5 %
T. ispahanicum (2)	Hap-2 (2) 100 %
T. turanicum (7)	Hap-4 (7) 100 %
T. aestivum (27)	Hap-1 (9) 33.33 %
	Hap-3 (6) 22.22 %
	Hap-4 (12) 44.45 %
T. sphaerococcum (18)	Hap-1 (18) 100 %
T. compactum (20)	Hap-1 (4) 20 %
	Hap-3 (6) 30 %
	Hap-4 (10) 50 %
T. spelta (22)	Hap-1 (3) 13.64 %
	Hap-3 (2) 9.09 %
	Hap-4 (17) 77.27 %
T. macha (10)	Hap-3 (3) 30 %
	Hap-4 (7) 70 %
T. vavilovii (3)	Hap-3 (2) 66.67 %
	Hap-4 (1) 33.33 %

primer durum_Ag5del_R2 (Wilhelm et al. 2009) to detect all dominant *Ppd-A1a* alleles (*Ppd-A1a.1*, *Ppd-A1a.2* and *Ppd-A1a.3*). This primer pair results in amplicons of 405, 463 and 372 bp for *Ppd-A1a.1*, *Ppd-A1a.2* and *Ppd-A1a.3*, respectively. PCR of the two *T. compactum* and two *T. spelta* accession with this primer pair resulted in amplicons of 806 bp in the *T.compactum* accessions only (PI 114638, PI 164160; Fig. 7a). These were cloned and sequenced (GenBank: KJ767779, KJ767780). Sequence alignment showed this novel allele includes a 684 bp deletion in the promoter region (Fig. 7b) of the *Ppd-A1* gene [position −302; from −302 to −986 bp, counting from the start codon of Chinese Spring (GenBank: DQ885753)]. This new putative *Ppd-A1* photoperiod insensitive allele was designated *Ppd-A1a.4*, according to the nomenclature of photoperiod insensitive *Ppd-A1* alleles proposed by Nishida et al. (2013).

Primer Ppd-A1proF1 was designed to include more of the left border of the *Ppd-A1* 5′ UTR (annealing site −1611 bp). As a result, PCR with Ppd-A1proF1/durum_Ag5del_R2 amplified 908 bp and 565 bp fragments from *Ppd-A1a.4* and *Ppd-A1a.2* (GenBank: KJ767781), respectively (Fig. 7a). However, no PCR products were observed from *T. spelta* accessions (UA0300218 and PI 348700). To try to overcome this, reverse primer TaPpd-B1int1R1 (Nishida et al. 2013, previously designed to amplify *Ppd-B1)* was used, as sequence alignment indicated this primer would also anneal to *Ppd-A1* (with a 1 bp mismatch). This reverse primer was used to expand the 3′ *Ppd-A1* border. It anneals at position 226–245 bp, and encompasses exon-1 and 55 bp of the first intron. Accordingly, pair primer Ppd-A1proF/TaPpd-B1int1R1 was optimized for investigation of *Ppd-A1* in *T. spelta* accessions UA0300218 and PI 348700. PCR products of 727, 636 and 1070 bp were obtained for *Ppd-A1a.2*, *Ppd-A1a.3* and *Ppd-A1a.4,* respectively. However, again, no amplicons were detected for the two *T. spelta* accessions (Fig. 7a).

The *T. compactum* (PI 114638) and *T. spelta* (PI 348700) accessions were grown in a glasshouse without vernalization under 16 h (LD) and 9.5 h (SD) photoperiods, and allelic variation at *Ppd1* homoeologous determined using DNA-markers (Beales et al. 2007; Diaz et al. 2012; Nishida et al. 2013; this study). None of the known dominant PI alleles at *Ppd-B1* or *Ppd-D1* were identified. While *T. spelta* accessions grown under SD did not flower by the end of the experiment (110 days), *T. compactum* (PI 114638) flowered 18 days later under SDs compared to LD. These observations confirm the decrease in sensitivity to photoperiod for *T. compactum* accessions with *Ppd-A1a.4*, and preservation of photoperiod sensitivity for *T. spelta* accessions with mutant variant *Ppd-A1*. These results indicate *Ppd-A1.a.4* is photoperiod insensitive, although for future confirmation of this assumption the genetic analysis will be required.

Discussion

In previous studies, Takenaka and Kawahara (2012, 2013) investigated 5823–7076 bp of genomic *Ppd-A1* sequence

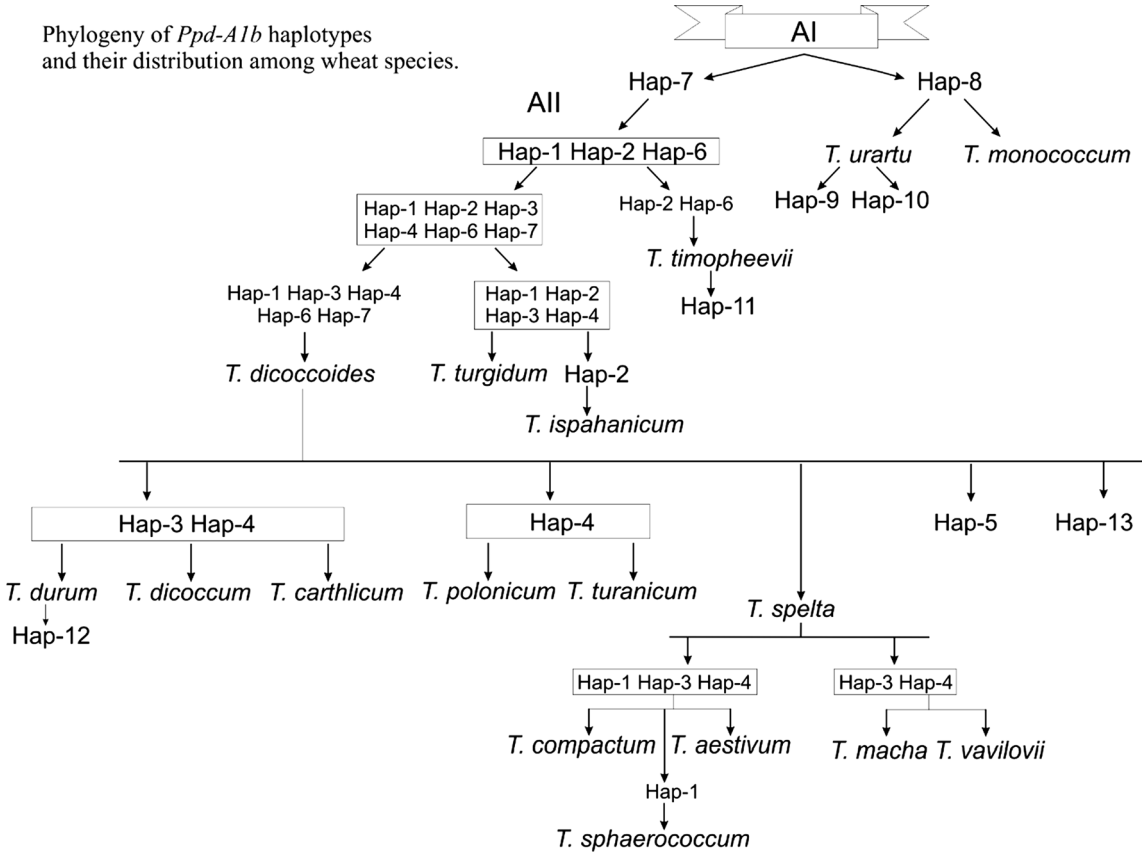

Fig. 5 Integration of *Ppd-A1b* haplotype phylogeny with their distribution among wheat species, for reconstruction of *Triticum* evolution

data in diploid and tetraploid wheat, identifying sixty-seven haplotypes. However, no method for haplotype detection other than DNA sequencing was proposed. In this study, a simple strategy to identify *Ppd-A1* haplotypes is presented based on PCR-HMA. This approach is able to identify 13 polymorphic, codominant DNA-markers for *Ppd-A1b* analysis. The procedure requires no special equipment or preparation of samples. The heteroduplex mobility assay was used to increase the informativeness of PCR-amplified *Ppd-A1b* 5′ UTR. Using this approach, five haplotypes were identified, allowing their distribution in hexaploid and tetraploid wheat to be investigated. A high diversity of haplogroups and AII haplotypes (Hap-1, Hap-3) in bread wheat (*T. aestivum*) varieties was identified here. Previously, *Ppd-A1* has been described as possessing low allelic diversity (Bentley et al. 2011; Beales et al. 2007; Wilhelm et al. 2009; Nishida et al. 2013). The PCR-HMA approach validated here provides polymorphic, codominant DNA-markers of *Ppd-A1b* for use in genetic analyses or marker assisted selection (MAS). Using PCR-HMA, accessions carrying Hap-1 and Hap-3 can be identified using any accession from the AI haplogroup as reference. We also identified haplotypes for 230 GenBank/EMBL/DDBJ accessions, which may also be used as references in PCR-HMA analyses.

It has previously been shown that electrophoretic mobility of heteroduplexes is proportional to the number of nucleotide mismatches when that level exceeds 5 %, and that mismatches near the center of fragments and clustered mismatches have an exaggerated influence on heteroduplex mobility (Upchurch et al. 2000). In this study the cumulative effect of neighbouring polymorphic sites on heteroduplex conformation and mobility has been identified ("cumulative anomaly"). We also observed the importance of mismatch position relative to the center of the molecule ("location anomaly"), and that mismatches which located within A/T- tracts ("A/T-tract anomaly") largely determine anomalous heteroduplex gel migration. It has previously been shown that anomalous migration is likely observed only for heteroduplexes formed between sequences with levels of divergence <5 %, as well as for some smaller DNA fragments (Delwart et al. 1993; Ganguly et al. 1993; Upchurch et al. 2000). Still, the maximum divergence between sequences which form heteroduplexes are analyzed in this study was 0.8 % (four mismatches across 452 bp). Overall, these effects can be characterized as

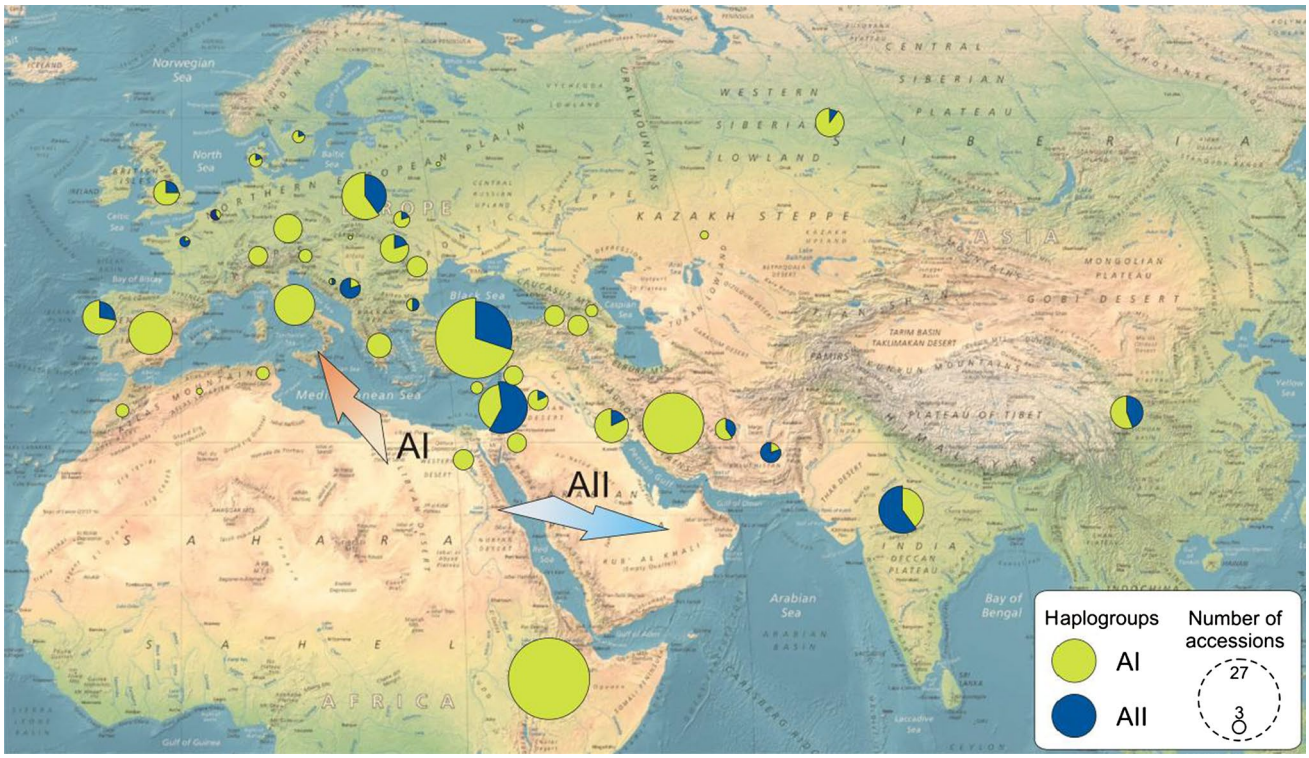

Fig. 6 Geographical distribution of the *Ppd-A1b* haplogroups in diploid, tetraploid and hexaploid wheat. The ratio of each haplogroup is shown [AI-*green* (*light circle*), AII-*blue* (*dark circle*)]. *Arrows* indicate two basic directions by which expanded wheat cultivation

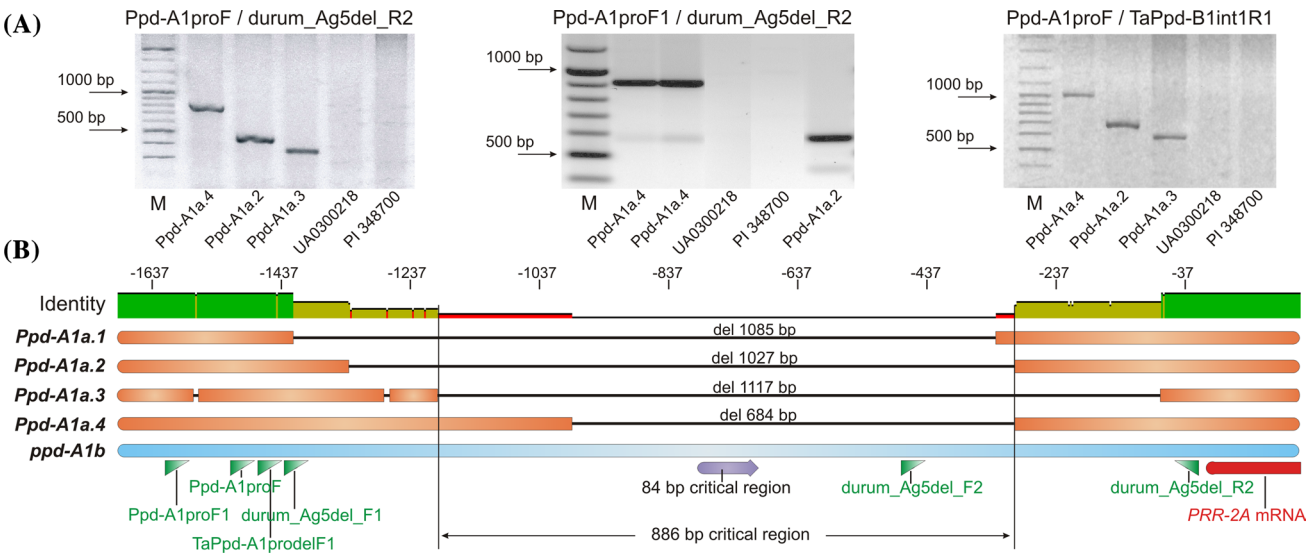

Fig. 7 **a** Amplification of *T. compactum* (PI 114638, PI 164160) with *Ppd-A1a.4* and *T. spelta* (UA0300218, PI 348700) accessions using primer pairs Ppd-A1proF/durum_Ag5del_R2, Ppd-A1proF1/durum_Ag5del_R2 and Ppd-A1proF/TaPpd-B1int1R1. The accessions of *T. durum* with *Ppd-A1a.2* (UA0200701) and *Ppd-A1a.3* (UA0201386) were used to testing of designed here PCR-markers. **b** Comparison the 5′ UTR of photoperiod insensitive alleles (*Ppd-A1a*) with the photoperiod sensitive allele *Ppd-A1b*. Deletion size, localization of primer annealing sites and "critical regions" are shown

"sequence-dependent" migration of heteroduplexes in PAA gels. In some cases, increased gel migration is inversely correlated with the number of mismatching nucleotides. This feature is widely used to evaluate of similarity and phylogenetic relationships between sequences (Delwart et al. 1993; Bowyer et al. 2000; Huang et al. 2010;

Upchurch et al. 2000). Our analyses find Hap-1/Hap3 migrates faster than Hap-3/Hap13 that confirms the later statement. The revealed features of anomalous migration of heteroduplexes can be useful to find optimal combinations between available haplotypes for HMA. For example, as it can be see from Fig. 3b, using Hap-4 as reference, the difference in migration between the Hap-1 and Hap-13 is not significant. However this difference can be increased when Hap-3 used as reference haplotype for hybridization.

Phylogenies can be estimated using distance based (distance-matrix) or character-based (maximum parsimony, maximum likelihood) methods. Distance-based algorithms are often used to infer relationships with high divergence between sequences, whereas parsimony is best used to reconstruct phylogeny for closely related sequences with a small number of changes (Mount 2008; Makarenkov et al. 2006). Here, we propose a character-based approach for inferring haplotype phylogenies and haplogroup definition (see Supplementary materials). This method is suitable for haplotype analysis (closely related sequences, intraspecific phylogenies) but requires additional algorithms to account for the events of reticulate evolution such as horizontal gene transfer, hybridization, homoplasy and recombination. A haplotype is a combination of closely linked intra-chromosome genetic markers that tend to be inherited together (Andersen and Lübberstedt 2003). In our case, haplotype is a combination of alleles (SNPs) located in same recombination unit. Therefore, horizontal gene transfer and hybridization do not influence haplotype evolution, due to absence of the recombination events between alleles of the different haplotypes. Furthermore, it assumes low homoplasy for closely related species (where the number of sequence changes per sequence position is small).

In previous studies, Takenaka and Kawahara (2012) divided *Ppd-A1* haplotypes into two groups (AI and AII) based on ~200 bp sequences located 1 kb upstream from the start codon. However, based on in silico PCR results using the primers they describe, we could not divide accessions into two haplogroups due to the identification of 11 PCR amplicons of different length. According to the criteria proposed by Takenaka and Kawahara (2012) for identification of haplogroup, amplicons of 235, 243, 244 bp are predictive for the AI haplogroup, whereas amplicons of 146, 153, 156, 245, 248, 251, 256, 279 bp predict haplogroup AII. In the present study, haplotypes were grouped into AI and AII haplogroups on the basis of their phylogenetic relationships determined by associations between alleles, and their use in the formation of haplotypes. Moreover, the 258T/C SNP which differentiates haplotypes of the AI and AII haplogroups was found to affect the intrinsic bending and curvature of DNA molecules, resulting in different electrophoretic mobility in PAA gels. Due to this, haplotypes could be unambiguously allocated to haplogroups during separation of *Ppd-A1b* products in PAA gels. Takenaka and Kawahara (2012, 2013) also suggested that haplogroup AII is older than AI, as they found all einkorn accessions to belong to the former (AII). However, as they show, AII possesses 2.5 times less haplotypes than AI, and with lesser geographic and species distribution. Furthermore, einkorn wheats are the diploid A genome progenitors [in particular *T. urartu* is thought to be the A genome donor for tetraploid (genome BBA^uA^u, GGA^uA^u) and hexaploid (BBA^uA^uDD) wheat species (Golovnina et al. 2007)], therefore their haplogroup must be predominant in wheat. Our study shows that the einkorn wheat *T. urartu* belongs to the AI haplogroup, which has a wide spread in wheat and it is therefore more likely that its haplotype-founder is evolutionary older than the haplotype which founded of AII haplogroup. In addition, we found in silico amplicons of *T. monococcum* include a 4 bp insertion but combination nucleotides of polymorphic positions forms the Hap-8 haplotype (AI). This supports previous reports that *T. monococcum* and *T. urartu* are very closely related (Johnson and Dhaliwal 1976).

Our assertion that the AI haplogroup is evolutionarily older than AII is based on the following observations: (1) AI contains more haplotypes. (2) AI is most widespread in terms of species and geographic distribution. (3) AI includes all haplotypes specific to einkorn wheat (*T. monococcum*, *T. urartu*) and is predominant in tetraploid wheat (which predates hexaploid wheat). (4) AI haplotypes contain an intact T-tract region in the *Ppd-A1*promoter. (5) From the results of the phylogenetic network reconstruction using "interallelic association" and MJ methods. This conclusion is also consistent with the geographical distribution of AI haplotypes, which were found to predominate in countries of the Fertile Crescent, and are found exclusively in Jordan (100 % accessions from this country possess AI), Lebanon (100 %), Iran (95 %) and Iraq (82 %). The Fertile Crescent is the wheat centre of origin and site of domestication (Matsuoka 2011; Kilian et al. 2009). This study shows the AI haplogroup is also widespread in northern regions. In contrast, AII is predominant in the southeast regions such as India, Pakistan, Israel and especially USA. *T. sphaerococcum,* which possesses the Hap-1 haplotype belonging to the AII haplogroup, is endemic to the Indian subcontinent (southern Pakistan and north-western India) (Ellerton 1939). The observed distributions of the AI and AII haplogroups is in good agreement with existing knowledge on domestication routes, by which wheat cultivation expanded in two directions: eastward to India and northwestward to the Mediterranean coastal region and Europe (Matsuoka 2011; Kilian et al. 2009). The domestication of *T. turgidum* was initiated from wild emmer wheat (*T. dicoccoides*) (Matsuoka 2011). Hap-7 was identified as the ancestral haplotype for polyploid wheat species, detected only *T. dicoccoides* accessions. Among tetraploid species,

only *T. turgidum* (4.65 %) and *T. dicoccoides* (16.28 %) include haplotype Hap-1, which is widespread in hexaploid wheat. This observation is consistent with the evolutionary origin of hexaploid wheat from *T. dicoccoides* (*T. turgidum* lineage wheats; Golovnina et al. 2007; Matsuoka 2011), parallel to *T. durum, T. polonicum, T. carthlicum*.

We observed a decrease of haplotype diversity after the genetic bottlenecks occurring due to allopolyploidization and speciation events. Allopolyploidization is often accompanied by a decrease in genetic variation due to the genetic bottleneck of the founder effect (Haudry et al. 2007; Feldman and Levy 2012). In tetraploid wheat, the AI haplogroup was found to be predominant. However, after formation of allohexaploid wheat species, the AII haplogroup became widespread. Genetic bottlenecks and founder effects are known to be associated with origin of *T. shaerococcum* and *T. polonicum*. All analyzed *T. sphaerococcum* accessions were found to possess the Hap-1 haplotype (haplogroup AII) predominant in hexaploid wheat. However, all *T. polonicum* accessions are characterized by Hap-4 (AI), predominant in tetraploid wheat. The low genetic variation found in this species, as well as for *T. carthlicum*, *T. turanicum*, can be explained by their later evolutionary emergence. Reproductive isolation (including isolation within cultivation systmes) may also explain the emergence of species-specific haplotypes in *T. timopheevii* (Hap-11), *T. urartu* (Hap-8, Hap-9, Hap-10), *T. dicoccoides* (Hap-5, Hap-7, Hap-13), *T. durum* (Hap-12), as well as specific 456 bp amplicons in *T. monococcum*.

In a previous study, Wilhelm et al. (2009) identified a ~900 bp region of the *Ppd1* promoter as critical for the control of gene expression. This region is defined by a series of deletions that have a most distal startpoint (*Ppd-A1a.2*) and most proximal end (*Ppd-A1a.3*) relative to the transcription initiation site. This region contains overlapping deletions of the dominant (photoperiod insensitive) *Ppd-A1a.1* (Nishida et al. 2013), *Ppd-A1a.2*, *Ppd-A1a.3* (Wilhelm et al. 2009) and *Ppd-D1a* alleles (Beales et al. 2007). Furthermore, a highly conserved ~100 bp region within the 900 bp region is thought to contain important motifs mediating down-regulation of *Ppd1* expression under short days (Wilhelm et al. 2009). The *Ppd-A1a.4* allele identified in this study lacks this 100 bp region, and most of the 900 bp "critical region", except for 207 bp at the 5′ end. Moreover, the 684 bp *Ppd-A1a.4* deletion is wholly spanned by the 1027 bp *Ppd-A1a.2* deletion, indicating it may have a similar affect on photoperiod sensitivity. However, unlike the *Ppd-A1a.2* deletion, the 684 bp *Ppd-A1a.4* deletion is flanked by a dinucleotide repeat ("GC"). This suggest that this deletion occurred via the nonhomologous end-joining of double-stranded DNA break repair mechanism, mediated by the single-strand annealing-like mechanism (Puchta 2005), as reported for the *Vrn1* genes (Cockram et al. Cockram et al. 2007a, b) and likely for *Ppd-A1a.1* (Nishida et al. 2013). While all four *Ppd-A1a* deletions share common deletions between nucleotides −986 and −336, the *Ppd-A1a.4* deletion is the shortest of all the known deletions. As a result, the left border of the "critical region" is shifted by 207 bp towards the start codon relative to the left border of the *Ppd-A1a.3* deletion.

As PCR analysis of the *Ppd-A1* promotor in *T. spelta* (accessions: UA0300218 and PI 348700) did not result in amplification products, it is assumed that these accessions either contain promoter mutations that encompass primer annealing sites or a large insertion, which could not be amplified at given PCR conditions. This result significantly increases probability of dysfunction in the regulation of *Ppd-A1* expression, and indicates new mutant variant of photoperiod sensitive (*ppd-A1b*, recessive) allele.

It is known that there is no genome specificity in *Ppd1* actions (Shaw et al. 2012). Due to this, the dominant *Ppd1* alleles from different genomes which possess similar mutations of *PRR* genes are characterized by the similar photoperiod insensitivity and effects on flowering time (Bentley et al. 2011; Shaw et al. 2012). The new photoperiod insensitivity allele *Ppd-A1a.4* identified here can be utilized as genetic source to breed early heading cultivars. Moreover, the contraction of the "critical region" it allows will help in the identification of the putative *Ppd1* promoter regulatory sequences, which currently remain undefined (Wilhelm et al. 2009; Nishida et al. 2013).

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8(11):554–560

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16(1):37–48

Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet 115(5):721–733

Bentley AR, Turner AS, Gosman N, Leigh FJ, Maccaferri M, Dreisigacker S, Greenland A, Laurie DA (2011) Frequency of photoperiod-insensitive *Ppd-A1a* alleles in tetraploid, hexaploid and synthetic hexaploid wheat germplasm. Plant Breed 130(1):10–15

Bentley AR, Jensen EF, Mackay IJ, Hönicka H, Fladung M, Hori K, Yano M, Mullet JE, Armstead IP, Hayes C, Thorogood D, Lovatt A, Morris R, Pullen N, Mutasa-Göttgens E, Cockram J (2013) Flowering time. In: Cole C (ed) Genomics and breeding for climate-resilient crops 2. Springer, Berlin, pp 1–67

Bowyer J, Verrills N, Gillings MR, Holmes AJ (2000) Heteroduplex mobility assay as a tool for predicting phylogenetic affiliation of environmental ribosomal RNA clones. J Microbiol Methods 41(2):155–160

Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the VNTR locus *D1S80* by the PCR followed by high-resolution PAGE. Am J Hum Genet 48(1):137–144

Campoli C, Drosse B, Searle I, Coupland G, von Korff M (2012) Functional characterisation of *HvCO1*, the barley

(*Hordeum vulgare*) flowering time ortholog of *CONSTANS*. Plant J 69(5):868–880

Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007a) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot 58(6):1231–1244

Cockram J, Mackay IJ, O'Sullivan DM (2007b) The role of double-stranded break repair in the creation of phenotypic diversity at cereal *VRN1* loci. Genetics 177(4):2535–2539

Cockram J, Thiel T, Steuernagel B, Stein N, Taudien S, Bailey PC, O'Sullivan DM (2012) Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS One 7:e45307

Delwart EL, Shpaer EG, Louwagie J, McCutchan FE, Grez M, Rübsamen-Waigmann H, Mullins JI (1993) Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. Science 262(5137):1257–1261

Diaz A, Zikhali M, Turner A, Isaac P, Laurie D (2012) Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLoS One 7(3):e33234

Diekmann S (1989) The migration anomaly of DNA fragments in polyacrylamide gels allows the detection of small sequence-specific DNA structure variations. Electrophoresis 10(5–6):354–359

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Ellerton S (1939) The origin and geographical distribution of *Triticum sphaerococcum* prec. and its cytogenetical behaviour in crosses with *T. vulgare* Vill. J Genet 38:307–324

Feldman M, Levy AA (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192(3):763–774

Ganguly A, Rock MJ, Prockop DJ (1993) Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90(21):10325–10329

Golovnina KA, Glushkov SA, Blinov AG, Mayorov VI, Adkison LR, Goncharov NP (2007) Molecular phylogeny of the genus *Triticum* L. Plant Syst Evol 264(3–4):195–216

Guo Z, Song Y, Zhou R, Ren Z, Jia J (2010) Discovery, evaluation and distribution of haplotypes of the *Ppd-D1* gene. New Phytol 185:841–851

Gusfield D (1997) Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge University Press, Cambridge

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, David J (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24(7):1506–1517

Huang T, Yeh Y, Tzeng DD (2010) Heteroduplex mobility assay for identification and phylogenetic analysis of anthracnose fungi. J Phytopathol 158(1):46–55

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinform 8(1):460–466

Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science 309(5732):293–297

Johnson BL, Dhaliwal HS (1976) Reproductive isolation of *Triticum boeoticum* and *Triticum urartu* and the origin of the tetraploid wheat. Am J Bot 63(8):1088–1094

Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, Jones G, Jones MK, Brown TA, Powell W (2008) Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol 25:2211–2219

Jones H, Civan P, Cockram J, Leigh FJ, Smith LMJ, Jones MK, Charles MP, Molina-Cano J-L, Powell W, Jones G, Brown TA (2011) Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. BMC Evolut Biol 11:e320

Kato K, Yokoyama H (1992) Geographical variation in heading characters among wheat landraces, *Triticum aestivum* L., and its implication for their adaptability. Theor Appl Genet 84(3–4):259–265

Kilian B, Özkan H, Pozzi C, Salamini F (2009) Domestication of the *Triticeae* in the Fertile Crescent. In: Muehlbauer GJ (ed) Genetics and genomics of the *Triticeae*. Springer, New York, pp 81–119

Koo HS, Wu HM, Crothers DM (1986) DNA bending at adenine-thymine tracts. Nature 320(6062):501–506

Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K (2009) Conformational analysis of nucleic acids revisited: curves+. Nucleic Acids Res 37(17):5917–5929

Law CN, Worland AJ (1997) Genetic analysis of some flowering time and adaptive traits in wheat. New Phytol 137:19–28

Makarenkov V, Kevorkov D, Legendre P (2006) Phylogenetic network construction approaches. Appl Mycol Biotechnol 6: 61–97

Marini JC, Levene SD, Crothers DM, Englund PT (1982) Bent helical structure in kinetoplast DNA. Proc Natl Acad Sci USA 79(24):7664–7668

Matsuoka Y (2011) Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant Cell Physiol 52(5):750–764

Mizuno T, Nakamichi N (2005) Pseudo-response regulators (PRRs) or true oscillator components (TOCs). Plant Cell Physiol 46(5):677–685

Mount DW (2008) Choosing a method for phylogenetic prediction. CSH Protoc pdb.ip49

Nakamichi N, Kita M, Niinuma K, Ito S, Yamashino T, Mizoguchi T, Mizuno T (2007) Arabidopsis clock-associated pseudo-response regulators PRR9, PRR7 and PRR5 coordinately and positively regulate flowering time through the canonical CONSTANS-dependent photoperiodic pathway. Plant Cell Physiol 48(6):822–832

Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, Laurie DA, Kato K (2013) 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 31(1):27–37

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612

Polzin T, Daneschmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. Oper Res Lett 31(1):12–20

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56(409):1–14

Scarth R, Law CN (1984) The control of day-length response in wheat by the group 2 chromosomes. Z Pflanzenzücht 92:140–150

Shaw LM, Turner AS, Laurie DA (2012) The impact of photoperiod insensitive *Ppd-1a* mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum aestivum*). Plant J 71(1):71–84

Stellwagen NC (2009) Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. Electrophoresis 30(1):188–195

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the

circadian clock and the control of flowering in Arabidopsis. Nature 410(6832):1116–1120

Takenaka S, Kawahara T (2012) Evolution and dispersal of emmer wheat (*Triticum* sp.) from novel haplotypes of *Ppd-1* (photoperiod response) genes and their surrounding DNA sequences. Theor Appl Genet 125(5):999–1014

Takenaka S, Kawahara T (2013) Evolution of tetraploid wheat based on variations in 5′ UTR regions of *Ppd-A1*: evidence of gene flow between emmer and timopheevi wheat. Genet Resour Crop Evol 60(7):2143–2155

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. Science 310(5750):1031–1034

Upchurch DA, Shankarappa R, Mullins JI (2000) Position and degree of mismatches and the mobility of DNA heteroduplexes. Nucleic Acids Res 28(12):E69

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 303(5660):1003–1006

Wilhelm EP, Turner AS, Laurie DA (2009) Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). Theor Appl Genet 118(2):285–294

Worland AJ, Börner A, Korzun V, Li WM, Petrovic S, Sayers EJ (1998) The influence of photoperiod genes on the adaptability of European winter wheats. Euphytica 100(1–3):385–394