The sequence polymorphism of *SBEIIa* gene in wheat (*Triticum* sp.)

Konovalov FA, Shaturova AS, Kudryavtsev AM

N.I.Vavilov Institute of General Genetics, 119991 Moscow, Russia; e-mail: konovalov@pisumsativum.org

INTRODUCTION

The *SBEIIa* gene in common wheat encodes a starch branching enzyme that converts growing amylose chains to amylopectin tree-like structures and maintains normal amylose/amylopectin ratio in endosperm starch[4,5,6]. In contrast with maize and other cereals where the expression of the paralogous *SBEIIb* gene is critical for maintaining wild-type amylose content in endosperm starch, the wheat *SBEIIb* expression in endosperm is limited, and its function is likely to be taken over by *SBEIIa*[5,6]. Antisense repression experiments clearly demonstrated the dramatic effects of simultaneous *SBEIIa* and *SBEIIb* inactivation on the seed amylose content[6]. As well as transgenic wheat non-transgenic mutant plants lacking the activity of all three *SBEIIa* homeologs (in A, B and D subgenomes) are expected to have elevated amylose content and may be used as a new industrial source of enzyme-resistant starch (RS), an important component of human nutrition. The partial mutants carrying non-functional alleles of *SBEIIa* homeologs may be searched for among the existing germplasm or produced by induced mutagenesis; however, there is no evidence that *SBEIIa* recessive mutations in a hexaploid genetic background may lead to some characteristic phenotype or are associated with clear changes in seed protein electrophoretic spectra. Instead, DNA sequence-based screening methods, like TILLING, can be used for the task[7].

To date the full sequence of wheat *SBEIIa* gene has been reported for the D-genome donor species, *Aegilops tauschii*[2,4]. Sequencing of *SBEIIa* homeologous copies in common wheat is an essential step in performing the mutation search at the DNA level; it also provides valuable information on wheat evolution and phylogeny, in particular the evolution of the starch branching enzyme family in cereals. The gene itself contains multiple small exons separated by introns of various size, making it a useful model for future studies of neutral vs. selective sequence changes. The gene region spanning introns 9-11 is also a good target for designing the subgenome-specific primers that can amplify the most functionally conserved downstream exon sequences[4], where the nucleotide substitutions have higher probability of affecting enzyme function.

MATERIALS AND METHODS

The plant material used in this study was obtained from N.I.Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR; St.Petersburg, Russia; Table 1).

Table 1. *The plant material used in the study.*

№	Species	Origin	VIR Cat. №
1	*T.aestivum* cv. Chinese Spring	China	K-44435
2	*T.aestivum* cv. Bezostaya 1	Russia	K-42790
3	*T.durum* cv. Langdon	USA	K-44404
4	*T.durum* (ICARDA IG6878)	Syria	–
5	*T.monococcum*	England	K-62499
6	*T.sinskajae*	Russia	K-48993
7	*T.boeoticum*	Armenia	K-58508
8	*T.urartu*	Armenia	K-33871
9	*T.araraticum* (ICARDA IG46434)		–
10	*T.turgidum* v. *melanotherum*	Turkey	K-16086
11	*T.turgidum* v. *rubralbum*	Chile	K-29716
12	*T.turgidum* v. *dinurum*	Switzerland	K-25086
13	*T.turgidum* v. *turgidum*	Italy	K-19336
14	*T.aethiopicum* v. *rufescens*	Yemen	K-25674
15	*T.aethiopicum* v. *arraseita*	Uzbekistan	K-54234
16	*T.araraticum* v. *thumaniani*	Azerbaijan	K-31121
17	*T.araraticum* v. *kurdistanicum*	Iraq	K-59941
18	*T.dicoccoides* v. *fulvovillosum*	Israel	K-61702
19	*T.dicoccoides* v. *namuricum*	Israel	K-61714
20	*T.turanicum* v. *notabile*	Iraq	K-14301
21	*T.turanicum* v. *ensigne*	Turkmenistan	K-56611
22	*T.sphaerococcum* v. *globosum*	India	K-23790
23	*T.sphaerococcum* v. *rotundatum*	Pakistan	K-23822
24	*T.macha* v. *rubiginosum*	Georgia	K-29576
25	*T.macha* v. *macha*	Georgia	K-38548
26	*T.timopheevii* v. *viticulosum*	Georgia	K-29537
27	*T.timopheevii* v. *viticulosum*	Russia	K-47792
28	*T.dicoccum* v. *arras*	India	K-46482
29	*T.persicum* v. *rubiginosum*	Armenia	K-25170
30	*T.polonicum* v. *kiritchenko*	Germany	K-25344
31	*T.polonicum* v. *gorskyi*	China	K-43335

The total DNA was extracted from individual plant

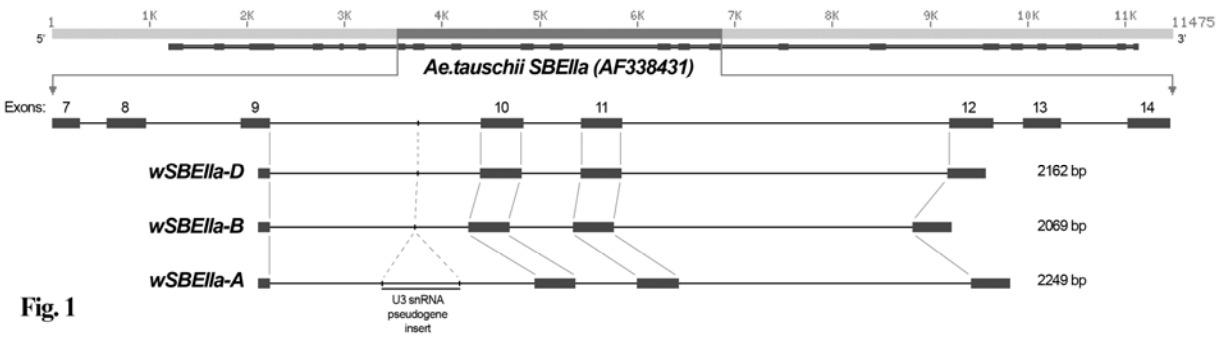

Fig. 1

leaves using one of the versions of CTAB protocol with minor modifications. PCR was performed using Fermentas TrueStart™ DNA polymerase in an MJ Research PTC-100 thermal cycler according to the standard protocol with 1.5 mM $MgCl_2$ and some minor optimization. The degenerate primer sequences for *SBEIIa* exons 9 and 12 are 5'-GGATWTATGAAWCACAYRTTGG and 5'-CCCATAGTTRAAYASACGRGAATC, respectively. For PCR screening of inserts the primer sequences were 5'-GCTTAAGATGGTTAGGGTTTCC and 5'-TTACCTATGAACAATATCCATAAGAAC.

The PCR products corresponding to exons 9-12 were cloned into a pGEM-T vector and sequenced on an ABI PRISM 3100-Avant system. Multiple clones were sequenced in order to correct for polymerase errors.

RESULTS AND DISCUSSION

We have isolated and sequenced three distinct copies of *SBEIIa* fragment (exons 9-12) using PCR with degenerate primers from common wheat cv. Chinese Spring DNA and also from the DNA of diploid wheat *T.monococcum* (Genbank entries EU024966–EU024969; Fig.1). The *T.aestivum* sequences were assigned to the corresponding subgenomes by their comparison with the sequences of *T.monococcum* and *Ae.tauschii* homologous *SBEIIa* fragments (data not shown).

The *SBEIIa* sequences of common wheat were subjected to multiple alignment using the Kalign algorithm. The exon-intron structure of *SBEIIa* homeologs is similar to *Ae.tauschii SBEIIa*. In total, 156 nucleotide positions in the alignment contain HSVs (homeolog-specific variations), with SSR-like polymorphic sites excluded. The average HSV frequency calculated for mean sequence length (2160 bp) is about 1 HSV per 13.9 base pairs. However, only 8 HSVs were found in exon sequences (exons 9-12) with a total length of 398 bp (the HSV frequency here is thus only 1 substitution per 49.75 bp), and the substitution frequency calculated for introns only reaches (2160-398)/(156-8)=11.91 bp^{-1}. While at some HSV sites (27) all three sequences differ from each other, in most cases only one sequence has a base substitution. It can be postulated that such HSVs occurred after the evolutionary divergence of the A, B and D-genome lineages. Using this type of HSVs only, it is possible to calculate the relative degree of sequence divergence from the 'consensus' one and to make a speculative comparison of base substitution rates in corresponding genomes[3]. However, the three genome lineages most possibly did not originate from a single ancestor simultaneously, so (given the substitution rates remain equal) the relative sequence differences can provide some information on species divergence order, as on a given timescale this factor may play a more important role in SNP accumulation than molecular clock rate changes. Out of 129 'subgenome-specific' HSVs there are 45 substitutions in the A-genome sequence, 55 in the B-genome and 29 in the D-genome homeolog. This can be compared with the data from completely different wheat homeologous sequences, *KN1*-like homeobox gene copies[3]; our ratio of 45:55:29 is not equal to 133:231:81, but in both cases the B genome contains the most divergent sequence, and D genome – the least divergent one.

In addition to the base substitutions described above there are several cases of insertion/deletion (indel) polymorphism between the homeologous sequences. They can be divided in two classes: small indels (<10 bp), 14 of which are distributed over the introns; and large indels (≥10 bp). In total there are 6 sites in the sequence alignment that contain such indels, and in two cases the nested large deletions covering different base pair numbers can be seen at the same site in two sequences. The first large indel is located in the first half of intron 9 and is ~150 bp long (A-genome homeolog has the gap which is filled by 149 bp in B genome and 152 bp in D genome; the difference is due to several small nested indels). BLAST searches did not reveal any resemblance between the indel sequence and known transposons, and the sequence is not palindromic. As the ~150-bp fragment is absent from the A genome only, it can be speculated that it is more likely to be a deletion; however, no direct repeats are found at the sequence ends or its flanking sites, and the possible mechanism for such a deletion event is uncertain.

The second medium-sized deletion site can be found in both the B and D genomes at bases 360 and 363, respectively. In the A genome it contains a 30-bp sequence which is totally absent in the B genome and is partially deleted in the D genome. The full deletion in the B genome could occur through recombination between short direct repeats (AAGA) at the sequence ends.

One of the inserts in the A-subgenome copy intron 9 is similar to wheat U3 snRNA gene (Genbank no. X63065;

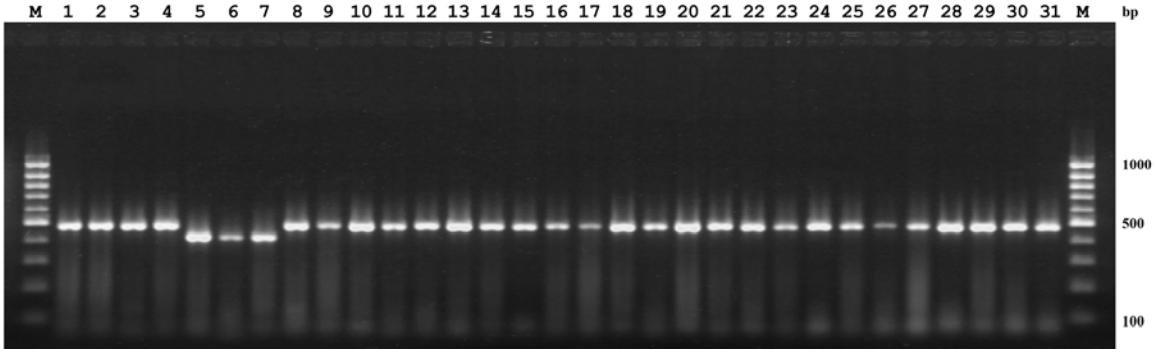

Fig. 2

91% identical over 207 bp) and by some features is likely to have arisen as an L1-processed pseudogene: it is reverse-oriented, covers the length of processed snRNA sequence, is flanked by direct 17-bp duplications and contains a short poly-T tail at 5'-end. Interestingly, the results of interaction between snRNAs and LINE retrotransposition machinery were observed in the human genome as well[1]. In *T.monococcum*, the pseudogene 3'-end also contains a 56-bp deletion that extends to the flanking intron sequence. The PCR screening of 31 *Triticum* accessions using the primers specific for A-genome demonstrated that U3-like processed pseudogene insert is present in *T.urartu* genome and in the A genome of all polyploid wheats, and the 56-bp deletion at the pseudogene 3'-end is a feature of diploid wheats carrying A^m genome, including *T.boeoticum*, *T.monococcum* and *T.sinskajae* (Fig. 2; the lane numbers correspond to the samples listed in Table 1). In the A^u genome lineage, from *T.urartu* to various tetra- and hexaploid wheat species, the deletion is absent, once again confirming that *T.urartu* is an A-genome donor of major cultivated wheat polyploids[2]. It can be speculated that the A genome of all wheat species used in the study originated from a single plant (more precisely, from a single cell) in which the integration of snRNA pseudogene by L1-class retrotransposon had taken place, as the probability of such an event occurring independently two or more times is likely to be negligible. Despite the fact that the excision of the insert by recombination between its end repeats is possible, the insertion-deletion site in *SBEIIa* may mark the two points of divergence in the evolutionary course of *Triticum/Aegilops* species, and it will be interesting to determine the insertion state among various *Aegilops* accessions as well.

ACKNOWLEDGEMENTS

We would like to thank Dr. Olga P. Mitrofanova (VIR, St.Petersburg, Russia) for selecting and providing the plant material, and also for insightful comments on our work.

The work was supported by Russian Foundation for Basic Research grant no. 08-04-01401-a.

REFERENCES

1. Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of L1. Genomics 80(4):402-406, 2002.
2. Feldman M. Origin of cultivated wheat // The World Wheat Book. A history of wheat Breeding / Eds A.P. Bonjiean, W.J. Angus. L.; P.; N.Y.: Tec&Doc/Intercept Ltd., P. 3-56, 2001.
3. Morimoto R, Kosugi T, Nakamura C, Takumi S. Intragenic diversity and functional conservation of the three homoeologous loci of the *KN1*-type homeobox gene *Wknox1* in common wheat. Plant Mol Biol 57(6): 907-924, 2005.
4. Rahman S, Regina A, Li Z, Mukai Y, Yamamoto M, Kosar-Hashemi B, Abrahams S, Morell MK. Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms. Characterization of a gene for starch-branching enzyme IIa from the wheat genome donor Aegilops tauschii. Plant Physiol. 125(3):1314-24, 2001.
5. Regina A, Kosar-Hashemi B, Li Z, Pedler A, Mukai Y, Yamamoto M, Gale K, Sharp PJ, Morell MK, Rahman S. Starch branching enzyme IIb in wheat is expressed at low levels in the endosperm compared to other cereals and encoded at a non-syntenic locus. Planta. 222(5):899-909, 2005.
6. Regina A, Bird A, Topping D, Bowden S, Freeman J, Barsby T, Kosar-Hashemi B, Li Z, Rahman S, Morell M. High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. Proc Natl Acad Sci USA. 103(10):3546-51, 2006.
7. Slade A, Fuerstenberg S, Loeffler D, Steine M, Facciotti D. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23(1):75-81, 2005.