BARD

FINAL REPORT

PROJECT NO. US-2926-97

**Enhancement of Essential Amino Acids in Cereal
Seeds: Four Approaches to Increased Lysine Content**

O.D. Anderson, G. Galili, A.E. Blechl

2002

Date: 5-9-02

BARD Project Number: US-2926-97

Project Title:

Enhancement of essential amino acids in cereal seeds: four approaches to
increased lysine content.

Investigators:
(Principal first)

Olin D. Anderson

Gad Galili

Ann E. Blechl

Affiliated Institution:

USDA, ARS, WRRRC, Albany, CA

Weizmann Institute of Science, Israel

USDA, ARS, WRRRC, Albany, CA

Project Start Date: Oct. 1, 1998

Type of Report: 1st Annual _____ 2nd Annual _____ Final X

Signature:

Principal Investigator

Signature:

Authorizing Official, Affiliated Institution

**BARD Final Scientific Report
Cover Page**

Publication Summary (numbers)

	Joint IS/US authorship	US Authors only	Israeli Authors only	Total
Refereed (published, in press, accepted)				0
Submitted, in review, in preparation				0
Invited review papers				0
Book chapters				0
Books				0
Master theses				0
Ph.D. theses				0
Abstracts				0
Not refereed (proceedings, reports, etc.)				0

Postdoctoral Training: List the names and social security/identity numbers of all postdocs who received more than 50% of their funding by the grant.

Kahana Anat: Identity Number: 02306812

Cooperation Summary (numbers)

	From US to Israel	From Israel to US	Together, elsewhere	Total
Short Visits & Meetings		1		1
Longer Visits (Sabbaticals)				0

Description of Cooperation:

The two laboratories joined their complementary expertise and biological resources to explore the limitation of parameters related to developing genetic engineering approaches to high-lysine wheat. The US lab cloned the wheat LKR/SDH gene of lysine catabolism for both wheat and barley using the orthologous corn gene and wheat ESTs identified to be from the LKR/SDH gene. The Israeli lab had previously isolated and characterized a partial sequence that was used to product anti LKR/SDH monoclonal antibodies that enabled to study the regulation of expression of the LKR/SDH gene, and also to study its activity. The Israeli lab constructed a chimeric gene encoding bacterial DHPS and AK (properties of the Israeli lab) under a wheat HMW-GS promoter (promoter provided by the US lab) and the US lab has transformed this gene into transgenic wheat using technology developed by them. Based on a cooperative development of expression system for wheat HMW-GS and their subsequence assembly studies, developed by both Israeli and US labs, the US lab engineered recombinant high-lysine HMW-GS. Engineering was designed to minimize interference in the

BARD Final Scientific Report
Cover Page

structure of these proteins and they function in bread making and other technological qualities of wheat. Transgenic wheat lines are being characterized and will be further analyzed by both the U.S. and Israeli laboratories.

Patent Summary (numbers)

	Israeli inventor (s) only	US inventor (s) only	Joint IS/US inventors	Total
Submitted				0
Issued (allowed)				0
Licensed				0

TABLE OF CONTENTS

PAGE

Table of Contents	1
Abstract	2
General Objectives	3
Relevant Data, Methodology, Results and Discussion	3
Evaluation of the Research Achievement	13
Description of the Coordination	14
Conclusions	14
List of Publications	15
Table 1	16
Table 2	17
Figure 1	18
Figure 2	19
Figure 3	20
Figure 4	22
Figure 5	26
Figure 6	28
Figure 7	35
Figure 8	36
Figure 9	37
Figure 10	38
Figure 11	39
Figure 12	40
Figure 13	46
Figure 14	47
Figure 15	49
Figure 16	50

**ABSTRACT: ENHANCEMENT OF ESSENTIAL AMINO ACIDS IN CEREAL SEEDS:
FOUR APPROACHES TO INCREASED LYSINE CONTENT**

Cereal seeds are the basis of the human diet, and their amino acid composition is thus of major nutritional and economic importance. Currently, deficiencies in essential amino acids are addressed, when possible, by additional protein sources or by supplementing animal feed with non-cereal protein or synthetic amino acids. A number of strategies have been suggested to make cereal flours more complete and balanced sources of amino acids, although systematic examination of such strategies is rare. This project proposed to begin such a systematic examination using four complementary and parallel approaches to increasing wheat seed lysine: 1) Modifying endogenous wheat seed proteins for increased lysine composition. 2) Overexpression of naturally occurring high-lysine proteins in the wheat endosperm. 3) Ectopic expression of proteins in the wheat endosperm. 4) Alteration of free lysine levels in the wheat endosperm. The results of these studies are expected to be wheat lines with increased lysine content and will establish a clearer understanding of the approaches most likely to enhance cereal seed protein quality. Progress is reported for all four objectives, with a significant foundation for further work on two of the objectives (modification of wheat storage proteins and lysine metabolism). Plans for continuing work on all four objectives are briefly outlined.

GENERAL OBJECTIVES: 1) Modification of endogenous polypeptide amino acid content. 2) Increases in endogenous polypeptides high in lysine. 3) Synthesis, processing, and stability of heterologous proteins in cereal endosperm. 4) Modulation of lysine metabolism in cereal endosperm.

RELEVANT DATA, METHODOLOGY, RESULTS AND DISCUSSION:

Objective #1: Modification of endogenous polypeptide amino acid content. Two targets for increasing lysine content in endogenous seed proteins are gliadins and the highly expressed high molecular weight glutenins. We have made a set of gene constructs, based on substituting coding sequences for engineered high lysine repeats into several protein backbones. Some were designed for expression in *Escherichia coli* so that proteins isolated from readily generated lysates could be used in preliminary tests of effects on dough quality. Other constructs were designed for transformation into and expression in wheat. Using the synthetic high molecular weight glutenin Dx5 construct containing just the α terminal non-repetitive domains (Anderson *et al.* (1996) *Gene* 174:51-58), we substituted all the arginine codons with lysine codons and placed restriction sites throughout the sequence to facilitate generation of variants containing different inserts (Fig 1). *E. coli* codon preferences were used to enhance expression in the bacteria. From this plasmid, two sets of glutenin constructs were generated. One set encoded variants with the normal synthetic Dx5 terminal sequences plus 16 or 32 copies of the Dx5 15-mer repeat, along with the high lysine repeats described below (Table 1). The other encoded variants containing the high lysine N-terminal non-repetitive region ("R2K") with 16 or 32 copies of the Dx5 15-mer repeat, and in some cases, also the high

lysine repeats. All versions of the glutenin were then cloned into the pET3A plasmid (Novagen, Madison, WI) for inducible expression in *E. coli*.

Three different high-lysine repeat sequences were used in these constructions (Fig. 2). "PEP" is a tetramer of the repeat PEPMPKPEPKPEPK, based on a naturally occurring high-lysine repeat found in a wheat proline-rich protein (Raines *et al.* (1991) *Plant Molecular Biology* 16:663-670). "MEEK" is an octamer of the repeat MEEKLKA and is based on a *de novo* high-lysine sequence that was well expressed in transgenic tobacco in another laboratory (Keeler *et al.* (1997) *Plant Molecular Biology* 34:15-29). "PQQ" is a tetramer of the repeat (PQQPGQEKGYYYP) that is identical to the synthetic Dx5 15-mer repeat (PQQPGQGQQGYYP; Anderson *et al.* (1996) *Gene* 174:51-58) except for the substitution of lysine and glutamate residues for glutamine and glycine, respectively. The glutamate substitutions were made to maintain charge neutrality.

All three of these repeats were cloned to permit oligomerization (Fig 3). A clone was obtained that contained the MEEK repeat in the orientation that permitted generation of repeats by using the strategy schematized in Figure 3A. The clones with the PEP and PQQ inserts were in the opposite orientation and thus had a *SpeI* site on one side of the insert and the original vector *SpeI* site on the other side. Cutting these clones with *DraIII* and *SpeI* in the oligomerization protocol would yield the insert with *SpeI* sites on either side, which is not usable in the oligomerization scheme (Figure 3B). To remedy this, we eliminated the vector *SpeI* site by cutting it and filling in with Klenow fragment DNA polymerase I. The PEP and PQQ inserts were then recloned into the vector lacking a *SpeI* site. We have generated dimers of all three repeats, and tetramers of both MEEK and PEP repeats.

The DNAs of most of the clones listed in Table I have been transformed into the *E. coli* expression host, BL21 (DE3) and cultures of the resultant clones have been grown and induced by IPTG. Induction results in the synthesis of the engineered proteins which were visualized by Coomassie blue staining of electrophoretic gels of the lysates (Fig. 4 A-D). Most of the lanes had strongly stained bands at the predicted size for the engineered glutenin. Curiously, none of the constructs with the MEEK repeat produced proteins of the predicted sizes (Fig. 4B, Lanes 5-7; Fig. 4D, Lanes 2-3). This could be because of a lack of solubility (e.g. formation of exclusion bodies or incompatibility with the purification protocol) or instability in the cloning system employed. In addition, some of the constructs containing the PQQ repeats resulted in polypeptides of different sizes from different clone isolates (Fig. 4A, lanes 1-2; Fig. 4D, lanes 5-6), indicating that these constructs can lose or gain coding sequences for PQQ repeats during growth in *E. coli*. In future experiments, it will be necessary to characterize PQQ-containing plasmids after growth to avoid use of a deletion derivative.

Future experiments will entail growth of larger cultures to allow isolation of quantities of engineered proteins that will be sufficient for use in physical/chemical characterizations and micro-mixograph experiments designed to measure their effects on dough properties.

As a second approach to Objective 1, we have constructed a synthetic high lysine version of a wheat α -gliadin gene. Lysines were substituted into positions where either an arginine or lysine is found in one or more member of the α -gliadin family of proteins (Anderson and Greene (1997) *Theor Appl Genet* 95:59-65). Restriction sites were placed as with the Dx5 construct to aid in production of permutations of the gliadin and high

lysine repeats (Fig 5). Several different constructs of the gliadin were obtained by cloning in different numbers and kinds of the high-lysine repeats (Table 2; Fig. 6). The various coding region derivatives have been cloned into a plasmid with seed-specific expression elements, constructed in our laboratory and called pGlu10(SFe)R853. It consists of the 5' flanking, promoter, and signal peptide-encoding regions from the high molecular weight glutenin Dy10, spliced in frame to the remainder of the coding region and 3' flanking regions of the wheat gene for Dx5 ("Dy10-Dx5", Table 2). Using the *StuI* and *FseI* sites in this vector and in the engineered high-lysine gliadin gene, the gliadin coding region was substituted for the majority of the Dx5 coding sequence, while maintaining the Dy10 signal peptide-encoding region spliced in frame to the new sequence (Fig. 6A). The regulatory elements provide for high expression specifically in cereal endosperm. High lysine repeats were then inserted into the *NheI* site, resulting in the generation of six additional gliadin derivatives whose identity and orientation were confirmed by sequencing (Fig. 6B –G). The various DNAs were transformed into wheat embryos by particle bombardment along with a second plasmid containing the bialaphos resistance gene under control of a maize ubiquitin promoter. Regenerated plantlets surviving the selection were screened by PCR amplification of genomic DNA using gene-specific primers from the Dy10 and gliadin portions of the construct. (Using untransformed Bobwhite wheat DNA, we verified that no endogenous gene amplifies with this combination of primers.) Successful amplification, indicative of wheat plants containing one or more copies of the integrated gliadin transgene, was obtained for 8 lines each for clones Bp1c and Bt9a and 5 lines for clone Bt15c (Table 2). Furthermore, a homozygote has been identified for one of the Bp1c lines, making it amenable to assay

for production of the modified gliadin. As homozygotes are identified in other lines, their seeds will be examined for modified gliadin expression.

Given the complexity of the storage protein content of wheat seeds, it is important to develop some sort of assay (for example, solubility partitioning, or retention in high pressure liquid chromatography or HPLC), to specifically identify the variant gliadins. For this, we needed a source of pure engineered protein. Isolation of such proteins is facilitated by inclusion of peptide motifs that have affinity for antibodies or other compounds (affinity "tags"). Thus we have initiated moving the gliadin construct coding regions into an *E. coli* expression plasmid (pRSET, Invitrogen, Carlsbad, CA) which, when induced, produces the protein fused to a hexa-histidine moiety that will bind specifically to a Nickel-complexed resin. Elution at a specific pH yields a solution of purified fusion protein from which the unmodified high-lysine gliadin can be recovered by specific proteolytic cleavage by enterokinase at its recognition site, which occurs between the hexa-histidine moiety and the gliadin sequence. The purified protein will then be ready for biochemical characterization based on mobility in HPLC or other physical properties. If enough pure protein can be obtained, we might even be able to study effects of these proteins on bread dough viscoelasticity. So far, one of the gliadin constructs (Bp1c insert) has been cloned into the pRSET fusion plasmid and transformed into an expression host. Expression and analysis of this construct is in progress. Several other gliadin constructs (Bt9a, Bt15c, Br2a) have been amplified with primers for insertion into pRSET, but remain to be transformed.

Objective #2: Increases in endogenous polypeptides high in lysine. A number of naturally occurring wheat seed endosperm proteins high in lysine have been identified (e.g. elongation factors EF-1 α and -1 β , dehydrins, triticins, extensins, and fructose bisphosphate aldolase). Among our EST libraries, we identified a number of cDNA clones with at least partial sequence identity to the EF-1 β sequence in Genbank. The inserts of retrieved clones were analyzed by restriction digestion to determine their length, and sequenced from both 5' and 3' ends to find full length cDNAs and to aid in design of primers for PCR amplification of the insert. One clone was full-length. We amplified this clone using PCR primers containing sequence of the insert corresponding to translational termination and initiation sites adjacent to restriction sites necessary for cloning into the pGlu10(SFe)R853 wheat endosperm-specific expression plasmid described earlier. To facilitate cleavage at the introduced restriction sites, the amplified fragment has been cloned in an intermediate plasmid and remains to be recloned into the wheat expression plasmid. We have also commenced a similar search for the much longer (> 1.6 kb) tritacin gene, but haven't found full-length cDNA clones yet. If we fail to find a full-length clone, one can be assembled by PCR amplification of selected clones from the endosperm cDNA library, a modification of the technique described by Anderson *et al.* (1996) *Gene* 174: 51-58.

Objective #3: Synthesis, processing, and stability of heterologous proteins in cereal endosperm. A number of non-endosperm proteins high in lysine have been identified by searching DNA sequence databases and the literature. Candidate ESTs have been chosen and are awaiting sequencing prior to transfer of full-length sequences into the

endosperm-specific expression plasmid. Experiments for this objective were reduced in priority to allow further progress on other objectives and will be resumed, depending on the results of those experiments.

Objective #4: Modulation of lysine metabolism in cereal endosperm.

This objective was to modulate lysine metabolism in the wheat endosperm and observe the effects on total seed lysine content. We used three approaches: study of the endogenous activity of key lysine catabolic enzymes, the characterization of the DNA sequences encoding these enzymes, and by-passing normal feed-back inhibition for lysine synthesis using feedback insensitive bacterial enzymes expressed in transgenic wheat.

The first approach was to study the expression of the first two enzymes of lysine catabolism, lysine ketoglutarate reductase (LKR) and saccharopine dehydrogenase (SDH), which are linked on a single bifunctional polypeptide. LKR/SDH enzyme has already been identified and characterized in a number of dicot and monocot species, including maize and rice. Although the function of lysine catabolism is still not clear, the presence of such a pathway may interfere with lysine accumulation in the grains. To test whether developing wheat grains also possess LKR and SDH activities, we subjected crude desalted extracts from developing grains to LKR and SDH activity assays. As shown in Fig. 7, the developing grains possessed significant activities of these two enzymes.

Amino acid sequence analyses suggests that the LKR/SDH proteins from different plant species are highly conserved (data not shown). Thus, in order to detect the

LKR/SDH protein in different plant species (including wheat), we prepared monoclonal antibodies against a recombinant *Arabidopsis* LKR/SDH expressed in yeast cells. One powerful monoclonal antibody was identified, which specifically recognizes the LKR domain of LKR/SDH. The ability of this antibody to recognize LKR/SDH protein from different plant species was tested in a Western blot. As shown in Fig. 8, a protein band with the expected size of LKR/SDH was identified in developing *Arabidopsis*, maize and wheat seeds (lanes A-C), but not in wheat leaves (lane D). This suggests that in wheat, expression of the *LKR/SDH* gene is significantly upregulated in developing grains.

Previous publications suggested that the LKR/SDH proteins of plants are homodimers. To test whether the wheat LKR/SDH is also a homodimer, we fractionated crude extract from developing wheat grains on a size column. As shown in Fig. 9, the LKR/SDH holoenzyme was eluted with an estimated size of ~150-160 kDa, which runs in between the expected sizes of a monomer (~116 kDa) and dimer (~230 kDa). It is therefore impossible yet to conclude whether the wheat LKR/SDH is a monomer or a homodimer.

We have previously shown that *Arabidopsis* possesses a composite *LKR/SDH* gene, which is transcribed to yield both bifunctional *LKR/SDH* and monofunctional *SDH* mRNAs. We therefore wished to test whether the wheat LKR/SDH gene is also capable of producing two such mRNAs. To address this, total RNA was extracted from developing wheat grains and hybridized with a probe derived from the SDH domain of the *LKR/SDH* cDNA (see below). As shown in Fig. 10, two hybridizing bands with the expected sizes of LKR/SDH and monofunctional SDH were obtained; suggesting that the wheat *LKR/SDH* gene is also a composite gene encoding both *LKR/SDH* and monofunctional *SDH* mRNAs.

Since no publically available wheat *LKR/SDH* gene had yet been reported, we first tried to obtain coding gene sequences for this enzyme. To search for *LKR/SDH* related sequences from wheat, we interrogated the databases with the amino acid sequence deduced from a previously identified maize mRNA. Several ESTs were identified and their combined derived amino acid sequence was compared to that of the maize *LKR/SDH* protein. As shown in Fig. 11, the wheat ESTs were derived from the C-terminal part of the wheat *LKR/SDH*, in the region encoding the SDH domain. A second, subsequent interrogation of databases using a newly released maize full-length cDNA sequence found two wheat *LKR/SDH* ESTs from the International Triticeae EST Cooperative (ITEC; <http://wheat.pw.usda.gov/genome/>). These ESTs were completely sequenced, and confirmed to be *LKR/SDH* sequences, but again were partial cDNAs mainly from the SDH terminal end of the complete sequence (Fig. 12). Since none of the available cDNA/EST clones contain a full-length *LKR/SDH* sequence, we determined to obtain the remainder of the sequence from genomic clones. Isolation and characterization of the complete gene is also necessary for a thorough study of the control of this pivotal gene.

Although it is known that *Arabidopsis* possesses only one *LKR/SDH* gene, the complexity of the *LKR/SDH* genes in wheat was tested in a Southern blot, using the *SDH* cDNA as a probe. As shown in Fig. 13, a low number of *LKR/SDH*-related DNA bands were revealed in the Southern blots containing DNA digested with several different restriction enzymes. This indicates that wheat also exhibits low complexity of the *LKR/SDH* genes, apparently one gene per diploid genome (total of three genes per hexaploid wheat).

To obtain full-length genes, we initially screened two lambda libraries (one based on a partial *Sau3A* digest and the other a partial *EcoRI* digest of cultivar Cheyenne genomic DNA) using one of these partial cDNAs as a probe, along with two maize *LKR/SDH* cDNA sequences obtained from Drs. Paulo Arruda and Marcelo J. Surpili, (Universidade Estadual de Campinas, São Paulo, Brazil). Fourteen positives were isolated.

During this work, the maize gene for *LKR/SDH* was reported to be 11.1 kbp (containing 24 introns). Since this size might preclude finding a complete wheat gene in a lambda clone if one assumes a similar or larger size for the wheat gene, we decided to screen bacterial-artificial-chromosome (BAC) libraries, which have a much larger insert size (up to 300 kb). Twelve BAC positives were purified from a *Triticum turgidum* library (collaboration with J. Dubcovsky, University of California, Davis, CA), as well as 16 BACs from a barley genomic library (Clemson University Genomics Institute). The purified DNAs from the *T. turgidum* BACs were subjected to fingerprint analysis using the restriction enzymes *HindIII* (Fig 14A) and *EcoRI*. The gel was also blotted and hybridized with the same probe as used for the screening, in order to identify bands containing the *LKR/SDH* coding sequence (Fig 14B). Two different contigs were obtained, presumably corresponding to the two genomes in the *T. turgidum* species (Fig. 15). Two BACs were chosen (1F14 and 6M7), one from each contig, which appeared to contain the most DNA surrounding the fragments hybridizing to the *LKR* gene probes. The DNA was subjected to random shearing and the resulting fragments ligated for production of sublibraries of these two BACs. These clones remain to be sequenced and assembled into contigs spanning the insert. Such genomic sequence information can be used in general studies on the role of the *LKR/SDH* gene in lysine metabolism and other

plant processes. Ultimately, the appropriate sequences from all of these wheat *LKR/SDH* clones will be used in constructing both over-expression and anti-sense gene constructs for transformation into wheat. These latter experiments will provide information on a wider range of gene expressions than normally found and will give a clearer understanding of the roles of the encoded enzyme and its effects both on lysine content and other aspects of cellular metabolism.

The third approach to increasing lysine seed composition is to bypass the normal feedback inhibition on increased lysine content. Our approach was to transform wheat with gene constructs expressing lysine-insensitive bacterial lysine biosynthetic genes. We previously reported isolation of the genes for bacterial dihydropicolinate synthetase (DHPS) and aspartate kinase (AK). The coding regions for these two genes have now been placed in a wheat endosperm specific expression construct (ProCi and ProCk, respectively, Figs. 16A and 16B) and have been transformed into wheat. The transformation of the ProCk construct, encoding AK, has yielded 4 transformed lines, while the transformation of the ProCi construct, encoding DHPS, has yielded one transformed line. PCR was used to identify transformants carrying the *DHPS* and *AK* genes. Once homozygous lines are identified, further study of enzyme activities and lysine composition will be carried out.

EVALUATION OF THE RESEARCH ACHIEVEMENT: The original ambitious nature and breadth of this research has been maintained, and continues to be the development of biotechnological approaches to increasing the amount of lysine in wheat seeds, thus enhancing the nutritional value of cereals for both human and animal

consumption. We have made good progress in objectives 1 and 4, as a number of constructs have been engineered for bacterial expression, and synthesis of the variant proteins in *E. coli* has been achieved. Likewise, several constructs have transformed into wheat. Further work will explore the use of the resources from Objectives 1 and 4 and will address Objectives 2 and 3 in more detail as appropriate.

DESCRIPTION OF THE COOPERATION: The project involves the interaction of the U.S. and Israeli laboratories plus collaborations with other laboratories. The U.S. laboratory has focused on generating synthetic gene constructs for both bacterial and wheat expression, cloning the *LKR/SDH* genes, and wheat transformation. The Israeli laboratory characterized wheat *LKR/SDH* enzymes, and constructed bacterial *DHPS* and *AK* genes for wheat expression in the seed-specific expression plasmid generated in the U.S. laboratory. Both the U.S. and Israel laboratories will utilize the transgenic lines to study the expression of engineered high-lysine proteins and seed lysine composition. Transgenic material will be transferred to the Israel laboratory for biochemical and genetic analysis.

CONCLUSIONS: The project has successfully generated sets of gene constructs for study of the effects of increased prolamine lysine composition, has initiated the generation of transgenic wheat lines with these constructs, has isolated the wheat and barley *LKR/SDH* genes, and has carried out a basic description of *LKR/SDH* enzyme activity in the developing wheat seed. A foundation has been established for future

research into the problem of increasing seed lysine, and a continuing collaboration between the participating laboratories will mine and extend these resources.

LIST OF PUBLICATIONS: No papers have yet been published from this project.

Emphasis has been on resource development, the use of which is anticipated to continue for at least several years. Planned papers directly resulting from this project include the construction of modified prolamine genes and their encoded proteins, the effects of increased lysine on HMW-glutenin functionality, expression of modified prolamine genes in transgenic wheat, structure of the wheat and barley LKR genes, the effects in transgenic wheat of over-expression of feedback-insensitive key enzymes in the lysine synthetic pathway, and LKR/SDH activity in wheat tissues.

Table 1 High molecular weight glutenin Dx5-based high lysine constructs designed for expression in *E. coli*. Non repetitive N terminal portion is either from the synthetic Dx5 or the higher-lysine R2K construct. The identity and number of the higher-lysine repeats are listed as well as the number of Dx5 synthetic 15-mer repeats. Transformation into the *E. coli* BL21 (DE3) expression host is indicated (3 constructs not yet transformed). Also shown are successful small-scale expression experiments and gel locations of the expressed band (gel / lane), where G1 indicates gel 1 and L3 indicates lane 3.

clone	plasmid	N term	high lysine repeat: (identity - number)	Dx5 15-mer repeats (number)	C term	BL21 line	expression	gel/lane (Fig.4)
16a	pET3a	Dx5	none	16	Dx5	yes	yes	G3 L1
32a	pET3a	Dx5	none	32	Dx5	yes	yes	G4 L9
Bt11a	pET3a	Dx5	pep-2	16	Dx5	yes	yes	G1 L7
Bt12d	pET3a	Dx5	pep-2	32	Dx5	yes	yes	G2 L2,3
Bt5c'a	pET3a	Dx5	PQQ-2	16	Dx5	no	not tested	
Bt6d	pET3a	Dx5	PQQ-2	32	Dx5	yes	yes	G4 L7
Br4d	pET3a	Dx5	MEEK-4	16	Dx5	yes	no	G2, L7
Br3a	pET3a	Dx5	MEEK-4	32	Dx5	yes	no	G2 L 5,6; G4 L8
Bq1e	pET3a	Dx5	MEEK-4	none	Dx5	yes	not tested	
H12	pET3a	R2K	none	16	Dx5	yes	yes	G3 L3
A9	pET3a	R2K	none	32	Dx5	yes	yes	G3 L4
Cbn2C7	pET3a	R2K	pep-1	16	Dx5	yes	no	
Cbn2B2	pET3a	R2K	pep-1	32	Dx5	yes	yes	G3 L6
Bt13b	pET3a	R2K	pep-2	16	Dx5	yes	yes	G2 L4
Bt14c	pET3a	R2K	pep-2	32	Dx5	yes	yes	G1 L6
Pa4n	pET3a	R2K	pep-4	16	Dx5	no	not tested	
Pa7l	pET3a	R2K	pep-4	32	Dx5	no	not tested	
Bt7b	pET3a	R2K	PQQ-2	16	Dx5	yes	yes	G1 L1,2
Bt8b, 8c	pET3a	R2K	PQQ-2	32	Dx5	yes	yes	G4 L6
Cbn2C6	pET3a	R2K	MEEK-1	16	Dx5	yes	no	G4, L3
Cbn1F2	pET3a	R2K	MEEK-1	32	Dx5	yes	no	G4, L2

Table 2. Alpha-gliadin based high lysine constructs that were cloned into the wheat endosperm expression vector pGlu10(SFe)R853. "Yes" in the wheat transformation column indicates that transformation experiments were attempted with this construct.

clone	gliadin	high lysine repeat	wheat transformations	positive lines
Bp1c	gliK	none	yes	8, 1 homozygous
Bt15c	gliK	pep-2	yes	5
Bt9a	gliK	PQQ-2	yes	8
Bw2f	gliK	PQQ-4	no	
Bu2b	gliK	pep-2--PQQ-2	yes	0
Bu7b	gliK	PQQ-4--pep-2	yes	0
Br2a	gliK	MEEK-4	yes	0

	<i>NdeI</i>														<i>SacI</i>	
Dx5R2K	H	M	E	G	E	A	S	E	Q	L	Q	C	E	K	E	L
Dx5R2K	<u>CAT</u>	<u>ATG</u>	GAA	GGT	GAA	GCA	TCT	GAA	CAG	CTG	CAG	TGC	GAA	AAA	<u>GAG</u>	<u>CTC</u>
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>HindIII</i>															
Dx5R2K	Q	E	L	Q	E	K	E	L	K	A	C	Q	Q	V	M	D
Dx5R2K	CAG	GAA	CTG	CAG	GAA	AAA	GAA	CTG	<u>AAA</u>	<u>GCT</u>	TGC	CAG	CAG	GTT	ATG	GAC
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>BspEI</i>															
Dx5R2K	Q	Q	L	K	D	I	S	P	E	C	H	P	V	V	V	S
Dx5R2K	CAG	CAG	CTG	AAA	GAC	ATC	<u>TCT</u>	<u>CCG</u>	GAA	TGC	CAC	CCG	GTT	GTT	GTT	TCT
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>MscI</i>															
Dx5R2K	P	V	A	G	Q	Y	E	Q	Q	I	V	V	P	P	K	G
Dx5R2K	CCG	GTT	<u>GCT</u>	<u>GGC</u>	<u>CAG</u>	TAC	GAA	CAG	CAG	ATC	GTT	GTT	CCG	CCG	AAA	GGT
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>KpnI</i> <i>MfeI</i>															
Dx5R2K	G	T	F	Y	P	G	E	T	T	P	P	Q	Q	L	Q	Q
Dx5R2K	<u>GGT</u>	<u>ACC</u>	TTC	TAT	CCG	GGT	GAA	ACC	ACC	CCG	CCG	CAG	<u>CAA</u>	<u>TTG</u>	CAG	CAG
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>EcoRI</i> <i>ScaI</i>															
Dx5R2K	K	I	F	W	G	I	P	A	L	L	K	K	Y	Y	P	S
Dx5R2K	AAA	ATC	TTC	TGG	<u>GGA</u>	<u>ATT</u>	CCG	GCT	CTG	CTG	AAA	<u>AAG</u>	<u>TAC</u>	<u>TAC</u>	CCG	TCT
Dx5	CGT	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>XmaI</i>															
Dx5R2K	V	T	C	P	Q	Q	V	S	Y	Y	P	G	Q	A	S	P
Dx5R2K	GTT	ACC	TGC	CCG	CAG	CAG	GTT	TCT	TAC	<u>TAC</u>	<u>CCG</u>	<u>GGT</u>	CAG	GCT	TCT	CCG
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>XbaI</i> <i>EcoRV</i> <i>SpeI</i>															
Dx5R2K	Q	K	S	R	D	I	T	S	S	S	Y	H	V	S	V	E
Dx5R2K	CAG	AAA	<u>TCT</u>	<u>AGA</u>	<u>GAT</u>	<u>ATC</u>	<u>ACT</u>	<u>AGT</u>	TCT	TCT	TAC	CAC	GTT	TCT	GTT	GAA
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>NheI</i> <i>FspI</i>															
Dx5R2K	H	Q	A	A	S	L	K	V	A	K	A	Q	Q	L	A	A
Dx5R2K	CAC	CAC	GCT	<u>CCT</u>	<u>AGC</u>	CTG	AJA	GTT	GCT	AAA	GCT	CAG	CAG	CTG	<u>GCT</u>	<u>CCG</u>
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>NaeI</i>															
Dx5R2K	Q	L	P	A	M	P	K	L	E	G	G	D	A	L	S	A
Dx5R2K	<u>CAG</u>	CTG	<u>CCG</u>	<u>GCT</u>	ATG	CCG	AAA	CTG	GAA	GGT	GGT	GAC	GCT	CTG	TCT	GCT
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	<i>BamHI</i>															
Dx5R2K	S	Q	ter	ter												
Dx5R2K	TCT	CAG	TGA	TAG	<u>GGA</u>	<u>TCC</u>										
Dx5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Fig. 1 Comparison of sequences of previously engineered Dx5 fused N- and C-terminal nonrepetitive domains (Anderson et al (1996) Gene 174: 51-58) and the new high lysine construct, Dx5R2K. Differences reflect not only changes from arginine to lysine codons, but changes introduced to produce unique restriction sites spaced throughout the gene, and to reflect codon frequency of highly expressed genes in *E. coli*. Nucleic acid sequences are presented along with associated translated amino acid sequence of the high-lysine construct. Identical nucleic acid sequence is represented by dots in the Dx5 sequence. Restriction enzyme sites are underlined and indicated above the sequences.

Fig. 2A "PEPMPK" Repeat

```

A S P E P M P K P E P K P E P K P E
GCT AGC CCA GAG CCG ATG CCA AAA CCG GAA CCA AAG CCG GAG CCA AAA CCG GAA
NheI

P M P K P E P K P E P K P E P M P K
CCA ATG CCG AAG CCA GAG CCG AAA CCA GAA CCG AAG CCA GAG CCG ATG CCA AAA

P E P K P E P K P E P M P K P E P K
CCG GAA CCA AAG CCG GAG CCA AAA CCG GAA CCA ATG CCG AAG CCA GAG CCG AAA

P E P K L T S
CCA GAA CCG AAG CTT ACT AGT
HindIII SpeI

```

Fig. 2B "MEEK" Repeat

```

A S M E E K L K A M E E K L K A M E
GCT AGC ATG GAA GAG AAA CTG AAG GCT ATG GAG GAA AAG CTG AAA GCT ATG GAA
NheI

E K M K A M E E K L K A M E E K L K
GAG AAA ATG AAA GCT ATG GAA GAA AAA CTG AAG GCT ATG GAG GAG AAA CTG AAG

A M E E K M K A M E E K L K A M E E
GCT ATG GAG GAA AAG ATG AAA GCT ATG GAG GAA AAA CTG AAG GCT ATG GAA GAG

K L K A T S
AAG CTT AAG GCT ACT AGT
HindIII SpeI

```

Fig. 2C "PQQ" Repeat

```

P Q Q P G Q G Q Q G Y Y P P Q Q
A S P Q Q P G Q E Q K G Y Y P P Q Q
GCT AGC CCG CAG CAG CCG GGT CAG GAA CAG AAA GGT TAC TAC CCG CCG CAG CAG
NheI

P G Q G Q Q G Y Y P P Q Q P G Q G Q
P G Q E Q K G Y Y P P Q Q P G Q E Q
CCG GGT CAG GAA CAG AAA GGT TAC TAC CCG CCG CAG CAG CCG GGT CAG GAA CAG

Q G Y Y P P Q Q P G Q G Q Q G Y Y P
K G Y Y P P Q Q P G Q E Q K G Y Y P
AAA GGT TAC TAC CCG CCG CAG CAG CCG GGT CAG GAA CAG AAA GGG TAC TAC CCG
RsaI

T S
ACT AGT
SpeI

```

Fig. 2 Sequences of three engineered high-lysine repeat sequences designed for insertion into the repeat regions of Dx5, Dx5R2K, or gliaK gene constructs. Each sequence starts with a *NheI* site, and ends with an in-frame *SpeI* site, so that the inserts can be cloned in the *XbaI-SpeI* region of the Dx5 genes and the *NheI-SpeI* region of the gliaK gene. They can also be oligomerized by the scheme presented in Fig. 3. DNA sequence is shown along with the translated amino acid sequence and relevant restriction sites.

(A) "PEP" (based on Raines et al, (1991) *Plant Molecular Biology* 16: 663-670).

(B) "MEEK" (based on Keeler et al (1997) *Plant Molecular Biology* 34: 15-29).

(C) "PQQ" is a tetramer of the original synthetic Dx5 repeat sequence (top line) of Anderson et al (1996) *Gene* 174: 51-58, modified by addition of 1 lysine and 1 glutamate residue (so that the original electrostatic charge is maintained) per repeat. The modified amino acid and DNA sequences are on the second and third lines, respectively.

Fig. 3 Generation of oligomers. The high-lysine repeat insert sequences (Fig. 2) and their orientations with respect to restriction sites on the plasmid are indicated by black arrows. The plasmid vector is pCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA). The locations of relevant restriction enzyme sites used in the generation of the oligomers are shown. Panel (A) illustrates a successful oligomerization strategy, whereby a plasmid containing a single MEEK repeat (Fig. 2) in the clockwise orientation is digested with both *Dra*III and either *Spe*I or *Nhe*I. The repeat-containing fragments are purified and then directionally ligated together to produce a plasmid with two tandem copies of the starting sequence, in parallel orientation and in reading frame. The restriction sites between the two copies are destroyed by the ligation, and so are indicated by (*Spe*I/*Nhe*I). This process can then be repeated to generate plasmids containing larger multiples of the repeat sequence. Panel (B) illustrates the strategy when applied to plasmids containing single repeats of PEP or PQQ (Fig. 2) in the counterclockwise orientation. The ligation is non-productive because the two *Spe*I sites flank the insert. To generate oligomers of these plasmids, the vector *Spe*I site that is not part of the repeat sequence was removed.

Figure 3
A

B

Fig 4 Expression of high lysine glutenin constructs. Constructs (see Table I) were transformed into the *E. coli* expression line BL21 (DE3), and individual clones were grown and induced. Cell pellets were sonicated in the presence of 50 mM Tris HCl, pH 8.0, 50 mM dithiothreitol, 50% normal propanol. The supernatant was dried and then suspended in gel loading dye (2% SDS, 10% glycerol, 50 mM dithiothreitol, 62.5 mM Tris HCl, pH 8.5, 0.006% pyronin Y) and electrophoresed on NuPAGE 10% BisTris gels using the MES buffer system (Invitrogen, Carlsbad, CA). M indicates molecular weight markers with sizes in kDa shown to the right. After electrophoresis, gels were stained with 0.25% Coomassie Brilliant Blue R-250 in 50% methanol/ 10% acetic acid and destained in 5% methanol/5% acetic acid before photography. Extracts are from the following constructs (lanes are in parentheses): A, Gel 1: (1 and 2) Bt7b; (3 and 4) Bt8b, (5) Bt6d, (6) Bt14c, (7) Bt11a. B, Gel 2: (1) pET3a-Dx5 termini only control, (2 and 3) Bt12d, (4) Bt13b, (5 and 6) Br3a, (7) Br4d. C, Gel 3: (1) 16a, (2) 32a, (3) H12, (4) A9, (5) Cbn1F2, (6) Cbn2B2, (7) Cbn2C6, (8) Cbn2C7. D, Gel 4: (1) pET3a- Dx5 termini control, (2) Cbn1F2, (3) Cbn2C6, (4) Cbn2C7, (5) Bt8b, (6) Bt8c, (7) Bt6d, (8) Br3a, (9) 32a

4A Gel 1

Fig 4 B Gel 2

Fig 4 C Gel 3

Fig 4 D Gel 4

pA42 ATG AAG ACC TTT CTC ATC CTT GCC CTC CTT GCT ATC GTG GCA ACC ACC GCC
 pA42 M K T F L I L A L L A I V A T T A

StuI AgeI XbaI
 gliK A S V K V P V P Q L Q P K N S
 gliK AG GCC TCT GTG AAG GTA CCG GTG CCA CAA CTG CAA CCA AAG AAC TCT
 pA42 ACA A.T G.A .A .GA .T .AT . .G . .C.A .T ---
 pA42 T T A . R Q . -

gliK R P S Q Q Q P Q K Q V P L V Q Q
 gliK AGA CCA TCT CAA CAA CAA CCA CAA AAG CAA GTG CCA CTG GTG CAA CAA
 pA42 ---GGGTT . .A
 pA42 E

XmaI BciVI
 gliK Q Q F P G Q Q Q Q Q F P P Q Q P Y
 gliK CAA CAA TTT CCC GGG CAA CAA CAA CAA TTT CCA CCA CAA CAA CCG TAT
 pA42AGG .A .T
 pA42 .

PstI
 gliK P K P Q P F P S Q Q P Y L Q L Q
 gliK CCA AAG CCA CAA CCA TTT CCA TCT CAA CAA CCA TAC CTG CAG CTG CAA
 pA42 .G C . .G . . .TAT
 pA42 . Q .

gliK P F P Q P Q - - - - - L P Y P Q
 gliK CCA TTT CCA CAA CCA CAA - - - - - CTG CCA TAC CCA CAA
 pA42G .G .G CCA TTT CCG CCA CAA .A . . .T .G .G
 pA42 P F P P Q

HindIII
 gliK P Q S F K P Q Q P Y P Q Q K P K
 gliK CCA CAA AGC TTT AAG CCA CAA CAA CCA TAC CCA CAA CAA AAG CCA AAG
 pA42 .G . . . TCA . . . CCATG C.A . . . C..
 pA42 P Q . . . Q

NheI EcoRV SpeI
 gliK Y L Q P Q Q P I A S D I T S S Q
 gliK TAC CTG CAA CCA CAA CAA CCA ATA GCT AGC GAT ATC ACT AGT TCT CAA
 pA42 .T .ATGATG .G
 pA42 .

gliK Q Q A Q Q Q Q Q Q Q Q Q Q Q Q Q
 gliK CAA CAA GCA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA CAA
 pA42 .
 pA42 .

ApoI
 gliK Q Q I L Q Q I L Q Q Q L I P C K
 gliK CAA CAA ATT TTG CAA CAA ATA CTG CAA CAA CAA CTG ATA CCA TGC AAG
 pA42C.TT TTG.
 pA42 . R

AatII ScaI
 gliK D V V L Q Q H N K A K A K S Q V
 gliK GAC GTC GTG CTG CAA CAA CAC AAC AAG GCA AAG GCA AAG TCT CAA GTA
 pA42 .T .T .C TTA .G C.T . . .GC .A . . .T
 pA42 I . H . S

gliK L Q Q S T Y Q L L K Q L C C Q K
 gliK CTG CAA CAA TCT ACT TAC CAA CTG CTG AAG CAA CTG TGC TGC CAA AAG
 pA42 TAGG .A T . .C.A . . .T . .T .T . .C.A
 pA42 Q Q

HinfI
 gliK L L Q I P E K S K C Q A I H K V
 gliK CTG CTA CAG ATT CCA GAG AAG TCT AAG TGC CAA GCA ATT CAC AAG GTG
 pA42 . . T.G . . .C .TG CC . .C .T .T .T
 pA42 Q N .

Fig. 6 A

Fig. 6 Plasmid maps of gliadin constructs prepared for wheat transformation (see also Table 2). Open arrows denote the Dy10 upstream region and Dx5 downstream region. Filled arrows indicate sequence from the gliK high lysine construct, and the horizontally hatched arrow indicates Bluescript sequences. A cross-hatched arrow corresponds to the MEEK repeat sequence, a vertical hatched arrow corresponds to the PQQ repeat sequence, and a thatched arrow corresponds to the PEP repeat sequence. Pertinent restriction sites are indicated, as are the translational initiation and termination sites. (A) Bp1c, (B) Br2a, (C) Bt9a, (D) Bw2f, (E) Bt15c, (F) Bu2b, (G) Bu7b

Fig. 6B

Fig. 6C

Fig. 6D

Fig. 6E

Fig. 6F

Fig. 6G

Fig. 7 LKR and SDH activities in developing wheat grains. (A) LKR activity; (B) SDH activity. Con, control reactions lacking lysine for LKR activity (A) or lacking saccharopine for SDH activity (B). Bars on top of the histograms represent standard errors.

Fig. 8 Western blot analysis of water soluble extracts from developing maize grains (A), *Arabidopsis* floral organs (B), developing wheat grains (C) and wheat leaves (D) with the anti *Arabidopsis* LKR monoclonal antibodies. The position of the LKR/SDH bands near the 116 kDa size marker is indicated.

Fig. 9 Estimation of the size of the wheat LKR/SDH holoenzyme. Crude extract from developing wheat endosperm was fractionated on a size column in parallel with size protein markers. Individual fractions were reacted in a Western blot with the anti *Arabidopsis* LKR monoclonal antibodies. The fraction containing the 160 kDa size marker is indicated on top. Only fractions 15 to 28 of the column (left to right) are indicated.

Developing wheat
grains RNA

Fig. 10 Northern blot analysis of total RNA from developing wheat grains with a probe derived from the *SDH* domain of a wheat *LKR/SDH* cDNA (Fig. 11). The positions of the bifunctional *LKR/SDH* and monofunctional *SDH* mRNAs are indicated on the right.

```

701 SQAAKDAGVTILCEMGLDPGIDHLSMKMIDEAHARKGKIKAFSTSYCGGL 750
      .  || .  ||
1  .....EFGTSGTSGL 10

751 PSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAK 800
      ||||| ||||| ||||| : ||||| : ||||| ||||| |||||
11 PSPAAANNPLAYKFSWSPAGAIRAGRNPAVYKFLGETINVDGSKLYESAK 60

801 RLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATLRYEGFSEIM 850
      ||||| ||||| ||||| ||||| : ||||| : ||||| |||||
61 RLRLPELPAFALEHLPNRNSL MYGDLYGISKEASTVYRSTLRYEGFSEIM 110

851 VTLSKTGFFDAANHPLLQDTSRPTYKGFLDELL . NNISTINTDLDEAS 898
      |. | ||||| ||||| : ||||| : ||||| ||||| |||||
111 AILAKIGFFDAENHPLLQETNRPTYRIFLNELLNVNNVSTSN TKVNGEET 160

899 GGYDDDLIARLLKLGCKKNKEIAVKT V KTIKFLGLHEETQIPKGCSSPFD 948
      ||: ||: ||: ||: || || ||||| ||||| ||||| |||||
161 GGHDELISKLMMLGHCKEKE LAVKILKTIKFLGLHEETQIPKDCSSAFS 210

949 VICQRMQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVE 998
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 VICQRMQRMAYGHNEQDMVLLHHEVEVEYPDGRPTKHKQTTLLEFGKTE 260

999 NGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILE 1048
      ||||| ||||| ||||| : ||||| : ||||| : ||||| : |||||
261 NGRSTTAMALTVGVPAAIGALLLLQNKVQRKGVIRPLEPEIYIPALEILE 310

1049 SSGIKLVEKVETZ..... 1061
      .. ||||| : |||||
311 TAGIKLIEKVETZEDVRMGZEST EYICCCNRGSEZIKZZFSPLZVKZANC 360

```

Fig. 11 Amino acid sequence comparison of the maize (top rows) and the wheat (bottom rows) LKR/SDH sequences. The numbers on the right and left represent the number of amino acid starting from the N-terminal beginning of the maize LKR/SDH and the partial SDH sequence of the wheat LKR/SDH. Vertical lines indicate identity and double dots indicate conservative changes.

m	M	G	S	A	A	T	E	G	N	D	T	L	L	G	N	G	V	51
	ATG	GGT	TCT	GCT	GCT	ACT	GAG	GGC	AAT	GAC	ACC	TTG	CTG	GGC	AAT	GGA	GTT	
m	V	G	I	L	A	E	T	C	N	M	W	E	R	R	A	P	L	102
	GTT	GGG	ATT	CTT	GCT	GAG	ACT	TGT	AAT	ATG	TGG	GAA	AGA	AGG	GCG	CCG	TTA	
m	T	P	S	H	C	A	R	L	L	L	G	G	G	K	N	G	P	153
	ACT	CCT	TCC	CAT	TGT	GCC	CGC	CTT	CTG	CTA	GGA	GGA	GGC	AAG	AAC	GGA	CCT	
m	R	V	N	R	I	I	V	Q	P	S	T	R	R	I	H	H	D	204
	CGA	GTA	AAC	CGG	ATT	ATT	GTG	CAG	CCA	AGC	ACA	AGG	AGG	ATC	CAT	CAT	GAC	
m	A	Q	Y	E	D	A	G	C	E	I	S	E	D	L	S	E	C	255
	GCT	CAG	TAT	GAG	GAT	GCA	GGA	TGC	GAG	ATT	TCA	GAA	GAC	CTG	TCA	GAA	TGC	
m	G	L	I	I	G	I	K	Q	P	K	L	Q	M	I	L	S	D	306
	GGC	CTT	ATC	ATA	GGC	ATC	AAA	CAA	CCC	AAG	CTG	CAG	ATG	ATT	CTT	TCA	GAT	
m	R	A	Y	A	F	F	S	H	T	H	K	A	Q	K	E	N	M	357
	AGA	GCG	TAC	GCT	TTC	TTT	TCA	CAC	ACA	CAC	AAA	GCC	CAA	AAA	GAG	AAT	ATG	
m	P	L	L	D	K	I	L	E	E	R	V	S	L	F	D	Y	E	408
	CCA	CTG	TTA	GAC	AAG	ATC	CTT	GAA	GAA	AGG	GTG	TCC	TTG	TTT	GAT	TAT	GAG	
m	L	I	V	G	D	D	G	K	R	S	L	A	F	G	K	F	A	459
	CTA	ATT	GTT	GGA	GAT	GAT	GGG	AAA	AGA	TCA	CTA	GCA	TTT	GGG	AAA	TTT	GCT	
m	G	R	A	G	L	I	D	F	L	H	G	L	G	Q	R	Y	L	510
	GGT	AGA	GCT	GGA	CTG	ATA	GAT	TTC	TTA	CAT	GGT	CTC	GGA	CAG	CGA	TAT	TTG	
m	S	L	G	Y	S	T	P	F	L	S	L	G	Q	S	H	M	Y	561
	AGC	CTT	GGA	TAC	TCG	ACT	CCA	TTT	CTC	TCT	CTG	GGA	CAA	TCT	CAT	ATG	TAT	
m	P	S	A	A	A	K	A	A	V	I	V	V	A	E	E	I		612
	CCT	TCG	CTC	GCT	GCA	GCC	AAG	GCT	GCA	GTC	ATT	GTC	GTT	GCA	GAA	GAG	ATA	
m	A	T	F	G	L	P	S	G	I	C	P	I	V	F	V	F	T	663
	GCA	ACA	TTT	GGA	CTT	CCA	TCC	GGA	ATT	TGT	CCG	ATA	GTG	TTT	GTG	TTC	ACT	
m	G	V	G	N	V	S	Q	G	A	Q	E	I	F	K	L	L	P	714
	GGA	GTT	GGA	AAC	GTC	TCT	CAG	GGT	GCG	CAG	GAG	ATA	TTC	AAG	TTA	TTG	CCC	
m	H	T	F	V	D	A	E	K	L	P	E	I	F	Q	A	R	N	765
	CAT	ACC	TTT	GTT	GAT	GCT	GAG	AAG	CTT	CCC	GAA	ATT	TTT	CAG	GCC	AGG	AAT	
m	L	S	K	Q	S	Q	S	T	K	R	V	F	Q	L	Y	G	C	816
	CTG	TCT	AAG	CAA	TCT	CAG	TCG	ACC	AAG	AGA	GTA	TTT	CAA	CTT	TAT	GGT	TGT	
m	V	V	T	S	R	D	M	V	S	H	K	D	P	T	R	Q	F	867
	GTT	GTG	ACC	TCT	AGA	GAC	ATG	GTT	TCT	CAC	AAG	GAT	CCC	ACC	AGA	CAA	TTT	
m	D	K	G	D	Y	Y	A	H	P	E	H	Y	T	P	V	F	H	918
	GAC	AAA	GGT	GAC	TAT	TAT	GCT	CAT	CCA	GAA	CAC	TAC	ACC	CCT	GTT	TTT	CAT	

m	E R I A P Y A S V I V N C M Y W E	969
	GAA AGA ATT GCT CCA TAT GCA TCT GTC ATC GTA AAC TGT ATG TAT TGG GAG	
m	K R F P P L L N M D Q L Q Q L M E	1020
	AAG AGG TTT CCA CCA TTA CTA AAT ATG GAT CAG TTA CAG CAA TTG ATG GAG	
m	T G C P L V G V C D I T C D I G G	1071
	ACT GGT TGT CCT TTA GTC GGC GTT TGT GAC ATA ACT TGT GAT ATT GGA GGT	
m	S I E F I N K S T S I E R P F F R	1122
	TCC ATT GAA TTT ATC AAC AAG AGT ACA TCA ATA GAG AGG CCT TTC TTT CGG	
m	Y D P S K N S Y H D D M E G A G V	1173
	TAT GAT CCT TCT AAG AAT TCA TAC CAT GAT GAT ATG GAA GGT GCC GGA GTG	
m	V C L A V D I L P T E F S K E A S	1224
	GTC TGC TTG GCT GTT GAC ATT CTC CCT ACA GAA TTC TCT AAA GAG GCC TCC	
m	Q H F G N I L S R L V A S L A S V	1275
	CAA CAT TTT GGA AAC ATA CTA TCT AGA CTT GTT GCT AGT TTG GCC TCA GTG	
m	K Q P A E L P S Y L R R A C I A H	1326
	AAG CAA CCG GCA GAA CTT CCT TCC TAC TTG AGA AGA GCT TGC ATT GCA CAT	
m	A G R L T P L Y E Y I P R M R N T	1377
	GCT GGC AGA TTA ACT CCT TTG TAT GAA TAT ATC CCT AGG ATG AGA AAT ACT	
m	M I D L A P A K T N P L P D K K Y	1428
	ATG ATA GAT TTG GCA CCC GCA AAA ACA AAT CCA TTG CCT GAC AAG AAG TAT	
m	S T L V S L S G H L F D K F L I N	1479
	AGC ACC CTG GTA TCT CTC AGT GGG CAC CTA TTT GAT AAG TTC CTT ATA AAT	
m	E A L D I I E T A G G S F H L V R	1530
	GAA GCT TTG GAC ATC ATT GAG ACA CTT GGA GGT TCA TTT CAC TTG GTT AGA	
m	C E V G Q S T D D M S Y S E L E V	1581
	TGT GAA GTT GGA CAA AGC ACG GAT GAT ATG TCA TAC TCA GAG CTT GAA GTA	
m	G A D D T A T L D K I I D S L T S	1632
	GGA GCA GAT GAT ACT GCC ACA TTG GAT AAA ATT ATT GAT TCC TTG ACT TCT	
m	L A N E H G G D H D A G Q E I E L	1683
	TTA GCT AAT GAA CAT GGT GGA GAT CAC GAT GCC GGG CAA GAA ATT GAA TTA	
m	A L K I G K V N E Y E T D V T I D	1734
	GCT CTG AAG ATA GGA AAA GTC AAT GAG TAT GAA ACT GAC GTC ACA ATT GAT	
m	K G G P K I L I L G A G R V C R P	1785
	AAA GGA GGG CCA AAG ATT TTA ATT CTT GGA GCT GGA AGA GTC TGT CGG CCA	
m	A A E F L A S Y P D I C T Y G V D	1836
	GCT GCT GAG TTT CTG GCA TCT TAC CCA GAC ATA TGT ACC TAT GGT GTT GAT	

m	D	H	D	A	D	Q	I	H	V	I	V	A	S	L	Y	Q	K	
w1	GAC	CAT	GAT	GCA	GAT	CAA	ATT	CAT	GTT	ATC	GTG	GCA	TCT	TTG	TAT	CAA	AAA	1887
										...	...	...	...	...	...	...	...	
											V							
m	D	A	E	E	T	V	D	G	I	E	N	T	T	A	T	Q	L	
w1	GAT	GCA	GAA	GAG	ACA	GTT	GAT	GGT	ATT	GAA	AAT	ACA	ACT	GCT	ACC	CAG	CTT	1938
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
										K		A			A		C	
m	D	V	A	D	I	G	S	L	S	D	L	V	S	Q	V	E	V	
w1	GAT	GTT	GCT	GAT	ATT	GGA	AGC	CTT	TCA	GAT	CTT	GTT	TCT	CAG	GTT	GAG	GTT	1989
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	V	I	S	L	L	P	A	S	F	H	A	A	I	A	G	V	C	
w1	GTA	ATT	AGC	TTG	CTG	CCT	GCT	AGT	TTT	CAT	GCT	GCC	ATT	GCA	GGA	GTA	TGC	2040
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	I	E	L	K	K	H	M	V	T	A	S	Y	V	D	E	S	M	
w1	ATA	GAG	TTG	AAG	AAG	CAC	ATG	GTA	ACG	GCA	AGC	TAT	GTT	GAT	GAA	TCC	ATG	2091
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	S	N	L	S	Q	A	A	K	D	A	G	V	T	I	L	C	E	
w1	TCA	AAC	TTG	AGC	CAA	GCT	GCC	AAA	GAT	GCA	GGT	GTA	ACT	ATA	CTT	TGT	GAA	2142
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	M	G	L	D	P	G	I	D	H	L	M	S	M	K	M	I	D	
w1	ATG	GGC	CTA	GAT	CCT	GGC	ATA	GAT	CAC	TTG	ATG	TCA	ATG	AAG	ATG	ATT	GAT	2193
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	E	A	H	A	R	K	G	K	I	K	A	F	T	S	Y	C	G	
w1	GAA	GCT	CAT	GCA	CGA	AAG	GGA	AAA	ATA	AAG	GCA	TTT	ACA	TCT	TAC	TGT	GGT	2244
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	G	L	P	S	P	A	A	A	N	N	P	L	A	Y	K	F	S	
w1	GGA	TTG	CCA	TCT	CCA	GCT	GCA	GCA	AAC	AAT	CCG	CTT	GCC	TAT	AAA	TTC	AGT	2295
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
m	W	N	P	A	G	A	L	R	S	G	K	N	P	A	V	Y	K	
w1	TGG	AAC	CCA	GCT	GGT	GCA	CTC	CGG	TCA	GGG	AAA	AAT	CCT	GCA	GTC	TAC	AAA	2346
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	

	F	L	G	E	T	I	H	V	D	G	H	N	L	Y	E	S	A	
m	TTT	CTT	GGT	GAG	ACG	ATC	CAT	GTA	GAT	GGT	CAT	AAC	TTG	TAT	GAA	TCA	GCA	2397
w1	...	...	..A	...	.TC	...	A..	..G	..C	...	AG.	..A	..A	..C	..G	...	...	
					I		N				S	K						
	K	R	L	R	L	R	E	L	P	A	F	A	L	E	H	L	P	
m	AAG	AGG	CTC	AGA	CTA	CGA	GAG	CTT	CCA	GCT	TTT	GCT	CTG	GAA	CAC	TTG	CCA	2448
w1	...	...	...	...	T..	.CG	..A	...	...	...	...	...	...	...	...	..A	...	
						P												
	N	R	N	S	L	I	Y	G	D	L	Y	G	I	S	K	E	A	
m	AAT	CGG	AAT	TCC	TTG	ATA	TAT	GGT	GAC	CTT	TAT	GGT	ATC	TCC	AAA	GAA	GCA	2499
w1	...	..A	...	...	...	..G	...	..A	...	..G	...	..G	...	...	...	...	...	
						M												
	S	T	I	Y	R	A	T	L	R	Y	E	G	F	S	E	I	M	
m	TCC	ACC	ATA	TAT	AGG	GCT	ACT	CTT	CGT	TAC	GAA	GGT	TTT	AGT	GAG	ATT	ATG	2550
w1	..T	...	G..	...	...	TCC	...	...	...	...	...	..A	...	...	...	..C	...	
			V			S												
	V	T	L	S	K	T	G	F	F	D	A	A	N	H	P	L	L	
m	GTA	ACC	CTT	TCC	AAA	ACT	GGG	TTC	TTT	GAT	GCT	GCA	AAT	CAT	CCA	CTG	CTG	2601
w1	.CT	.T.	..G	G.A	...	.T.	...	..T	...	...	...	..A	...	...	...	...	..A	
	A	I		A		I						E						
	Q	D	T	S	R	P	T	Y	K	G	F	L	D	E	L	L	N	
m	CAA	GAT	ACT	AGT	CGT	CCA	ACA	TAT	AAG	GGT	TTC	CTT	GAT	GAA	CTA	CTG	AAT	2652
w1	...	..A	...	.A.	...	...	...	...	.G.	AA.	..T	..C	A..	...	..C	..T	...	
		E		N					R	N			N					
	-	-	N	I	S	T	I	N	T	D	L	D	I	E	A	S	G	
m	---	---	AAT	ATC	TCC	ACA	ATT	AAC	ACG	GAC	TTA	GAT	ATT	GAA	GCC	TCT	GGT	2697
'	GTC	AAT	...	...	...	...	TC.	...	..A	A.G	A..	A.C	GG.	...	.AG	...	..A	
	V	N					S			K	I	N	G		E			
	G	Y	D	D	D	L	I	A	R	L	L	K	L	G	C	C	K	
m	GGA	TAC	GAT	GAT	GAC	CTG	ATT	GCC	AGA	CTG	TTG	AAG	CTC	GGG	TGT	TGC	AAA	2748
w1	...	C.T	...	...	..A	...	G..	T.A	...	..C	A..	.T.	..T	...	CA.	...	...	
		H			E		V	S			M	M			H			
	N	K	E	I	A	V	K	T	V	K	T	I	K	F	L	G	L	
m	AAT	AAG	GAA	ATA	GCT	GTT	AAG	ACA	GTC	AAA	ACC	ATC	AAG	TTC	TTG	GGA	CTA	2799
w1	G.A	...	...	C..	...	..C	...	.T.	C..	...	..A	...	...	...	...	..G	..G	
w2	...	...	...	C..	...	..C	...	.T.	C..	...	..A	...	...	...	...	..G	..G	
	E			L				I	L									
	H	E	E	T	Q	I	P	K	G	C	S	S	P	F	D	V	I	
m	CAT	GAA	GAG	ACT	CAA	ATA	CCT	AAG	GGT	TGT	TCG	AGC	CCA	TTT	GAT	GTG	ATT	2850
w1	...	..G	...	..A	..G	..T	...	...	.A.	...	..A	..T	G..	..C	AG.	..T	...	
w2	...	..G	...	..A	..G	..T	...	...	.A.	...	..A	..T	G..	..C	AG.	..T	...	
									D				A		S			

	C	Q	R	M	E	Q	R	M	A	Y	G	H	N	E	Q	D	M	
m	TGC	CAG	CGA	ATG	GAA	CAG	AGG	ATG	GCC	TAT	GGC	CAC	AAT	GAG	CAA	GAC	ATG	2901
w1	...	..A	...	...	...	...	..A	...	...	...	...	...	...	...	..G	..T	...	
w2	...	..A	...	...	...	...	..A	...	...	...	...	...	...	...	..G	..T	...	

	V	L	L	H	H	E	V	E	V	E	Y	P	D	G	Q	P	A	
m	GTA	CTG	CTC	CAC	CAC	GAA	GTC	GAG	GTG	GAA	TAC	CCG	GAC	GGG	CAA	CCC	GCC	2952
w1	...	...	..G	...	..T	...	...	...	...	..G	...	..C	..T	...	..G	...	A..	
w2	...	...	..G	...	..T	...	...	...	...	..G	...	..C	..T	...	..G	...	A..	

R

	E	K	H	Q	A	T	L	L	E	F	G	K	V	E	N	G	R	
m	GAA	AAG	CAC	CAA	GCG	ACG	CTA	CTG	GAG	TTC	GGG	AAG	GTT	GAA	AAT	GGC	AGG	3003
w1	..G	...	...	...	...	...	..G	A..	...	...	..A	...	AC.	..G	..C	...	...	
w2	..G	...	...	...	...	...	..G	A..	...	...	..A	...	AC.	..G	..C	...	...	

M T

	S	T	T	A	M	A	L	T	V	G	I	P	A	A	I	G	A	
m	TCC	ACC	ACT	GCC	ATG	GCG	CTG	ACC	GTC	GGC	ATT	CCA	GCA	GCA	ATA	GGG	GCC	3054
w1	..G	...	..C	...	...	..C	..C	...	...	..G	G.A	..G	...	..G	...	..A	...	
w2	..G	...	..C	...	...	..C	..C	...	...	..G	G.A	..G	...	..G	...	..A	...	

V

	L	L	L	L	K	N	K	V	Q	T	K	G	V	I	R	P	L	
m	CTG	CTA	TTG	CTA	AAG	AAT	AAG	GTC	CAG	ACG	AAA	GGA	GTG	ATC	AGG	CCT	CTG	3105
w1	...	..C	...	..C	C..	..C	...	..T	...	..A	...	..G	...	...	C..	...	...	
w2	...	..C	...	..C	C..	..C	...	..T	...	..A	...	..G	...	...	C..	...	...	

Q K R

	Q	P	E	I	Y	V	P	A	L	E	I	L	E	S	S	G	I	
m	CAA	CCG	GAA	ATC	TAC	GTT	CCA	GCA	TTG	GAG	ATC	TTG	GAG	TCG	TCG	GGC	ATC	3156
w1	...	...	..G	..Y	...	A.C	..T	..G	C..	...	...	...	..A	G..	...	...	...	
w2	...	...	..G	..C	...	A.C	..T	..G	C..	...	...	...	..A	G..	...	...	...	

A

	K	L	V	E	K	V	E	T	ter						
m	AAG	CTG	GTT	GAG	AAA	GTG	GAG	ACT	TGA	AAGTT	CCCTG	ATACA	CAGAT	AAAGA	
w1	...	GC.	A.A	...	..G	...	...	..C	...	GAGGA	TGTCA	GGATG	GGATG	AGAAT	
w2	...	G..	A.A	...	..G	...	...	..C	...	GAGGA	TGTCA	GGATG	GGATG	AGAAT	

V I R

m	TAGTA	TGATA	TAGCA	GGGCA	CATGT	ATCTT	TTGTA	TTAAC	TCCGT	TCTGG	AATAT
w1	CCACC	GAGTA	TATAT	GCTGC	TGCAA	CAGAG	GCAGC	GAGTA	AATAA	AATAT	GATGA
w2	CCACC	GAGTA	TATAT	GCTGC	TGCAA	CAGAG	GCAGC	GAGTA	AATAA	AATAT	GATGA

m	ATATT	TGTGA	ACTAA	AACGT	GACAA	ATAAA	AGAAA	TAAAA	AGAAC	GGAAG	TAGTA
w1	TTCTC	GCTGT	TGTAA	GTCCA	ATGAG	CGAAC	TGTAT	GTATG	TATGT	GCCTA	TCTAT
w2	TTCTC	GCTGT	TGTAA	GTCCA	ATGAG	CGAAC	TGTAT	GTATG	TATGT	GCCTA	TCTAT


```

m  TATTG TAAAA AAAAA AAAAA AAA
w1 TGTGT ATATA CCGAA TWCTG TCGCC GGTKT AAAAA AAAAA AAAAA AAAAA AA
w2 TGTGT ATATA CCGAA TCTGT CGCCG GTTGA TTCGG TTGGT GAACT TCCCA AAAAA AAAAA
AAAAA A

```

Fig. 12 Sequence of two partial wheat cDNAs encoding lysine-ketoglutarate reductase/saccharopine dehydrase (w1 and w2), compared to the complete coding sequence of the gene from maize (m; Accession # AF003551). Nucleotide residues identical to that of the maize sequence are indicated by a period (.). Putative encoded amino acid sequence of the maize gene is written above the maize nucleic acid sequence, while the deduced amino acid residues of the two wheat ESTs, if different from the maize sequence, are written below the appropriate nucleic acid sequence. The putative polyadenylation signals are underlined in the 3' untranslated regions. Numbers at the end of lines refer to nucleotide number starting at the translational initiation codon in the maize sequence.

Fig. 13 Southern blot analysis of genomic wheat DNA with the probe derived from the *SDH* domain of the wheat *LKR/SDH*. Genomic DNA was digested with restriction enzymes indicated on top of the panel. Each digestion contains four different repeats of the digestion products.

Fig. 14A

A 1% agarose electrophoretic gel of the positive *T. turgidum* LKR BACs digested with *HindIII* run in a cooled circulating buffer chamber for 17.5 hrs. at 50 V and stained with Vistra Green™ (Amersham Pharmacia Biotech).

Fig. 14B

A Southern Blot of the gel of the positive *T. turgidum* LKR BACs digested with *HindIII*, using the same probe as used for the initial screening.

Fig. 15

A and B represent two respective contigs of positive *T. turgidum* LKR BACs produced with Image (im3, Sanger Center, UK) and FPC (Clemson University Genomics Institute) software packages using Size Data setting, Tolerance of 8, and Cut-off value of $1e-05$. Asterisks indicate canonical clones, containing 90% of all restriction bands present in the contig. The tilde represents a clone having >90% of restriction bands the same size, within defined tolerance, as those in the canonical clone.

Fig. 16 Plasmids containing the genes encoding (A) dihydropicolinate synthetase (DHPS), and (B) aspartate kinase (AK) used for transformation into wheat. Bacterial coding sequences were ligated in frame to the wheat seed-specific promoter sequence from the gene encoding high molecular weight glutenin Dy10 (Glu-Pro), yielding plasmids ProCi (DHPS gene) and ProCk (AK gene). Tocs is an OCS terminator sequence. SSu represents the signal sequence from Pea rbcS3A, accession number X04333. Restriction sites are shown.