Theor Appl Genet (2006) 112: 618–626
DOI 10.1007/s00122-005-0160-7

ORIGINAL PAPER

Steven A. Brooks · Li Huang · Marie N. Herbel
Bikram S. Gill · Gina Brown-Guedira · John P. Fellers

Structural variation and evolution of a defense-gene cluster in natural populations of *Aegilops tauschii*

Received: 16 March 2005 / Accepted: 13 November 2005 / Published online: 10 January 2006

Abstract Genetic mapping and sequencing of plant genomes have been useful for investigating eukaryotic chromosome structural organization. In many cases, analyses have been limited in the number of representatives sampled from specific groups. The degree of intraspecific genome diversity remains in question. The possibility exists that a single model genome may have limited utility for identifying genes in related members of the species or genus. Crop improvement programs have particular interests in disease resistance genes that are harbored by wild relatives of modern cultivated crops. These genes are evolutionarily dynamic and under selective pressure by a broad range of pathogenic organisms. Using resistance gene analogs as models for gene evolution, intraspecific genome comparisons were made among populations of wild diploid wheat (*Aegilops tauschii*). We observed that deletion haplotypes are occurring frequently and independently in the genome. Haplotypes are geographically correlated and maintenance of gene complements in localized populations indicates selective advantage. Furthermore, deletion haplotypes are not detrimental to plant health, since genes without adaptive value in alternate environments are eliminated from the genome. Deletion haplotypes appear to be a common form of allelic variation in plants, and we address the consequences on genome restructuring and gene evolution.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-0160-7 and is accessible for authorized users.

Communicated by B. Keller

S. A. Brooks · G. Brown-Guedira · J. P. Fellers (✉)
Plant Science and Entomology Unit, Department of Plant Pathology, USDA-ARS, Manhattan, KS 66506, USA
E-mail: jpf@pseru.ksu.edu
Tel.: +1-785-5322367
Fax: +1-785-5326167

L. Huang · M. N. Herbel · B. S. Gill
Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

Present address: S. A. Brooks
DBNRRC, USDA-ARS, Stuttgart, AR 72160, USA

Introduction

Partial and complete genome-sequencing projects of plant and animal species are revealing fine-scale resolution of gene arrangement within eukaryotic chromosomes. However, the effects of genome structure on the evolution of novel genes remain unknown. Because genomes of related species can vary in size and organization, novel evolutionary mechanisms that are unique to or shared among specific genome lineages have yet to be discovered. Interspecific genome comparisons have revealed mechanisms of whole genome diversification, whereas intraspecific genome comparisons are required to answer questions on haplotype diversity and to glean information regarding evolution at specific genetic loci.

When comparing genomes, it is expected that for any specific gene in an individual organism there exist alleles of that gene at orthologous loci in other individuals of the same population and the species as a whole. This expectation was recently found to be false, as Fu and Dooner (2002) found deletion haplotypes in maize inbred lines. Using intraspecific bacterial artificial chromosome (BAC) sequence comparisons, gene loss was found to occur frequently and independently in maize lineages. Therefore, micro-colinearity (or synteny) can fail within a single species and is variable within different regions of a single genome.

With a high degree of plasticity and tolerance of gene loss in a genome, whole gene complements become a polymorphic trait. This is an unexpected form of variation and has dramatic consequences on gene evolution through genome restructuring. We expect the possibility of this type of variability in a diploid genome only if the genes involved are of minor importance. However, other

explanations are possible. If genes are locally adapted, in flux with dramatic selection pressure, or buffered by paralogous duplications within the genome, then deletion of entire gene loci will have little effect on overall plant fitness.

To examine the consequences of genome structure on gene evolution we investigated haplotype differences at a highly variable locus in *Aegilops tauschii* Coss., the D-genome donor of bread wheat (*Triticum aestivum* L.). Previously, Brooks et al. (2002) reported a 34 kilobase (kb) cluster of genes within a region of chromosome 1DS. We have continued the work by focusing on the three putative defense-related genes in *Ae. tauschii*. Accession TA1703 was used as the bases for detailed comparative analyses of coding sequence (CDS) 3, a leucine zipper-nucleotide binding site-leucine rich repeat (LZ-NBS-LRR) class resistance gene analog (RGA); CDS 5, an NBS-LRR class RGA; and CDS 7, an *N*-hydroxycinnamoyl/benzoyltransferase (HCBT)-like putative pathogenesis related (PR) gene. Using these genes, we determined the extent of gene conservation in the cluster among *Ae. tauschii* accessions collected from a wide geographical range. Mechanisms for structural diversification are shown to occur in this self-pollinated diploid species of wheat.

Materials and methods

Plant material and genomic DNA preparation

All accessions of *Ae. tauschii*, *Hordeum vulgare*, *Triticum monococcum*, and *T. aestivum* were supplied by the Wheat Genetics Resource Center at Kansas State University. Seedling leaf tissue was harvested at the three-leaf stage for genomic DNA isolation using a modified CTAB DNA extraction protocol (Hulbert and Bennetzen 1991).

PCR primer design

The complete DNA sequence of *Ae. tauschii* BAC clone M11 was used for primer design (GenBank accession number AF446141, Brooks et al. 2002). Primers were designed for PCR from genomic DNA and cDNA (RT-PCR) templates using MacVector™ 6.5.3 (Oxford Molecular Ltd., Madison, WI, USA), selecting 22–24mers with a T_m of 55–65°C. Positive RT-PCR control primers were designed from a putative actin unigene (TC21432) from the TIGR wheat gene index (TaGI) (http://www.tigr.org). A complete list of primers and sequences is provided in Table 1.

PCR conditions for genomic DNA

PCR was performed using the Herculase™ Enhanced DNA Polymerase Kit, reagents and protocol (Stratagene, La Jolla, CA, USA). A 50 µl total reaction volume was used with 400 ng of template DNA, 5 µl of 10× reaction buffer (Stratagene), 1 µl of 25 mM dNTPs, 200 ng of each gene specific primer, 2.5 units of Herculase™ DNA polymerase, and 3% dimethyl sulphoxide (final concentration). MJ Research (Waltham, MA, USA) PTC-225 thermal cycler conditions were as follows: initial denaturing step at 95°C for 2 min. Ten cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 3 min; 10 s per cycle was added to the extension (72°C) step for the next 20 cycles.

RNA isolation and RT-PCR

Plant tissue for total RNA isolation was collected from *Ae. tauschii* accession TA1703 at various growth stages (seedling leaf tissue only from TA2468). Leaf, shoot and root tissue was collected from seedlings at the three-leaf stage and adult plants at jointing. Spike tissue was

Table 1 PCR primer pairs used to amplify genomic DNA and cDNA

Forward primer	Sequence 5′–3′	Reverse primer	Sequence 5′–3′	Predicted amplicon size (nt)
Actin_F	GGTATCGTGAGCAACTGGGATG	Actin_R	GTGAAGGAGTAACCTCTCTCGGTG	383
CDS7_F2	CTTCAACACCTGGTCCGAACTAAG	CDS7_B2	TACACCGTCGCCTTGCCATCTATC	723
CDS7_F13	ATTTTCTTCCCCTCCGTCCC	CDS7_B15	GGTCAAACTTTTAGGGACGGAGG	4,340
CDS7_F14	ATGGTGGCTCTGTATGGTGTGTC	CDS7_B27	CAAACAGTTCGTGATTGAGCCAG	3,294
CDS7_F3	TATGCGACCGTCATCCGAAC	CDS7_B11	TTGGACTCGCAAAACGGCAG	2,913
F1	ATGGCAGGGTGTAGTATGGCAGAG	B3	TTGGAGGTGTGTGGTTAGCAGTG	2,758
F11	TTGGAACTGAGGACAAGTGCCC	RGAyr10_rev	CACTAACCCAATCACACCAAG	1,590
F14	ATAGCGTGGCTGCTGGAATG	B21	CCTTGTTCAATGCTTGCTCGC	1,776
F22	AGAGTGGAAGACCTGTGGGATACC	B24	GGGGTTTAGGGGTGATTTATGG	1,500
F26	CGAAGAGCGAGCAAGCATTG	B25	GCTGCTGATAAGTGTGAGGCATTG	1,902
F3	TCTCTATTGTTGGCTTTGGAGGC	B9	GCTCGCCTTATTAGCAACATCG	2,053
F30	CTCCCACCGTTTCTTTTTAGTCC	B32	ATAAGGAGAGGCTTTAGGCATCG	1,482
F31	GCATCTTGTCCGCTTTGACG	B32	ATAAGGAGAGGCTTTAGGCATCG	850
F35	CAACCTATTCAAATGTCCTTCGCC	B35	AATCTCTCCACTCAATGTTGGGC	1,921
F36	GGATTTTGCTTCCTTTCAGCCC	B37	CATCGCTTCTCTTTACAAACCGC	1,786
F46	GCCCAAACCTATCATTGTTGTCTC	B46	GCCATACTACACCCTGCCATTG	1,248
ORF7_F	AAGTCCGAAGAATCCCTTGAGC	ORF7_R	TTCTAACCCTCGCAACTCTCTCAG	3,343
patchF	GGTTCAGTATCAAGAGTAGGGG	patchR	CCCACCAAACAACGGATTAGATAG	1,261

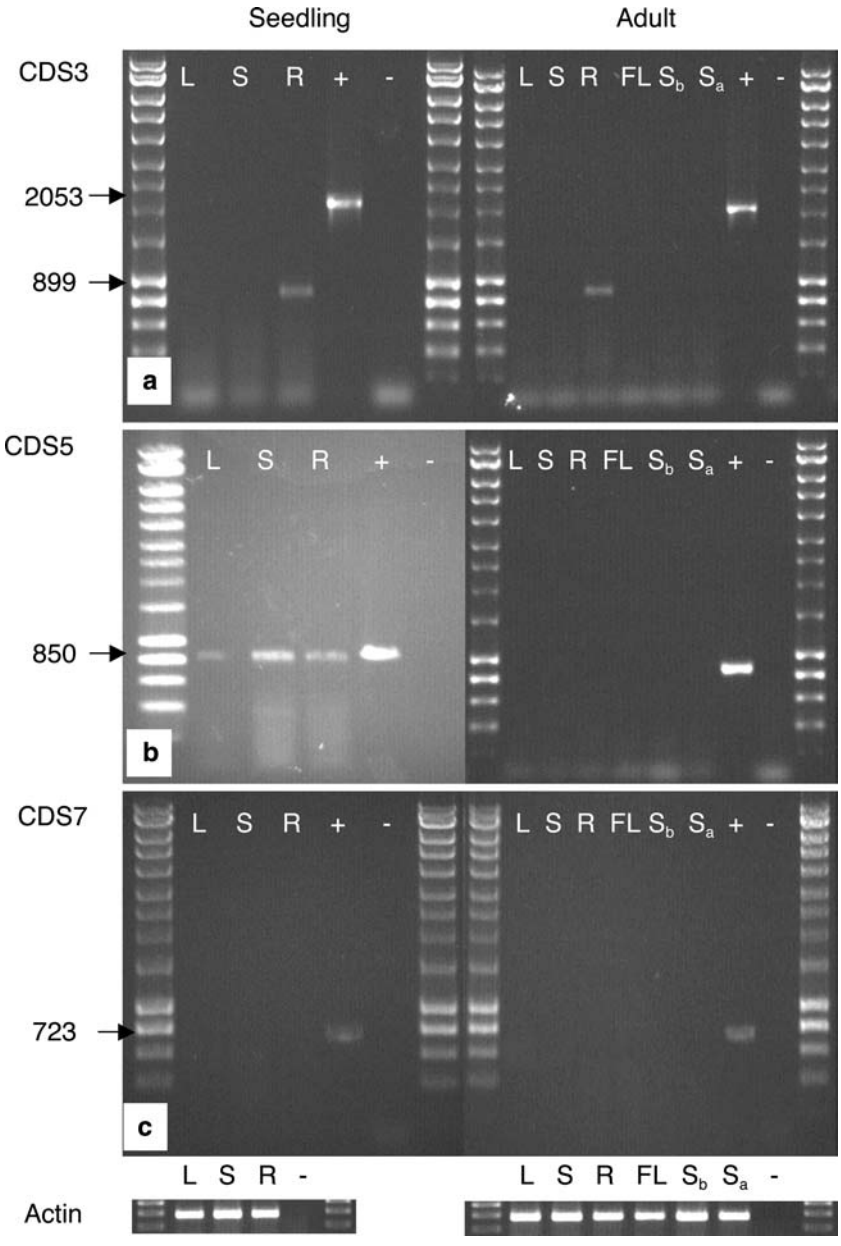

Fig. 1 Gene expression results by RT-PCR for CDSs 3 (**a**), 5 (**b**), and 7 (**c**) in TA1703. Tissue source for seedling and adult stages are as follows: leaves (*L*), shoots (*S*), roots (*R*), flag leaves (*FL*), booted spikes (S_b), and spikes 2 days post-anthesis (S_a). Positive (+) PCR control of BAC M11, and negative (−) water control. No plasmid control was available for actin, however, the predicted amplicon was obtained for all sample cDNAs. Numbers at left indicate size of each amplicon in nucleotides

harvested in the boot 2 days post-anthesis. Flag leaves were also collected 2 days post-anthesis. Tissue was harvested from three replications, four plants per replication and handled individually for each tissue and each growth stage. Harvested tissue was flash frozen in liquid nitrogen, ground with a mortar and pestle and preserved with RNA*later*™ (Ambion, Austin, TX, USA) according to the manufacturers protocol.

Total RNA was purified from 100 mg of preserved tissue using the RNeasy® Plant Mini Kit (QIAGEN) and protocol. The optional RNase-Free DNase Set (QIAGEN) was used to ensure elimination of contaminant DNA. One microgram of total RNA and 500 ng of oligo $d(T)_{18}$ primer were used for synthesis of first-strand cDNA using the SUPERSCRIPT™ II RNase H Reverse Transcriptase Kit (Invitrogen, Carlsbad, CA, USA) and protocol. One microliter of the completed reaction was used directly for PCR. Twenty-five microliters PCR reactions included 1.25 units of Taq DNA polymerase, 1× PCR buffer, 2.5 mM $MgCl_2$ (Sigma, St. Louis, MO, USA), 0.25 mM dNTPs, and 100 ng of each gene specific primer. PTC-225 thermal cycler conditions are as follows: initial denaturing step at 92°C for 3 min. Thirty-five cycles of 92°C for 1 min, 60°C for 1.5 min, and 72°C for 2 min; and a final extension step of 72°C for 10 min. Primers for actin were designed for use as a positive RT-PCR control. No cloned actin gene was available for positive PCR control, however, all samples amplified the predicted 383 bp product with ActinF and ActinR primers.

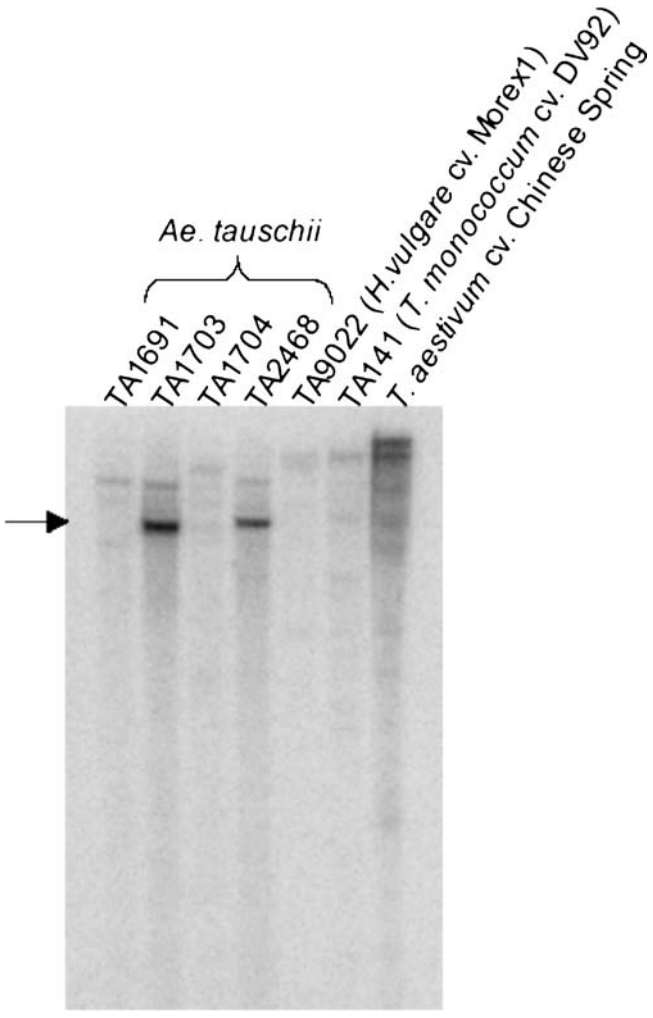

Fig. 2 Southern hybridization of *Aegilops tauschii* accessions TA1691, TA1703, TA1704, TA2468; *Hordeum vulgare* cv. Morex1; *Triticum monococcum* cv. DV92; and *Triticum aestivum* cv. Chinese Spring. Genomic DNA was digested with *Eco*RI and CDS 7 probe was used for hybridization. Positive signal is detected in accessions TA1703 and TA2468 only and indicated by the *arrow* at left

Southern hybridization

Southern blotting, probe labeling, hybridization, and filter washing were conducted as previously described (Huang and Gill 2001). Five micrograms (*Ae. tauschii*, *H. vulgare*, *T. monococcum*) and 15 µg (*T. aestivum*) of genomic DNA were used to prepare blots. The probe of CDS 7 was amplified by PCR (CDS7_F2-CDS7_B2, respectively) from BAC M11 DNA and purified by spin-column chromatography.

TA cloning and plasmid preparations

PCR products were cloned using the TOPO TA Cloning® Kit, reagents and protocol (Invitrogen). Amplified products were ligated directly into plasmid vector pCR®2.1-TOPO and transformed into TOP10 One Shot® chemically competent *E. coli* cells (F$^-$*mcr*A *Δ*(*mrr-hsd*RMS-*mcr*BC) *Φ*80*lacZΔ*M15 *Δlac*74 *rec*A1 *deo*R *ara*D139 *Δ*(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G) (Invitrogen). Plasmids were prepared for sequencing using the QIAprep® Spin Miniprep Kit, reagents and protocol.

Sequencing, sequence analysis and contig assembly

For each segment of genomic DNA, five different colonies were chosen and purified plasmids were sequenced directly with primers M13 reverse and T7 by the Kansas State University DNA Sequencing Facility. Base quality scores called by *phred* version 0.990722.f (Ewing et al. 1998; Ewing and Green 1998), contigs assembled using *phrap* version 0.990319 (http://www.phrap.org), and contigs edited with *Consed* version 11.0 (Gordon et al. 1998). Complete DNA sequences are available on the NCBI web site (http://www.ncbi.nlm.nih.gov). GenBank accession numbers are as follows: TA2468_contig2 AY613782, TA2468_contig1 AY613783, TA1691_contig2 AY613784, TA1704_contig1 AY613785, TA1691_contig1 AY613786.

Sequence annotation and comparison

FASTA files of assembled contigs were exported from *Consed* for sequence analysis. FGENESH 1.1 (http://www.softberry.com) was used for coding sequence (CDS) prediction with monocot genomic DNA parameters. Putative polypeptide sequences were defined by results of BLAST searches against the NCBI non-redundant database (Altschul et al. 1997; http://www.ncbi.nlm.nih.gov/BLAST/). Sequence comparisons and alignments were made using the MacVector™ ClustalW algorithm. (Mention of a trademark of a proprietary product does not constitute a guarantee of warranty of the product by the United States Department of Agriculture, and does not imply its approval to the exclusion of other products that may also be suitable.)

Results

Gene expression profiles

The relevance of the defense-related genes in TA1703 (BAC M11 donor) was explored by RT-PCR to determine whether these genes were expressed. CDS 3 was amplified from cDNA with primers F3 and B9 designed to flank the predicted intron (Table 1). The predicted amplicon from genomic DNA template was 2,053 bp, as amplified from BAC M11. An 899 bp amplicon was predicted and amplified only from root-derived cDNA from both seedlings and adults (Fig. 1a). Results verify intron prediction (Brooks et al. 2002) and the lack of

contaminating genomic DNA. CDS 5 was amplified from cDNA with primers F31 and B32 (Table 1), which were designed within the predicted single exon gene. Transcript was detected in seedling leaf, shoot and root tissue, but not in any adult tissue (Fig. 1b).

Amplification of CDS 7 with primers CDS7_F2 and CDS7_B2 was negative for all cDNA samples (Fig. 1c). An additional set of primers designed in a different exon produced the same result (data not shown), indicating no basal expression of CDS 7 detectable by RT-PCR. An additional experiment was designed to monitor CDS 7 expression following inoculation with leaf rust (*Puccinia triticina*), but no expression was detected. A search of the NCBI EST database with the first predicted exon of CDS 7 (positions 39,402–40,773 of BAC M11) produced significant, but not exact alignments to ESTs in an etiolated wheat seedling root cDNA library (GenBank accession BE446216; E value $= 9.9\times10^{-27}$), a *Blumeria graminis* infected barley EST library (GenBank accession BI952010; E value $= 2\times10^{-30}$), and a water stressed *Sorghum bicolor* EST library (GenBank accession BE592145; E value $= 3\times10^{-17}$). Since no transcripts or exact database matches were found, the expression of CDS 7 was uncertain.

Southern hybridization was used to determine whether CDS 7 was conserved in other *Ae. tauschii* accessions and related cereals. Genomic DNA was isolated from four *Ae. tauschii* accessions, TA1691, TA1703 (M11 donor), TA1704, and TA2468, *H. vulgare* cv Morex1, diploid A^MA^M *T. monococcum*, and hexaploid AABBDD *T. aestivum*. CDS 7 was used as a probe and results indicate that CDS 7 was only present in TA1703 and TA2468 and not present in *H. vulgare*, *T. monococcum*, or hexaploid wheat (Fig. 2). Weak hybridizations were present in

Table 2 Forty-two *Ae. tauschii* accessions, the country of origin, and their 1DS haplotype. An "X" indicates the presence of the indicated CDS

TA number	Taxonomic group[a]	Haplotype	CDS 3	CDS 5	CDS 7	Country
1691	*meyeri*	A	X			Unknown
1630	*tauschii*	E				Afghanistan
2404	*typica*	E				Afghanistan
2398	*typica*	E				Afghanistan
2388	*typica*	E				Afghanistan
2575	*typica*	C	X	X	X	Armenia
1681	*tauschii*	C	X	X	X	Azerbaijan
1601	*tauschii*	B	X	X		China
2483	*typica*	C	X	X	X	Iran
2452	*strangulata*	C	X	X	X	Iran
2468	*strangulata*	C	X	X	X	Iran
2472	*typica*	C	X	X	X	Iran
1703	*tauschii*	C	X	X	X	Iran
2482	*meyeri*	C	X	X	X	Iran
2488	*typica*	B	X	X		Iran
2375	*typica*	B	X	X		Iran
2492	*typica*	A	X			Iran
2450	*strangulata*	D		X		Iran
2460	*typica*	D		X		Iran
2455	*strangulata*	E				Iran
1717	*tauschii*	E				Iran
2526	*typica*	E				Iran
10147	*eusquarrosa*	C	X	X	X	Kazakhstan
10106	*eusquarrosa*	D		X		Kyrgyzstan
2385	*typica*	E				Pakistan
1704	*tauschii*	B	X	X		Tajikistan
10108	*eusquarrosa*	E				Tajikistan
10155	*eusquarrosa*	E				Tajikistan
10156	*eusquarrosa*	E				Tajikistan
10157	*eusquarrosa*	E				Tajikistan
10158	*eusquarrosa*	E				Tajikistan
10159	*eusquarrosa*	E				Tajikistan
1652	*tauschii*	E				Tajikistan
1588	*tauschii*	C	X	X	X	Turkey
1634	*tauschii*	E				Turkey
10175	*eusquarrosa*	C	X	X	X	Turkmenistan
1675	*tauschii*	C	X	X	X	Turkmenistan
10189	*anthera*	B	X	X		Uzbekistan
10199	*eusquarrosa*	B	X	X		Uzbekistan
10213	*eusquarrosa*	E				Uzbekistan
10123	*eusquarrosa*	E				Uzbekistan
10193	*eusquarrosa*	E				Uzbekistan

[a] *A. t.* form strangulata only, *A. t.* ssp. eusquarrosa forms typica, anthera, and meyeri

other species and may indicate similar sequence, as evident by the presence of similar ESTs.

Haplotype analysis

Differences in haplotypes were explored by evaluating a broader set of accessions in collections from the origin of diversity for *Ae. tauschii*. Forty-two accessions of *Ae. tauschii* collected from 12 different countries (Table 2) were used for analysis. Gene specific primers for CDSs 3, 5, and 7 (Table 1) were used to amplify target sequences from plant genomic DNA and to amplify probes from BAC M11 DNA for Southern hybridization. All accessions were scored for presence or absence of each gene by PCR and Southern hybridization. The two methods produced identical results in all samples (data not shown).

Five haplotypes of the defense-gene cluster were detected (Fig. 3, Table 2). Haplotype A is represented by CDS 3 only, haplotype B by both CDSs 3 and 5, haplotype C by all three CDSs, haplotype D by CDS 5 only, and haplotype E by the absence of all three genes. Results indicate that haplotypes are not associated with specific subspecies or varieties of *Ae. tauschii*, rather they are geographically correlated (Fig. 3). One member of haplotype A, five members of haplotype B, ten members of haplotype C, all D haplotypes and 18 E haplotypes were mapped geographically. The sampling of the genotypes is limited and the specific site of collection is not known for five of the accessions, TA1634, TA1691, TA1703, TA1704, and TA10147, thus have been omitted from the map. Strong inferences about distribution are difficult, however, B and D haplotypes appear to be geographically widespread. Haplotype C (ten accessions) is clustered in the regions surrounding the Caspian Sea. Haplotype E is primarily distributed in the eastern most range of the species, except for three accessions present in Iran and one in far Western Turkey.

Sequence analysis

The extent of sequence differences among members of specific haplotypes was evaluated using PCR and DNA sequencing. The RGA cluster was amplified from TA1691, TA1704, and TA2468 while genomic DNA of TA1703 and purified BAC M11 plasmid DNA were included as positive PCR controls. Primers were designed to produce overlapping amplicons (Table 1). Assemblies of contiguous DNA sequences were constructed spanning the complete coding sequences of CDSs 3 and 5, and the major exon of CDS 7.

TA1691 (haplotype A, Fig. 4) contained only one gene of the cluster described in TA1703. A 9,601 bp contiguous DNA sequence was assembled spanning CDS 3 into the 3′ untranslated region (UTR) of CDS 5. An additional 1,482 bp of the CDS 5 3′ UTR was obtained, separated from contig 1 by a 303 bp sequence gap. Comparison of the sequence of these contigs relative to TA1703 revealed 18 single nucleotide polymorphisms

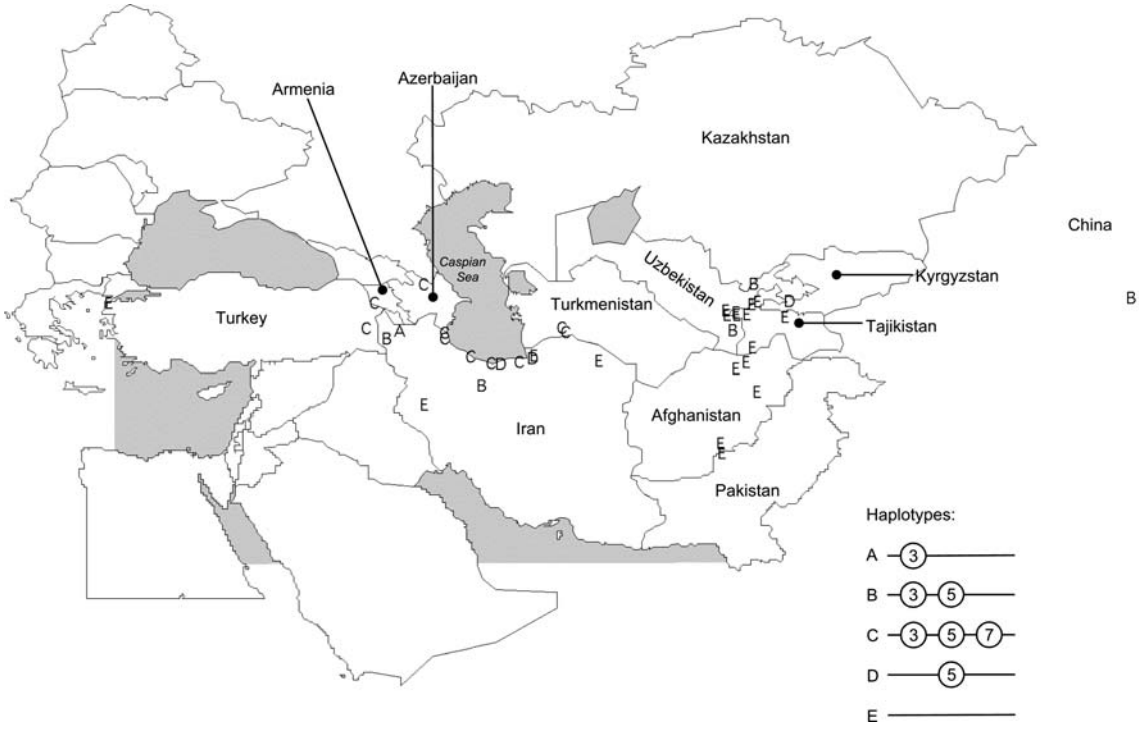

Fig. 3 Map of the geographical distribution for 37 of the 42 *Ae. tauschii* haplotypes

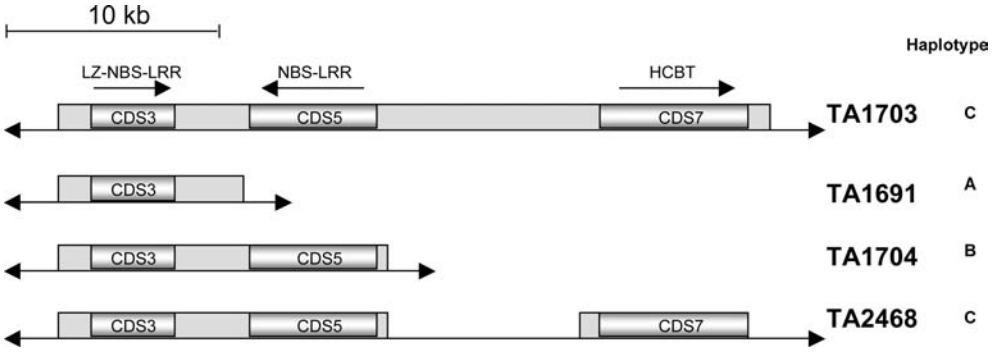

Fig. 4 Graphical representation of CDSs 3, 5, and 7 for *Aegilops tauschii* accessions TA1703, TA1691, TA1704, and TA2468. *Shaded bar* indicates sequenced regions and CDS position is shown by *labeled boxes*. *Arrows* above the sequence maps indicate direction of transcription

(SNPs), one 8 bp insertion/deletion (InDel) and one 32 bp InDel (Table 3 supplemental). Two SNPs were detected within the coding sequence of CDS 3 (positions 3,057 and 5,248), with one (3,057) mis-sense mutation altering the amino acid sequence of CDS 3. TA1703 encodes a leucine and TA1691 encodes phenylalanine 32 aa upstream of the P loop in the putative LZ. The 32 bp InDel in TA1691 (between positions 8,541 and 8,542) removes the polyadenylation signal sequence for CDS 5. The remaining sequence obtained for CDS 5 included the 3′ UTR, stopping 47 bp from the 3′ end of the single exon coding sequence. Thus, providing evidence that a deletion event occurred within the CDS 5 gene. The 2,541 bp of sequence homology to the 3′ UTR of CDS 5 in TA1703 was found. Only eight SNPs distinguish the two sequences (Table 3 supplemental).

TA1704 (haplotype B, Fig. 4) contains both RGAs found in TA1703. An 11,607 bp contiguous DNA sequence was assembled spanning CDS 3 into the 3′ UTR of CDS 5. An additional 3,299 bp of CDS 5 was obtained including a portion of the 3′ UTR and all of the single exon, separated from consensus 1 by a 499 bp sequence gap in the 3′ UTR. Comparison of the sequence in consensus 1 relative to TA1703 revealed 62 SNPs, and eight InDels (Table 4 supplemental). Twenty-one SNPs and two InDels occurred between the promoter and exon 1 of CDS 3, after which there is 100% conservation of exon 1, the intron, and exon 2 up to position 6,064 of consensus 1. A 288 bp InDel occurs in the 3′ end of exon 2 at position 6,065 involving a 78 bp duplication of sequence from positions 6,392 to 6,469 (Fig. 5a). Exon 2 is extended with a replacement of the final nine aa with a 47 aa sequence, resulting in a 827 aa polypeptide (Fig. 5). A 694 bp InDel occurs within the 3′ UTR of CDS 3 from positions 6,519 to 7,210 (Fig. 5), and does not have similarity to sequences described herein. The InDel does have 90% nucleotide identity to an *Ae. tauschii* LZ-NBS-LRR class RGA that is linked to *Lr21* (GenBank accession AF509533), suggesting an illegitimate recombination event between non-allelic homologs in the distal region of chromosome 1DS.

TA2468 (haplotype C, Fig. 4) contains all three CDSs described in TA1703. A contiguous DNA sequence of 15,042 bp was assembled spanning CDSs 3 and 5. Comparison of this sequence relative to TA1703 revealed seven SNPs and two InDels (Table 5 supplemental). All polymorphisms occur between the promoter and first exon of CDS 3. Otherwise, there is 100% nucleotide identity between the two sequences. Additionally, a 6,062 bp contiguous sequence was assembled spanning the 5′ end of CDS 7, including the promoter, major exon 1

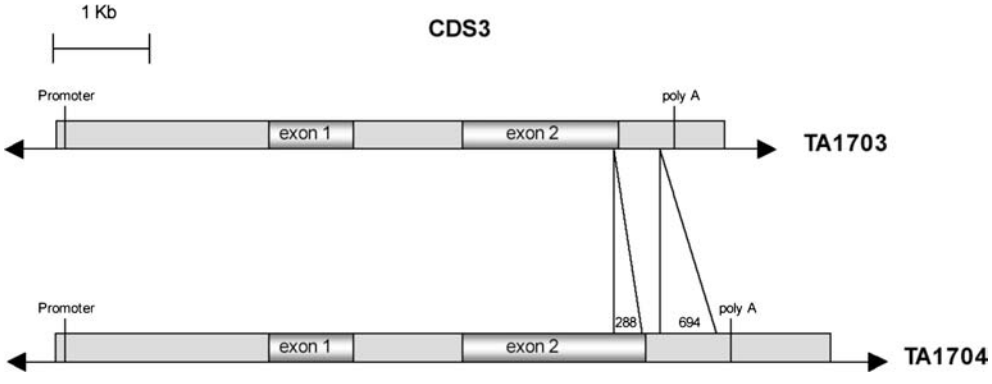

Fig. 5 Diagrammatic representation of CDS 3 coding regions in TA1703 and TA1704. Position of promoter, exons, and polyadenylation sequences are shown. Labels 288 and 694 indicate sequence InDels between TA1704 and TA1703

(encoding 457 aa of the 558 aa polypeptide), and most of intron 1. Within this region, eight SNPs and three InDels occurred relative to TA1703 (Table 5 supplemental). There is 100% conservation of the promoter sequence. Only one polymorphism was detected within exon 1 which was a SNP that did not alter the amino acid sequence of the polypeptide. All other polymorphisms occurred in non-coding regions of the gene.

Discussion

Previously, we described the defense-gene cluster in TA1703 and suggested that the arrangement may be more than a random association of genes (Brooks et al. 2002). The idea of conservation of RGA clusters as functional units was proposed. Here we describe varying levels of conservation of genes and gene sequence and have shown that gene deletions occur at this locus without any apparent detrimental effects. Thus, indicating that a buffering capacity for gene loss exists in the diploid wheat progenitors prior to formation of polyploids. The CDSs are located in an agronomically important and gene-rich region of the *Ae. tauschii* genome (Spielmeyer et al. 2000). Even though the exact function of these genes is unknown, this defense-gene cluster provides a unique genomic locus to study gene conservation.

Haplotype is correlated with geographic distribution

The Caspian region of Iran is regarded as the ancestral origin of the *Ae. tauschii* species and the region of hexaploid wheat formation and evolution (Lubbers et al. 1991; Dvorak et al. 1998). This region holds the greatest level of genetic diversity for the species. Subsets of *typica* genotypes are dispersed from diverse ancestral populations to become the widest distributed species-type with the greatest level of polymorphism (Lubbers et al. 1991). The data presented here is in agreement with previous work on genetic diversity in *Ae. tauschii*. Structural alterations are maintained within populations and haplotypes have become geographically correlated with species dispersal.

Haplotype C was only detected in the ancestral (Caspian) region of Iran and the predominance of this haplotype infers a selective advantage for the locus. If this is true then a hypothesis is necessary to explain the alternate haplotypes. Deletion of the entire coding sequence in the *rpm1*$^{-}$ allele of *Arabidopsis* has been observed (Tian et al. 2003). A fitness cost associated with the functional RPM1^{+} allele prevents it from being driven to fixation in *Arabidopsis* populations. Therefore, the susceptible allele is maintained because the deletion eliminates the possibility of fitness costs associated with production of susceptible protein variants. Tian et al. (2003) also speculate that resistance alleles (or loci) with the greatest fitness costs are least likely to be driven to fixation, and resistance and susceptibility alleles are maintained due to the costs and benefits associated with resistance. Given that alleles fluctuate in response to environmental pressure, it is expected that dominant haplotypes are indicative of selective pressure. Therefore, we can speculate that in the Caspian region of Iran, significant pressure must exist to maintain a dominance of haplotype C. However, at this time we do not know the selection agent.

A clustering of haplotype E was observed in the easternmost region of the species distribution. Is the dominance of this haplotype due to a lack of environmental pressure, or was distribution to this region an outcome of a population bottleneck? Dvorak et al. (1998) postulate that two geographically isolated ancestral populations of *Ae. tauschii* existed, one in Caspian-Iran, the other in north-central Iran (genepools: 'strangulata' and 'tauschii,' respectively). Genepool 'tauschii,' lacking the diversity associated with the 'strangulata' population, is believed to be the earliest colonizer of eastern and western limits of the species distribution. Evidence exists for the close relationship and lack of diversity among accessions collected from Afghanistan and Pakistan (Lubbers et al. 1991). This is consistent with our observations on the distribution of haplotype E, which is observed in Iran, but dominant in the most eastern distributions of the species. Therefore, it appears that loss of the entire locus (haplotype E vs. C) does not present a significant detriment to overall plant fitness.

Haplotypes are derived from independent deletion events

It was not clear if insertion or deletion was responsible for the differences between haplotypes C and E. To address this issue DNA sequence data from partial haplotypes was used to provide greater detail to the degree of diversity. The sequence of haplotype A accession TA1691 provided evidence for a breakpoint within CDS 5 where the remnant 3′ UTR had a high degree of sequence conservation with TA1703. An InDel was also observed in the 3′ UTR of CDS 3 in haplotype B accession TA1704 that appears to be the result of recombination with a non-allelic homolog. Finally, sequence analysis of all three haplotypes demonstrated that the order of genes remained constant even with variability in gene number.

Deletion mechanics

The questions remain of how the deletions arise? Sequence similarities to other RGAs on 1DS suggest illegitimate recombination between non-allelic homologs of CDS 3 in accession TA1704. This type of mispairing during meiosis can lead to duplication/deficiency gametes which could have produced the deletions observed in this region. Others have suggested non-allelic

recombination as the cause of deletions of genetic material, and indicate this is coincident with transposon turnover (Fu and Dooner 2002; Isidore et al. 2005; Scherrer et al. 2005). Deletion of DNA from the *Arabidopsis* genome is an outcome of illegitimate recombination between LTR retrotransposons (Devos et al. 2002), and transposable elements influence the stability of chromosome regions in barley making them more amenable to illegitimate recombination (Wicker et al. 2005).

Consequences for genome evolution

Bennetzen and Ramakrishna (2002) speculate that haplotype variation through gene deletion may be the result of the ancient polyploid nature of the maize genome, where deletions are tolerated and common in genome restructuring or the deletions are insulated by paralogous duplication of genes within the genome. Others have used these findings to support the dominance model for heterosis (Fu and Dooner 2002). Hemizygous hybrids from parental lines that lack different genes have good combining ability and display hybrid vigor. Inbreeding depression results from deletion haplotypes accumulating in the absence of the ability to fix hybrid vigor through recombination. Our observations are in a self-pollinating plant that lacks inbreeding depression, and recent observations in the wheat A genome (Isidore et al. 2005) and the barley H genome (Scherrer et al. 2005) have shown deletion events occurring at the *Lr10* and *Rph7* loci. Although we cannot make conclusions on the effects for heterosis, it appears that deletion haplotypes probably have little effect on inbreeding depression. Rather, the plasticity of the genome buffers phenotypic effects of gene loss.

Locus specific gene loss

The high degree of variability at the *Rph7* locus (Scherrer et al. 2005**)** has lead to the questioning of the genome specific deletion haplotypes in maize (Fu and Dooner 2002; Song and Messing 2003). It is evident that wheat and barley also have this variation and is not unique to any one-plant species. Occurrence of multiple independent deletion events is very frequent in *Ae. tauschii* at the M11 locus and supports a locus specific model for gene loss. Bearing this in mind, there are immediate implications other than evolution that have effects on studies in structural genomics. A structural comparison of genomes between species is limited by resolution without knowledge of the diversity that exists within a single species. Localized failure of synteny further complicates the exploitations of model genomes for genetic analysis in related species, such as wheat to rice comparisons.

Acknowledgements We extend our thanks to Angie Matthews for sequencing and sequence analysis, Dr. Craig Webb for assistance with PCR conditions optimization, Dr. Wanlong Li for use of Southern blots, and Deven See for assistance with Southern hybridization. We extend special thanks to Jon Raupp of the Wheat Genetics Resource Center at Kansas State University for preparing the map presented in Fig. 3. This manuscript is submitted as Journal number 05–48-J of the Kansas Agricultural Experiment Station and the United States Department of Agriculture.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bennetzen JL, Ramakrishna W (2002) Exceptional haplotype variation in maize. Proc Natl Acad Sci USA 99:9093–9095

Brooks SA, Huang L, Gill BS, Fellers JP (2002) Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. Genome 45:963–972

Devos K, Brown JKM, Bennetzen J (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Dvorak JM, Luo C, Yang ZL, Zhang HB (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 97:657–670

Ewing B, Green P (1998) Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res 8:175–185

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578

Gordon D, Abajian C, Green P (1998) *Consed*: a graphical tool for sequence finishing. Genome Res 8:195–202

Huang L, Gill BS (2001) An RGA-like marker detects all known *Lr21* leaf rust resistance gene family members in *Aegilops tauschii* and wheat. Theor Appl Genet 103:1007–1013

Hulbert SH, Bennetzen JL (1991) Recombination at the *Rp1* locus of maize. Mol Gen Genet 226:377–382

Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15:526–536

Lubbers EL, Gill KS, Cox TS, Gill BS (1991) Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34:354–361

Scherrer B, Isidore E, Klein P, Kim J, Bellec A, Chalhoub B, Keller B, Feuillet C (2005) Large intraspecific haplotype variability at the *Rph7* locus results from rapid and recent divergence in the barley genome. Plant Cell 17:361–374

Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci USA 100:9055–9060

Spielmeyer W, Moullet O, Laroche A, Lagudah ES (2000) Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii* the D-genome donor of wheat. Genetics 155:361–367

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature 423:74–77

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-elF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194