Mutational analysis of the major soybean UreF paralogue involved in urease activation

Joe C. Polacco1,4,6,*, David L. Hyten2, Mônica Medeiros-Silva1,7, David A. Sleper3,4 and Kristin D. Bilyeu4,5

1 Biochemistry Department, 117 Schweitzer Hall, University of Missouri, Columbia, MO 65211, USA and Department of Biophysics and Center of Biotechnology, Universidade Federal do Rio Grande do Sul, CEP 91501-970 Porto Alegre, RS, Brazil
2 Genomics and Improvement Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA
3 Division of Plant Sciences, 110 Waters Hall, and Center for Soybean Biotechnology, University of Missouri, Columbia, MO 65211, USA
4 Interdisciplinary Plant Group, 371 Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA
5 USDA-ARS, Plant Genetics Research Unit, 110 Waters Hall, Columbia, MO 65211, USA
6 Current address: Biochemistry Department and Interdisciplinary Plant Group, Univeristy of Missouri, Columbia, MO 65211, USA
7 Permanent address: Graduate Program in Cellular and Molecular Biology, Center of Biotechnology, Universidade Federal do Rio Grande do Sul, CEP 91.501-970, Porto Alegre, Brasil

* To whom correspondence should be addressed. E-mail: polaccoj@missouri.edu

Received 15 December 2010; Revised 29 January 2011; Accepted 31 January 2011

Abstract

The soybean genome duplicated ~14 and 45 million years ago and has many paralogous genes, including those in urease activation (emplacement of Ni and CO2 in the active site). Activation requires the UreD and UreF proteins, each encoded by two paralogues. UreG, a third essential activation protein, is encoded by the single-copy Eu3, and eu3 mutants lack activity of both urease isozymes. eu2 has the same urease-negative phenotype, consistent with Eu2 being a single-copy gene, possibly encoding a Ni carrier. Unexpectedly, two eu2 alleles co-segregated with missense mutations in the chromosome 2 UreF paralogue (Ch02UreF), suggesting lack of expression/function of Ch14UreF. However, Ch02UreF and Ch14UreF transcripts accumulate at the same level. Further, it had been shown that expression of the Ch14UreF ORF complemented a fungal ureF mutant. A third, nonsense (Q2*) allelic mutant, eu2-c, exhibited 5- to 10-fold more residual urease activity than missense eu2-a or eu2-b, though eu2-c should lack all Ch02UreF protein. It is hypothesized that low-level activation by Ch14UreF is ‘spoiled’ by the altered missense Ch02UreF proteins (‘epistatic dominant-negative’). In agreement with active ‘spoiling’ by eu2-b-encoded Ch02UreF (G31D), eu2-b/eu2-c heterozygotes had less than half the urease activity of eu2-c/eu2-c siblings. Ch02UreF (G31D) could spoil activation by Ch14UreF because of higher affinity for the activation complex, or because Ch02UreF (G31D) is more abundant than Ch14UreF. Here, the latter is favoured, consistent with a reported in-frame AUG in the 5’ leader of Chr14UreF transcript. Translational inhibition could represent a form of ‘functional divergence’ of duplicated genes.

Key words: Eu2, Nickel, SNP, soybean, urease, UreF, Universal Soy Linkage Panel.

Introduction

Soybean expresses an active seed urease, synthesized in the developing embryo (Torisky and Polacco, 1990; Torisky et al., 1994) and termed the embryo-specific urease (ESU). High seed urease activity is a trait shared with other members of the leguminosae (Bailey and Boulter, 1971) and the cucurbitaceae (Fahmy et al., 1994). Loss of ESU has no apparent physiological effect (Stebbins et al., 1991). A possible non-metabolic role of urease(s) is in seed and embryo defence (Polacco and Holland, 1993). Consistent with a role in defence ESU and jackbean seed urease were
reported to be insecticidal, due to the action of a sub-peptide released by cathepsin-type proteases in the gut of susceptible insects (Carlini and Polacco, 2008). ESU is encoded by the **Eu** gene (Glyma05g27840) (Meyer-Bothling and Polacco, 1987; Torisky et al., 1994).

Soybean expresses a second urease, the tissue-ubiquitous urease (UU), essential for assimilation of urea, either endogenously generated or environmentally available. UU is encoded by **Eu4** (Glyma11g37250), which is expressed in all tissues examined, including callus and the developing embryo (Torisky et al., 1994; Goldraij et al., 2003). Loss of UU activity leads to accumulation of urea in seeds and in leaves to the extent that leaf tips become necrotic (Stebbins et al., 1991), a trait also observed in soybean plants deprived of nickel (Eskew et al., 1983), a component of the urease active site (Carter et al., 2009). Endogenous urea is produced by arginase action, as during mobilization of seed protein reserves (Goldraij and Polacco, 1999). Whether purine degradation in plants liberates urea has been long debated (Muñoz et al., 2001, 2006; Todd and Polacco, 2004, 2006). In tropical legumes, ureides (purine degradation products) are the major forms of fixed nitrogen transported out of the nodule, so that conversion of some or all of the ureide pool to urea would make urease essential for full assimilation of fixed N. However, in the non-legume, *Arabidopsis thaliana*, enzymatically catalysed purine turnover has been demonstrated not to liberate urea (Werner et al., 2009), an observation that agrees with our earlier results in soybean plants (Winkler et al. 1985, 1987). In addition, it was shown that isogenic urease-negative mutant and urease-positive soybean plants, when supported by fixed N, exhibited no differences in growth or in total accumulated seed protein. Further, urease-negative callus did not have impaired utilization of the ureide, allantoin (Stebbins and Polacco, 1995). However, unstable degradation intermediates can liberate urea non-enzymatically (Werner et al., 2009), and this instability may explain our observation of increased urea accumulation in fixing soybean plants (Stebbins and Polacco, 1995; Todd and Polacco, 2004).

Normal function of **Eu2** and **Eu3** is necessary for full enzyme activity of both ESU and UU (Meyer-Bothling et al., 1987). The ‘pleiotropic’ **Eu3** gene (Glyma08g08970) is the plant orthologue of bacterial **ureG** (Freyermuth et al., 2000). In most bacteria, UreG, and associated UreD and UreF proteins, are essential for construction of the urease active site: the addition of Ni and CO2 to form a di-Ni carbamate on an active site lysine (Carter et al., 2009). In *Arabidopsis* UreD, UreF, and UreG are essential, and possibly sufficient, for activation of urease (Witte et al., 2005). Bacterial UreG has GTase activity (Carter et al., 2009) and UreF has been proposed to be a GTase activase (Salomone-Stagni et al., 2007). Plant UreG has a nucleotide-binding site and, in addition, can bind Ni, due most likely to a His-rich N-terminal extension not found on bacterial orthologues (Freyermuth et al., 2000). In soybean, the dual ‘substrates’ for **Eu3**/UreG and **Eu2**, namely the ESU and UU apo-isozymes, are distinguishable by tissue distribution, level of expression, and the propensity to form trimers or hexamers (Polacco and Havir, 1979; Polacco et al., 1985). Activated ESU and UU differ in *Km*, pH optimum, and inhibitor sensitivity (Kerr et al., 1983; Polacco and Winkler, 1984).

The **Eu2** product has not heretofore been identified. The elucidated complete genome sequence of soybean (Schmutz et al., 2009) contains two closely related paralogues each for **UreD** (Glyma02g20690 and Glyma20g17990) and for **UreF** (Glyma02g44440 and Glyma14g04380). Duplicate genes are not surprising since the soybean genome underwent two duplications within the past 45 million years (Schlueter et al., 2004). Thus, we postulated that **Eu2** is a single-copy gene encoding a different function—possibly a Ni-carrier protein related to UreE (reviewed in Carter et al., 2009). Alternatively, **Eu2** could represent the only functional parologue of duplicated **UreD** or of **UreF**. The two annotated **UreF** paralogues are on chromosomes 2 and 14, and we previously reported that cDNA for the chromosome 14-encoded UreF was functional in the fission yeast Schizosaccharomyces pombe (Bacamawmo et al., 2002). Here, it is shown that the chromosome 2 **UreF** paralogue encodes the major functional UreF in soybean: two different mutant **eu2** alleles co-segregate with mutations leading to radical alterations in highly conserved residues in Ch02UreF, alterations resulting in <1% urease activity. However, a null mutant of **eu2** contains significant basal urease activity, 5–10 times that of the missense mutants. The argument is considered that the **eu2** missense alleles encode UreF variants that compete against Chr14UreF participation in the activation complex, while the **eu2-c** null mutant ‘frees up’ apo-urease for partial activation by the ‘silent’ paralogue (Ch14UreF).

### Materials and methods

**Nomenclature and plant material**

Bacterial and fungal genes are designated in lower case italics (*ureG*) whereas soybean genes are capitalized (**Eu3** or **UreG**). Mutant alleles are always lower case (as in the soybean **eu2-a**, **eu2-b**, and **eu2-c** alleles). Protein products are capitalized and non-italicized (**Eu2** or **UreF**).

The **eu2** mutant, first reported in Meyer-Bothling et al. (1987) and most recently in Bacamawmo et al. (2002), is here termed **eu2-a**. The allelic mutation in isolate EN24 (Polacco et al., 1989), was termed **eu2**<sub>EN24</sub> in Bacamawmo et al. (2002) and is here designated **eu2-b**. **eu2-a** and **eu2-b** were recovered, respectively, from populations treated with ethyl methane sulphonate and N-methyl-N-nitrosourea. A third allele, **eu2-c**, described here for the first time, was recovered in the same ethyl methane sulphonate-treated population of ‘Williams’ that yielded **eu2-a** (Meyer-Bothling et al., 1987). **eu2-a** was out-crossed twice to ‘Williams 82’ and propagated for at least six generations of single-seed descent. **eu2-b** was induced in ‘Strayer 2233’ and was not out-crossed.

**Soybean crosses**

Crosses were made in early 2005 at Costa Rica Soybean Seeds located near Upala, Costa Rica. F1 plants were grown at the Bradford Research and Extension Center of the University of Missouri, and single plant threshed. Crosses of **eu2-a** (‘Williams 82’ background) and

**Data availability and acknowledgments**

The SNPs were obtained from the Soybase database (Wessler et al., 2007). The authors would like to thank the Soybase database for access to the soybean sequence.

**Ethical responsibility**

This research did not involve the use of human or animal subjects. All procedures were approved by the University of Missouri Animal Care and Use Committee.

**Conflict of interest**

None declared.
eu2-b (‘Strayer 2233’) were made reciprocally with ‘Maverick’ as the Eu2 (urease-positive) parent. Hybrid F1 individuals were detected by urease phenotype, flower colour, and pubescence. Upon confirmation of segregation of urease-positive and urease-negative seeds, several F2 populations were pooled for each cross. In this study, F2 progeny of the crosses eu2-a x Maverick and eu2-b x Maverick were analysed (maternal parent given first) by bulked segregant analysis. Reciprocal crosses gave the same F2 progeny ratios.

**Urease phenotyping**

Seed urease activity in parents and F2 populations was determined by the non-destructive, semi-quantitative ‘seed chip’ assay (Meyer-Bothling et al., 1987): a sliver taken from the hilum, and free of the seed coat, was placed in a glass test tube to which was added 0.1 or 2 ml of indicator solution [0.1 M urea, 10 mM KPO4, 1 mM EDTA, pH 7.0, 5 μg ml–1 cresol red and 0.02% (w/v) NaN3 to inhibit microbial growth]. At room temperature, wild-type, urease-positive, slivers turned the solution from yellow (pH 7.0) to pink (pH~7.5) to ‘red’ (pH=8.0) usually within 15 min, due to consumption of H+ in urea conversion to 2 NH4+ and HCO3-.

To confirm the phenotypes of all individuals included in bulked phenotypic classes, urease activity was again determined by ‘chip’ analysis of approximately one-third of a cotyledon from 5-day seedlings germinated from ‘chipped’ seeds.

**Urease assays**

Urease activity in mature seeds was determined by analysis of urea-derived ammonia. Two seeds of the same genotype (Table 2) were crushed with a hammer and, after removal of seed coat, combined and ground in a mortar with 20 vol of KTE (0.1 M TES, 1 mM EDTA, combined acidic forms taken to pH 7.4 with KOH). After chilling and spinning at ~14,000g for 5 min, the supernatant under the lipid cap was re-centrifuged. It was diluted 1:10 to 1:50 for ‘wild type’ (Williams 82), and used undiluted for eu2-a, eu2-b, and eu2-c. Extracts (0.1 ml) were added to 1.9 ml KTE+10 mM urea, incubated at 37 °C for 5 h and added to 3.1 ml of 133 mM NaOH. Following addition of 0.1 ml of Nessler’s reagent, A436 was determined after 10 min and converted to ammonium by comparison with identically treated standards of (NH4)2SO4 in KTE. Single F2 (Fig. 2) or F3 (Fig. 3) seeds were crushed with a hammer, ground to a powder, half of which was used for sequence characterization of Ch02UreF after DNA extraction. The remainder was assayed as above, and appropriate dilutions of their extracts were made.

Urease activity in seedling trifoliates was determined by analysis of 14CO2 production from [14C]urea. First trifoliate leaves of two 10-day-old plants, grown in nutrient solution in ‘inoculation pouches’ in a growth chamber, were harvested, quick frozen in liquid N2 and stored at ~80 °C. They were ground in liquid N2 and then in 3 ml of KTE and centrifuged at ~14,000g for 5 min. Aliquots (0.1 ml) of the supernatant were added to 1 ml of KTE+10 mM [14C]urea (10-3 μCi/ml), which was incubated at 37 °C for up to 4 h. Reactions were stopped by addition of 0.2–0.5 ml of 1 M H2SO4. Acid-released 14CO2 was trapped and quantified as described by Meyer-Bothling and Polacco (1987).

Seed urease activity was expressed as μmol urea hydrolysed h–1 mg protein–1 and trifoliate leaf activity as nmol urea hydrolysed h–1 mg protein–1. Protein was determined by Coomassie binding, following the instructions of the manufacturer (Bio-Rad) of the Bradford Reagent, using bovine serum albumin as a standard.

**The Golden Gate assay**

The Golden Gate ‘Highly Parallel SNP Genotyping’ assay for single nucleotide polymorphism (SNP) alleles as described in Fan et al. (2003) was performed essentially according to protocols provided by the manufacturer (Illumina, San Diego, CA, USA). Details for soybean are given in Hyten et al. (2008); they employed an array of 1536 possible segregating SNPs on the Universal Soy Linkage Panel (USLP1.0). Briefly, single-stranded genomic DNA is a template for multiplex allele-specific primer extension and is inactivated by incubation at >1500 loci. These products, in turn, are templates for PCR-produced allele-specific oligonucleotide probes (with one of two fluorescent tags for each allele) which hybridize to beads containing ‘address’ sequences for each SNP locus. Each well of an array contains complementary address sequences on beads at the ends of ~50,000 optical fibres, ~30 fibres per locus. The addresses on each bead type are previously ‘deciphered’. In the case of homozygosity for an SNP (in the DNA sample) individual fibres will be labelled with one probe or, in the case of heterozygosity for an SNP, with two differentially tagged probes.

**Bulked segregant analysis**

DNA was prepared (DNeasy Plant Maxi kit; Qiagen) from pooled leaves of 24 or 9 urease-negative F2 individuals from crosses of Maverick by eu2-a or eu2-b, respectively. Each pool was analysed separately for SNP analysis. Similarly, DNA was prepared and analysed from 36 or 20 pooled urease-positive F2 individuals, respectively, from the Maverick × eu2-a and Maverick × eu2-b crosses, as well as from 10 individuals of each of the three parents.

**Identification of genomic region with disproportionate eu2 donor loci**

Polymorphic SNP loci were defined as those for which the parental lines contained alternative SNP alleles. SNP information for bulk parental DNAs and for the bulk urease-positive and negative F2 lines contained alternative SNP alleles in a 2:1 ratio favouring the urease-positive parent. This ‘skew’ of the SNP data was the criterion for the urease-positive bulk to corroborate the co-segregation of the urease-negative trait with tightly linked SNPs. Unlinked polymorphic loci were those for which both bulked F2 phenotypes scored as heterozygous.

**qRT-PCR of UreF transcripts**

Steady-state RNA levels for the chromosome 2 and 14 UreF genes were determined by quantitative RT-PCR from field-grown Williams 82 young trifoliates developing leaves and seeds (embryo plus seed coat) that averaged either 7.8 or ~10 mm in length. RNA was prepared using the TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). Each RNA sample was adjusted to 50 ng μl–1 water and subjected to DNase I (Ambion, Austin, TX, USA) treatment for 30 min at 37 °C in 1× DNase I buffer (10 mM Tris, pH 7.5, 2.5 mM MgCl2, and 0.5 mM CaCl2). DNase I was inactivated by incubation at 75 °C for 15 min. RNA templates (100 ng) were used in 20 μl of a one-step reverse transcriptase PCR using the Quantitect SYBR Green RT-PCR kit (Qiagen, Valencia, CA, USA) supplemented with 0.2-titanium Tag polymerase (Clontech, Mountain View, CA, USA). A common reverse primer was used for both chromosome 2 and chromosome 14 UreF genes: UreFxr2:atgcctttcgacccacttc. The chromosome 2 UreF-specific primer was UreF2F3: ctgagaagctcgctgcttg, and the chromosome 14 UreF-specific primer was UreF14F3: ctggaagcagctgc. Control primers were PEPC16fw: TTCCTTTATCGAAGA-TACGGATTTAGCT (phosphophenylphosphate carboxylase) and PEPC16rev: GTGTCTATTTTGGCCGAGCG (Tuteja et al., 2004). Primer efficiency was calculated from PCR amplification of serial dilutions of genomic Williams 82 DNA templates without a reverse transcriptase step. qRT-PCR was performed with a 30-min 50°C
step followed by a single 95 °C/15-min step. PCR proceeded with 35 cycles of 95 °C for 20 s, 63 °C for 20 s, and 72 °C for 20 s with a fluorescence read at the end of each cycle. A melting curve was produced as the temperature increased from 70 °C to 90 °C by a fluorescence read every 0.2 °C with a 1-s hold during the melt. For each primer pair, RNA samples were assayed in triplicate, quantified using the appropriate standard curve, and normalized relative to the PEPC16 control. Control reactions were run with RNA samples but no reverse transcriptase, and negligible levels of contaminating DNA template were observed.

Sequence analysis

The DNA region encompassing the soybean chromosome 2 UreF gene was aligned with the chromosome 14 UreF gene to enable selection of primers capable of discriminating amplifying only the chromosome 2 UreF sequence. Primers were UreF1: GTTGA-TAAAAGTTCTGGGCT and UreFt1: ACATGCATCGAAA-TATGATGTGCA. Since neither UreF gene in soybean contains an intron in the translated region, amplification of the chromosome 2 UreF gene was performed using the bulked pooled and parental sample DNA templates described above but diluted in water 10-fold. PCR was carried out using Ex Taq according to the manufacturer’s recommendations (Takara, Otsu, Shiga, Japan) using the following conditions: 95 °C for 3 min followed by 35 cycles of 95 °C for 20 s, 60 °C for 20 s, and 72 °C for 1.5 min. PCR products were verified for size by electrophoresis on a 1.2% FlashGel (Lonza, Basel, Switzerland). PCR products were isolated with the Qiaprep Spin Miniprep kit (Qiagen) and sequenced at the University of Missouri DNA Core facility. Templates for PCR products were isolated with the Qiaprep Spin Miniprep kit (Qiagen) and sequenced at the University of Missouri DNA Core facility.

Results

Mapping strategy to identify Eu2

UreD, UreF, and UreG form a urease activation complex in bacteria (Carter et al., 2009) and each protein is essential for in vivo urease activation in Arabidopsis (Witte et al., 2005). The sequenced soybean genome (Schmutz et al., 2009) indicates two predicted soybean paralogues encoding UreD, two encoding UreF, and only one for UreG, the latter identified to be the soybean Eu3 gene (Freyermuth et al., 2000). Since mutation in Eu2 results in the same phenotype as mutation in Eu3, namely the elimination of the activity of both ureases, mutation in an accessory gene was suspected. However, the strategy of sequencing two UreD and two UreF-coding cDNAs in the eu2 background was not pursued because (i) genetic redundancy and the severe urease phenotype seemed to eliminate UreD and UreF, and (ii) mapping Eu2 could reveal other candidate urease accessory genes, such as a plant UreE. In Klebsiella aerogenes, and other bacteria, UreE is a Ni chaperone essential for full activation of urease (Mulrooney et al., 2005). Hence, Eu2 was mapped to enable identification of linked candidate genes.

Bulked segregant analysis of polymorphic SNPs maps Eu2 to chromosome 2

eu2 mutants were each crossed to a urease-positive parent, ‘Maverick’: A. eu2-a (‘Williams 82’)×Maverick, and B. eu2-b (‘Strayer 2233’)×Maverick. F2 progeny were threshed from urease-positive F1 plants and urease-positive and urease-negative F2 segregants were pooled. Table 1 shows a typical segregation pattern of urease-positive and urease-negative F2 seeds, assayed by the non-destructive ‘seed chip’ assay (Meyer-Bothling et al., 1987). The seed chip assay, as performed, did not distinguish eu2/Eu2 from Eu2/Eu2 segregants. For crosses A and B (Table 1), pools of 24 or 9 urease-negative and 36 or 20 urease-positive individuals were germinated and analysed by bulked segregant analysis. DNA from trifoliate leaves was pooled and prepared for multiplex SNP analysis on the USLP1.0.

For this analysis, ‘skewed’ loci were searched for: those loci that demonstrated the following characteristics: (i) an identical genotype for the urease-negative bulk and the mutant parent; (ii) a contrasting polymorphic genotype for the Maverick parent; (iii) a heterozygous genotype call for the urease-positive bulk. For the skewed loci, the urease-positive bulk presumably consisted of a 2:1 mixture of heterozygous and wild-type genotypes, but the technical limitations of the analysis made it difficult to categorize this result for any locus. The USLP1.0 provided 660 combined polymorphic loci including 214 common to the two crosses. Of these, 19 skewed loci were identified for population A and 25 for population B, but the majority of these were dispersed throughout the genome with regions of non-skewed polymorphic loci separating them. However, for both populations the distal end of chromosome 2 contained a number of skewed loci uninterrupted by other non-skewed polymorphic loci. For population A, this region encompassed 6.8 million base pairs and for population B, 3.7 million base pairs in the same chromosome 2 region. There were polymorphic loci proximal to the skewed region in

Table 1. F2 transmission of mature seed chip urease traits. Reciprocal crosses, not shown here, gave essentially the same ratios.

| Cross (seed chip class) | F2 mature seed chip class | n | \(\chi^2\) | \(P^b\) | A. eu2-a/eu2-a×Maverick (3) | 72 | 28 | 0 | 100 | 0.48 | 0.49 |
|------------------------|---------------------------|---|----------|--------|-----------------------------|----|----|---|-----|-------|-----|
| B. eu2-b/eu2-b×Maverick (4) | 70 | 0 | 28 | 98 | 0.67 | 0.41 |

a A non-destructively removed seed chip drives urea hydrolysis accompanied by a pH increase which is detected by the conversion of cresol red from yellow (pH 7.0) to pink (pH ~7.5) to red/vermillion (pH >8.0). This conversion takes the time courses: I (WT), red at 0.5 h; III, pink at ~10 h, red at 24–48 h; IV, yellow at 24 h, pink at 48 h, red at >48 h.
both cases, but only one marker distal to the last skewed locus, and it was not polymorphic for either population.

Mutant alleles of soybean chromosome 2 UreF co-segregate with two allelic eu2 bulked segregants

The soybean genome contains two UreF genes. That on chromosome 14 (Glyma14g04380) has previously been characterized and demonstrated to activate the host urease in the fission yeast (Bacanamwo et al., 2002). The second UreF is on chromosome 2 (Glyma02g44440) and is located within the mapped region for eu2 (ORF beginning at nucleotide 49017308), but has not been characterized previously. Since the mapped markers associated with lesions in eu2-a and eu2-b flanked UreF on chromosome 2, and since no other candidate urease accessory genes were located in this region of chromosome 2, we examined the UreF gene sequence from the chromosome 2 paralogue in the bulked samples and the parental lines. As shown in Fig. 1, the chromosome 2 UreF paralogue unambiguously yielded a radical change in a glycine for each eu2 allele: G148R for eu2-a and G31D for eu2-b. In each case a single G to A transition in positions 442 and 92 of the Ch02UreF coding sequence (Glyma02g44440 available at http://www.phytozome.net) was responsible for the amino acid change.

Though there is only 20–30% sequence identity between plant and bacterial UreF homologs, Cao et al. (2010) have presented evidence for conservation of the arrangement of helical domains between UreF from rice and the bacterial UreF of Helicobacter pylori and of K. aerogenes. Based on the experimentally determined structure of H. pylori UreF (Lam et al., 2010) the changes in Ch02UreF involved conserved G residues. G148R (eu2-a) corresponds to H. pylori position 164 (three of six bacterial residues=G), in the middle of helix g, and G31D (eu2-b) to H. pylori position 46 (five of five residues=G) between helices a and b. Replacement of G residues with the much more bulky, and charged, R and D residues is consistent with loss of >99% urease activity in both eu2-a and eu2-b (Table 2 and Polacco et al., 1999). However, the degree of reduction of urease activation depends not solely on the nature of the amino acid change, but also on the new interactions of the altered UreF with other components of the activation complex.

While the nucleotide sequence of bulked urease-negative F2 segregants (24 or 9 individuals from crosses of Maverick by eu2-a or eu2-b, respectively) showed complete inheritance of the mutant base, bulked urease-positive F2 segregants presented a mixture of the wild-type base and the mutant base in an approximation of the expected ratio of 2:1 (data not shown).

A novel ‘null’ eu2 mutant with higher residual urease activity

The eu2-c mutant was identified by its pleiotropic reduction of embryo-specific and ubiquitous urease activity, albeit with higher levels than found in eu2-a and eu2-b. In crosses eu2-c complemented the eu3-a mutant (not shown). However, in crosses to eu2-b, the F1 was urease negative and the F2 segregated no urease-positive individuals, only those resembling the parents with respect to the seed chip urease assay (Fig. 2). Compared with eu2-a, which gives a fully

---

Fig. 1. Alignment of soybean UreF paralogues encoded by Ch02UreF, Ch14UreF, and two eu2 alleles of Ch02UreF, eu2-a and eu2-b (Polacco et al., 1989) were mapped by the USLPl1.0. Positions of amino acid changes are indicated by a + for eu2-b and a ‘×’ for eu2-a. Identical amino acid residues are highlighted in black while similar amino acid residues are highlighted in gray. The eu2-c allele has a C to T transition in the second codon, resulting in a stop (not shown). The At1g21840 sequence is the single UreF gene of Arabidopsis (Bacanamwo et al., 2002).
of the Ch02 sequence resulting in a stop in the second codon (Q2*). The mutation contains a C to T mutation in position 4 of its coding sequence, resulting in a translational termination, is expected to result in a urease activity than the average of the b/b and c/c activities and, when quantified, eu2-c has higher urease specific activity than the eu2-a and eu2-b missense mutants in both seeds (mainly ESU) and seedling trifoliates (all UU) (Table 2).

Models to explain why a mutant apparently lacking all Ch02UreF protein has measurable urease activity include: (i) downstream initiation of protein synthesis at methionine in position 78 (Fig. 1), or (ii) unimpeded activation by Ch14UreF—a low-level activation without interference from the ‘spoiler’ missense forms of Ch02UreF encoded by eu2-a and eu2-b. Since highly conserved glycine31, the mutated residue in eu2-b, falls in the middle of amino acids 1–77, downstream translation initiated at methionine78 in model 1 is not favoured. In model 2 the G148R (eu2-a) and G31D (eu2-b) forms of Ch02UreF effectively prevent Ch14UreF from participating in the urease activation complex.

Fig. 2. Analysis of urease levels among F2 segregants in a cross of eu2-b×eu2-c. Twelve seeds from a single plant were chosen at random, ground, and on a portion, genomic sequence was determined for the Ch02UreF ORF. Each mutant site characteristic of eu2-b or eu2-c showed either homozygosity (wild type or mutant) or heterozygosity (mutant/wild type), and both sites were consistent within each individual (e.g. homozygous wild type for one site corresponded to homoygous mutant for the other). The remainder of ground seed, in a ‘blind’ experiment, was assayed for urease activity (b/b, eu2-b/eu2-b; c/c, eu2-c/eu2-c; bc, eu2-b/eu2-c). The horizontal line is the average of specific activities of the three eu2-b/eu2-b and three eu2-c/eu2-c individuals.

positive seed chip assay (yellow to red) in 24–48 h (Table 2), the eu2-c seed chip phenotype exhibited more urease activity. On the decreasing activity scale of eu2-b (b/b), the seed chip phenotype of eu2-c is 2, giving a faster reaction than eu2-b (Table 2). The eu2-c allele was found to contain a C to T mutation in position 4 of its coding sequence resulting in a stop in the second codon (Q2*). The C to T transition (CAA to TAA) is consistent with the mutagenic action of ethyl methane sulphonate.

If Ch02UreF is responsible for all UreF function, then lack of the Ch02UreF gene product in eu2-c, presuming efficient translatinal termination, is expected to result in a urease-null phenotype. However, not only is the seed chip phenotype of eu2-c the strongest of the three eu2 allelic mutants but, when quantified, eu2-c has higher urease specific activity than the eu2-a and eu2-b missense mutants in both seeds (mainly ESU) and seedling trifoliates (all UU) (Table 2).
two doses of eu2-b [Ch02UreF (G31D)] are enough to abolish urease activation almost completely. Within the Ch02UreF locus, however, eu2-b does not show any semi-dominance over Eu2, with respect to the final level of active urease, as shown among 24 F2:F3 sibs from the cross, eu2-b×Maverick (Fig. 3). The eight Eu2/Eu2 individuals showed no increase in seed urease activity compared with 16 Eu2/eu2-b siblings. There is no clear predictor, at this point, of a ‘pecking order’ among the various allelic forms of Ch02UreF for access to the urease activation complex.

Transcription patterns of chromosome 2 compared with 14 UreF paralogues

Our model asserts that Ch14UreF is partially functional in the eu2-c/eu2-c null background that lacks Ch02UreF. The model therefore predicts that Ch14UreF is expressed at the transcript and protein level. Ch14UreF transcript levels were measured by quantitative RT-PCR in wild-type (Williams 82) field-grown soybean employing template RNA from young trifoliate leaves and from immature seeds at two different stages. Figure 4 shows that transcripts from both UreF genes were equally represented, and that each was elevated ~10-fold in developing seeds compared with young trifoliate leaves—consistent with the reported higher levels of UreG protein in developing embryos compared with leaves (Freyermuth et al., 2000). Digital expression values (individual sequence reads) compiled for soybean ESTs (http://soybase.org/soyseq/) indicate that Ch14UreF (Glyma14g04380) transcripts are ~5-fold more common than Ch02UreF (Glyma02g44440) transcripts (across leaves, pods, embryos, roots, and nodules). Thus,
there is no indication that the Ch14UreF transcript is a minor one.

Discussion

The Illumina USLP1.0 platform was used to analyse the genetic basis of pleiotropic loss of soybean urease activities in two eu2 mutants. This strategy was chosen over sequence analysis of predicted UreD and UreF transcripts in eu2 mutant backgrounds because such a strategy would ‘miss’ mutation in a distinct gene, such as a Ni chaperone that has not yet been annotated in soybean or in any plant. In addition, sequence analysis of an expressed functional UreF (on chromosome 14) in eu2 backgrounds (Bacanamwo et al., 2002) seemed to preclude the possibility that the urease-negative phenotype of eu2 could be due to mutations in a potentially redundant paralogous gene. The ‘unbiased’ bulked segregant analysis pointed directly to Chr2UreF as the mutated gene on the gene/genomic assembly—a demonstration of the power of SNP mapping and bulked segregant analysis for identifying genes underlying phenotypes.

Table 3 summarizes current understanding of the urease structural and accessory genes of soybean. All are single copy except for UreD and UreF. These results suggest that Ch02UreF is necessary for full activation of each urease, the ESU and the UU isozymes. Only specific elimination of Ch02UreF will indicate whether it plays a significant role, although limited urease activity in the absence of Ch02UreF (in the eu2-c mutant) indicates that it does not. Certainly, there has not been a ‘specialization of function’ whereby each soybean UreF protein is involved in activation of one urease isozyme in all tissues, or of both ureases in specific tissues—missense mutation in Chr2UreF results in >99% loss of activity of the ESU and UU (Table 2), the latter in all tissues examined (Meyer-Bothling et al., 1987; Polacco et al., 1999). The ‘divergent’ Ch14UreF product has 95% deduced amino acid identity with that of Ch02UreF (225 of 238 amino acids are identical). None of the changes occur in conserved positions within the protein, and none was predicted to have a deleterious effect on the protein (Ng and Henikoff, 2001). Hence, we predict no radical changes relative to function in Ch14UreF. Indeed, a demonstration of the functionality of the Ch14UreF-encoded ORF is its partial correction of a ureF lesion in the fission yeast, S. pombe (Bacanamwo et al., 2002).

In the ‘spoiler’ hypothesis, whereby either of two missense forms of Ch02UreF proteins interferes with basal activation by Ch14UreF, neither lower levels of, nor lowered intrinsic urease activation by, Ch14UreF were invoked. Here, it is suggested that the Ch14UreF transcript [found at levels equal to that encoding Ch02UreF (Fig. 4)] is inefficiently translated, and that the two missense forms of Ch02UreF out-competes activation by Ch14UreF by preventing the less abundant Ch14UreF from binding to the activation complex. Recently, Ca et al. (2010) analysed intron predictions for the S’ leader of UreF and UreF paralogues from 11 plant species (the ORF is always intronless). All, with two exceptions, contained a S’ leader intron whose splicing removes AUG codons upstream of the start codon. One of the exceptions is the Ch14UreF of soybean: upon splicing, it retains an out-of-frame AUG codon upstream of the start AUG. The authors speculated that this paralogue is translated less efficiently than the Ch02UreF transcript. Quantification of the UreF proteins, aided by antibodies, will test this hypothesis of a lowered level of Ch14UreF relative to Ch02UreF protein.

That the missense eu2-b allele actively ‘spoils’ or interferes with activation by Ch14UreF is suggested by its ‘dominance’ with respect to the null eu2-c allele. eu2-b copy number is inversely related to urease activity level. Two ‘bad’ copies of Ch02UreF (G31D) are enough to abolish urease activation almost completely (Fig. 2). However, the eu2-b allele exhibits no ‘spoiler’ effects on its wild-type allele (Fig. 4). Among 24 F2:F3 sibs from the cross, eu2-b×Maverick, 8 Eu2/Eu2 and 16 Eu2/eu2-b individuals averaged urease specific activities (U mg protein\(^{-1}\)) of 29.4 (SD ± 8.7) and 27.3 (SD ± 3.4), respectively. The model makes no prediction of the ‘pecking order’ among the various allelic forms of Ch02UreF for access to the urease

Table 3. Urease genes\(^a\) of soybean, identified by mutational and phenotypic analyses (columns 1 and 2) and by genome sequence annotation (column 3). Matches of mutant and annotated loci are described in the references (last column).

| Genetic locus | Encodes: | Glyma no./annotation | Mutation disrupts urease activity of: | Refs |
|---------------|----------|----------------------|--------------------------------------|------|
| Eu1           | Urease (ESU) | Glyma05g27840/urease | ESU                                  | Meyer-Bothling and Polacco, 1987; Torisky et al., 1994 |
| Eu4           | Urease (UU) | Glyma11g37250/urease | UU                                   | Meyer-Bothling and Polacco, 1987; Torisky et al., 1994 |
| Eu3           | UreG      | Glyma08g08970/UreG   | ESU and UU                           | Freyermuth et al., 2000 |
| Eu2           | UreF      | Glyma02g44440/UreF   | ESU                                   | This work |
|               |           | Glyma14g04380/UreF   |                                       |      |
|               |           | Glyma02g0690/UreD    |                                       |      |
|               |           | Glyma20g17990/UreD   |                                       |      |

\(^a\) A ‘third’ urease in the genome (Glyma08g10850) is most likely non-ureolytic because it has deletions in the ORF, as well as a mutated residue essential for activity.
activation complex. However, the missense mutant versions of Ch02UreF are hypothesized to impede access of Ch14UreF to the urease activation complex thereby preventing activation.

Analysis of the soybean genome (Schmutz et al., 2009; <http://www.phytozone.net/soybean>) indicates that a segmental duplication event likely led to the Ch02UreF and Ch14UreF paralogues since they are on regions of shared synteny. UreF is postulated to bind to UreD on bacterial apo-urease and subsequently interact with and activate the UreG GTPase (Carter et al., 2009). UreD–urease interaction has been demonstrated in rice (Cao et al., 2010). Apparently, the stoichiometry of the activation complex subunits is critical since overexpression of UreD and UreF interferes with bacterial urease activation (Lee et al., 1992; Park et al., 1994). Possibly, there was selective pressure for lowered accumulation of UreF protein, manifested by alternatively, and apparently defectively, spliced transcripts of Ch14UreF. Witte et al. (2005) demonstrated in vitro activation of Arabidopsis urease by mixed extracts of Escherichia coli expressing apo-urease and two of three accessory proteins (i.e. UreD+UreF and UreD+UreG). The co-expression of apo-urease and UreD was essential for activation, thus reinforcing the role of UreD as an apo-urease chaperone.

There are two ureD paralogues in soybean (Table 3), and lack of urease-negative UreD lesions among the many urease-negative soybean mutants recovered suggests that both UreD genes are functional in urease activation. A prediction is that both are expressed at low levels, in agreement with multiple alternatively, and apparently defectively, spliced transcripts of UreD in Arabidopsis (Witte et al., 2005).

The tools are now available for soybean to be a model plant for better understanding of divergence of function of paralogues arising either from recent polyploidy or segmental duplication. These tools will help to optimize the combination of alleles of paralogous genes for soybean improvement.

Acknowledgements

We thank Paul Little for excellent technical assistance, Perry Cregan for encouragement in SNP mapping, Mike McMullen for providing invaluable advice on analysis of SNP data, and Doug Randall for lab space. MMS was supported by a predocotoral fellowship from CAPES, Brazil.

References

Bacanamwo M, Witte C-P, Lubbers MW, Polacco JC. 2002. Activation of the urease of Schizosaccharomyces pombe by the UreF accessory protein from soybean. Molecular Genetics and Genomics 268, 525–534.

Bailey C, Boulter JD. 1971. Urease, a typical seed protein of the Leguminosae. In Harborne JB, Boulter D, Turner BL, eds. The chemotaxonomy of the Leguminosae. New York, : Academic Press, 485–502.

Cao FQ, Werner AK, Dahncke K, Romeis T, Witte C- P. 2010. Identification and characterization of proteins involved in rice urea and arginine catabolism. Plant Physiology 154, 98–108.

Carlini CR, Polacco JC. 2008. Toxic properties of urease (perspectives). Crop Science 48, 1665–1672.

Carter EL, Flugga N, Boer JL, Mulrooney SB, Hausinger RP. 2009. Interplay of metal ions and urease. Metallomics 1, 207–221.

Eskew DL, Welch RM, Carey EE. 1983. Nickel: an essential micronutrient for legumes and possibly all higher plants. Science 222, 621–623.

Fahmy AS, Mohamed MA, Kamel MY. 1994. Ureas in the Cucurbitaceae, distribution and properties. Phytochemistry 35, 151–154.

Fan JB, Oliphant A, Shen R, et al. 2003. Highly parallel SNP genotyping. Cold Spring Harbor Symposium on Quantitative Biology 68, 69–78.

Freyermuth SK, Bacanamwo M, Polacco JC. 2000. The soybean Eu3 gene encodes a Ni-binding protein necessary for urease activity. The Plant Journal 21, 53–60.

Goldraij A, Beamer LJ, Polacco JC. 2003. Interallelic complementation at the ubiquitous urease coding locus of soybean. Plant Physiology 132, 1801–1810.

Goldraij A, Polacco JC. 1999. Arginase is inoperative in developing soybean seeds. Plant Physiology 119, 297–304.

Hyten DL, Song Q, Choi L-Y, Yoon M-S, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB. 2008. High-throughput genotyping with the Golden Gate assay in the complex genome of soybean. Theoretical and Applied Genetics 116, 945–952.

Kerr PS, Blevins DG, Rapp B, Randall DD. 1983. Soybean leaf urease, comparison with seed urease. Physiologia Plantarum 57, 339–345.

Lam R, Romanov V, Johns K, Batteke KP, Wu-Brown J, Guthrie JL, Hausinger RP, Pai EF, Chirgadze NY. 2010. Crystal structure of a truncated urease accessory protein UreF from Helicobacter pylori. Proteins: Structure, Function, and Bioinformatics 78, 2839–2848.

Lee M, Mulrooney SB, Renner MJ, Markowicz Y, Hausinger RP. 1992. Klebsiella aerogenes urease gene cluster: sequence of ureD and demonstration that 4 accessory genes (ureE, ureF, and ureG) are involved in nickel metallocenter biosynthesis. Journal of Bacteriology 174, 4324–4330.
Meyer-Bothling LE, Polacco JC. 1987. Mutational analysis of the embryo-specific urease locus of soybean. *Molecular and General Genetics* 209, 439–444.

Meyer-Bothling LE, Polacco JC, Cianzio SR. 1987. Pleiotropic soybean mutants defective in both urease isozymes. *Molecular and General Genetics* 209, 432–438.

Mulrooney SB, Ward SK, Hausinger RP. 2005. Purification and properties of the Klebsiella aerogenes UreE metal-binding domain, a functional metallochaperone of urease. *Journal of Bacteriology* 187, 3581–3585.

Muñoz A, Piedras P, Aguilar M, Pineda M. 2001. Urea is a product of ureidoglycolate degradation in chickpea. Purification and characterization of the ureidoglycolate urea-lyase. *Plant Physiology* 125, 828–834.

Muñoz A, Raso MJ, Pineda M, Piedras P. 2006. Degradation of ureidoglycolate in French bean (Phaseolus vulgaris) is catalysed by a ubiquitous ureidoglycolate urea-lyase. *Planta* 224, 175–184.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Research* 11, 863–874.

Park IS, Carr MB, Hausinger RP. 1994. In vitro activation of urease apoprotein and role of urea as a chaperone required for nickel metallocenter assembly. Proceedings of the National Academy of Sciences USA 91, 3233–3237.

Polacco JC, Freyermuth SK, Gerendás J, Cianzio SR. 1999. Soybean genes involved in nickel insertion into urease. *Journal of Experimental Botany* 50, 1149–1156.

Polacco JC, Havir EA. 1979. Comparisons of soybean urease isolated from seed and tissue culture. *Journal of Biological Chemistry* 254, 1707–1715.

Polacco JC, Holland MA. 1993. Roles of urease in plant cells. *International Review of Cytology* 145, 65–103.

Polacco JC, Judd AK, Dybing JK, Cianzio SR. 1989. A new mutant class of soybean lacks urease in leaves but not in leaf-derived callus or in roots. *Molecular and General Genetics* 217, 257–262.

Polacco JC, Krueger RW, Winkler RG. 1985. Structure and possible ureide degrading function of the ubiquitous urease of soybean. *Plant Physiology* 79, 794–800.

Polacco JC, Winkler. 1984. Soybean leaf urease, a seed enzyme? *Plant Physiology* 74, 800–803.

Salomone-Stagni M, Zambelli B, Musiani F, Ciurli S. 2007. A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis. *Proteins: Structure, Function, and Bioinformatics* 68, 749–761.

Schlueter JA, Dixon P, Granger CG, Grant D, Clark L, Doyle JJ, Shoemaker RC. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47, 868–876.

Schmutz J, Cannon SB, Schlueter J, et al. 2009. Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.

Stebbins NE, Holland MA, Cianzio SR, Polacco JC. 1991. Genetic tests of the roles of the embryonic ureases of soybean. *Plant Physiology* 97, 1004–1010.

Stebbins NE, Polacco JC. 1995. Urease is not essential for ureide degradation in soybean. *Plant Physiology* 109, 169–175.

Todd CD, Polacco JC. 2004. Soybean cultivars ‘Williams 82’ and ‘Maple Arrow’ produce both urea and ammonia during ureide degradation. *Journal of Experimental Botany* 55, 867–877.

Todd CD, Polacco JC. 2006. AtAAH, an allantoate amidohydrolase from Arabidopsis thaliana. *Planta* 223, 1108–1113.

Torisky RS, Griffin JD, Yenofsky RL, Polacco JC. 1994. A single gene (Eu4) encodes the tissue-ubiquitous urease of soybean. *Molecular and General Genetics* 242, 404–414.

Torisky RS, Polacco JC. 1990. Soybean roots retain the seed urease synthesized during embryo development. *Plant Physiology* 94, 681–689.

Tuteja JH, Clough SJ, Chan W-C, Vodkin LO. 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. *Plant Cell* 16, 819–835.

Werner AK, Romeis T, Witte C-P. 2009. Ureide catabolism in Arabidopsis thaliana and Escherichia coli. *Nature Chemical Biology* 6, 19–21.

Winkler RG, Bleivins DG, Polacco JC, Randall DD. 1987. Ureide catabolism of soybeans. II. Path of catabolism in intact leaf tissue. *Plant Physiology* 83, 585–591.

Winkler RG, Polacco JC, Bleivins DG, Randall DD. 1985. Enzymic breakdown of allantoate in developing soybeans. *Plant Physiology* 79, 787–793.

Witte C-P, Rosso MG, Romeis T. 2005. Identification of three urease accessory proteins that are required for urease activation in Arabidopsis. *Plant Physiology* 139, 1155–1162.