DOI: 10.1534/genetics.107.080366

Evidence on the Molecular Basis of the *Ac/ac* Adaptive Cyanogenesis Polymorphism in White Clover (*Trifolium repens* L.)

Kenneth M. Olsen,[1] Shih-Chung Hsu and Linda L. Small

Department of Biology, Washington University, St. Louis, Missouri 63130-4899

Manuscript received August 10, 2007
Accepted for publication February 20, 2008

ABSTRACT

White clover is polymorphic for cyanogenesis, with both cyanogenic and acyanogenic plants occurring in nature. This chemical defense polymorphism is one of the longest-studied and best-documented examples of an adaptive polymorphism in plants. It is controlled by two independently segregating genes: *Ac/ac* controls the presence/absence of cyanogenic glucosides; and *Li/li* controls the presence/absence of their hydrolyzing enzyme, linamarase. Whereas *Li* is well characterized at the molecular level, *Ac* has remained unidentified. Here we report evidence that *Ac* corresponds to a gene encoding a cytochrome P450 of the CYP79D protein subfamily (*CYP79D15*), and we describe the apparent molecular basis of the *Ac/ac* polymorphism. CYP79D orthologs catalyze the first step in cyanogenic glucoside biosynthesis in other cyanogenic plant species. In white clover, Southern hybridizations indicate that *CYP79D15* occurs as a single-copy gene in cyanogenic plants but is absent from the genomes of *ac* plants. Gene-expression analyses by RT–PCR corroborate this finding. This apparent molecular basis of the *Ac/ac* polymorphism parallels our previous findings for the *Li/li* polymorphism, which also arises through the presence/absence of a single-copy gene. The nature of these polymorphisms may reflect white clover's evolutionary origin as an allotetraploid derived from cyanogenic and acyanogenic diploid progenitors.

CYANOGENESIS (cyanide release following tissue damage) occurs in >2650 plant species, including ferns, gymnosperms, monocots, and dicots (Seigler and Brinker 1993). In its most common form, cyanogenesis involves the interaction of two compounds, cyanogenic glycosides and their hydrolyzing enzymes; these are separated in intact tissue and brought into contact with tissue disruption. White clover (*Trifolium repens* L., Fabaceae) is polymorphic for cyanogenesis, with both cyanogenic and acyanogenic plants occurring in natural populations (Armstrong *et al.* 1913; Ware 1925). This polymorphism arises through the presence/absence of both of the underlying cyanogenic components. Inheritance of cyanogenesis in white clover follows a simple Mendelian two-locus, two-allele model (Coop 1940; Melville and Doak 1940; Corkill 1942): the gene *Ac/ac* controls the presence/absence of cyanogenic glucosides; and the independently segregating gene *Li/li* controls the presence/absence of linamarase, a cyanogenic β-glucosidase (reviewed by Hughes 1991). Plants that carry at least one dominant (functional) allele at both genes (*Ac* and *Li*) are cyanogenic, while the occurrence of two nonfunctional alleles at either gene confers the acyanogenic phenotype.

White clover is a common legume of fields, lawns, and pastures in mesic temperate climates. It is a native of Eurasia but has been introduced worldwide as a forage crop and lawn plant. It is a perennial and an insect-pollinated, obligate outcrosser; plants also spread vegetatively by stolons. Chromosome number and genetic map data indicate that white clover is an allotetraploid (*e.g.*, Barrett *et al.* 2004). While *Ac* and *Li* have not been mapped, both cyanogenesis genes are believed to be present in only one of the two parental genomes (*e.g.*, Williams and Williamson 2001); this suggests that the species originated through the hybridization of a cyanogenic and an acyanogenic Trifolium species (Williams and Williamson 2001; Badr *et al.* 2002).

The ecological genetics of white clover cyanogenesis has been studied for >60 years, and this system represents one of the best-documented examples of an adaptive polymorphism in plants (reviews by Hughes 1991; Hayden and Parker 2002; Olsen *et al.* 2007). Cyanogenic plants are generally found to be strongly favored in the presence of generalist herbivores, which avoid eating them (*e.g.*, Ennos 1981; Dirzo and Harper 1982a,b; Pederson and Brink 1998; Saucy *et al.* 1999; Viette *et al.* 2000). The factors favoring acyanogenic plants appear to be tied to resource allocation trade-offs and differential fitness in cool environments. Populations of white clover show strong clinal variation in cyano-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU370164–EU370205.

[1]*Corresponding author:* Department of Biology, Campus Box 1137, Washington University, 1 Brookings Dr., St. Louis, MO 63130-4899. E-mail: kolsen@wustl.edu

genesis, with frequencies of cyanogenic plants closely correlated with minimum winter temperatures in both the native and the introduced range (*e.g.*, Daday 1954a,b, 1958; De Araujo 1976; Till-Bottraud *et al.* 1988; Ganders 1990; Majumdar *et al.* 2004). Comparisons of cyanogenic and acyanogenic plants have revealed trade-offs between cyanogenesis and several fitness-related traits, including flowering rate (Daday 1965; Foulds and Grime 1972; Dirzo and Harper 1982b; Kakes 1989), drought tolerance (Foulds and Grime 1972), and resistance to rust (Dirzo and Harper 1982b). Thus, the predominance of acyanogenic plants in colder climates may reflect differential competitive ability under these conditions, possibly as a function of decreased herbivore abundance (Kakes 1989).

Understanding the genetic basis of adaptation is a major focus of modern evolutionary biology, and the white clover cyanogenesis polymorphism offers an attractive system for addressing this question. The simple Mendelian inheritance of the polymorphism makes it tractable for examining its molecular genetic basis and molecular evolution. At the same time, unlike polymorphisms in many model genetic systems (*e.g.*, *Arabidopsis thaliana*, crop species), the ecology of this polymorphism is very well studied; >50 articles on white clover cyanogenesis have been published since the 1940s, and the system has become a textbook example of variation maintained by opposing selective forces (*e.g.*, Dirzo and Sarukhan 1984; Silvertown and Charlesworth 2001). This ecological foundation makes it possible to more fully understand the ecological and evolutionary relevance of underlying molecular genetic variation.

In a previous study (Olsen *et al.* 2007), we examined the molecular genetic basis and population genetics of the *Li/li* polymorphism, which is responsible for the presence/absence of linamarase. In the present study, we extend our molecular evolutionary analysis to the *Ac* gene, which heretofore has remained unidentified at the molecular level. We address (1) the molecular identity of *Ac*; (2) the molecular genetic basis of the *Ac/ac* polymorphism, underlying the presence/absence of cyanogenic glucosides in white clover plants; and (3) the molecular population genetics of *Ac*, as inferred through comparisons of DNA sequences at this gene and three unlinked, neutrally evolving genes.

Our analysis of *Ac/ac* is based on a well-established foundation of information on plant cyanogenic glucoside biosynthesis. This pathway has been studied extensively in several species, including the model legume *Lotus japonicus* (Forslund *et al.* 2004; reviewed by Bak *et al.* 2006), which is in the same subfamily as white clover. Plant cyanogenic glucosides are synthesized from amino acids in a highly channeled set of reactions catalyzed by two membrane-bound cytochrome P450's and a soluble glycosyltransferase (Forslund *et al.* 2004; Bak *et al.* 2006). These three proteins are believed to form a metabolon (multienzyme complex) that acts to channel the otherwise toxic and highly reactive intermediates through the pathway to the formation of the end product (Bak *et al.* 2000, 2006; Kristensen *et al.* 2005). Cyanogenic white clover plants (as well as *L. japonicus*, cassava, lima bean, flax, and rubber tree) synthesize two closely related cyanogenic glucosides: linamarin (1-cyano-1-methylethyl β-d-glucopyranoside) and lotaustralin (R-1-cyano-1-methylpropyl β-d-glucopyranoside); these are generated from valine and isoleucine, respectively (Butler and Butler 1960). A single metabolon is responsible for the synthesis of both forms of cyanogenic glucoside (Collinge and Hughes 1984; Forslund *et al.* 2004). White clover plants that possess only nonfunctional *ac* alleles are unable to synthesize either linamarin or lotaustralin (Corkill 1942; Hughes and Conn 1976).

The highly reactive nature of intermediates in the cyanogenic glucoside biosynthesis pathway suggests that the *ac* phenotype would be very unlikely to arise through the disruption of an intermediate step in the pathway, as this would lead to the accumulation of toxic intermediates (Bak *et al.* 1999, 2000; Kristensen *et al.* 2005). In *L. japonicus* and cassava, the first step in the pathway is catalyzed by a cytochrome P450 belonging to the CYP79D protein subfamily (Andersen *et al.* 2000; Forslund *et al.* 2004). Therefore, our hypothesis in conducting the present study was that a gene encoding a white clover CYP79D protein (or, alternatively, a gene controlling the expression of this protein) is responsible for the *Ac/ac* polymorphism.

MATERIALS AND METHODS

Samples: White clover's current cosmopolitan distribution represents a recent and rapid population expansion, both in the Old World (following the development of agriculture within the last 10,000 years) and worldwide (following European colonization within the last 500 years). Seeds from a geographically diverse sample of wild *T. repens* populations were obtained from the white clover germ plasm collection maintained by the U.S. Department of Agriculture (http://www.ars-grin.gov). Forty-two white clover samples were included in the study, representing 26 *Ac* plants (synthesizing cyanogenic glucosides) and 16 *ac* plants (lacking cyanogenic glucosides) (Table 1). Seeds were lightly scarified, germinated, and grown in standard greenhouse conditions at Washington University. A single plant was grown per U.S. Department of Agriculture (USDA) accession number. Two related, cyanogenic Trifolium species were also included for use as outgroups in DNA sequence analyses (Table 1).

Plants were assayed for the presence/absence of cyanogenic constituents using a modified Feigl–Anger hydrogen cyanide (HCN) assay (Feigl and Anger 1966), as described by Olsen *et al.* (2007). The presence/absence of functional alleles at *Ac* and *Li* can be inferred independently through the exogenous addition of either linamarase or linamarin, respectively, to an acyanogenic leaf sample (see Olsen *et al.* 2007). Liberated cyanide is normally detectable in the Feigl–Anger assay within 1 hr of leaf tissue incubation with test reagents. Because we

observed wide quantitative variation in cyanogenic glucoside content among *Ac* plant samples (discussed below), all leaf samples were incubated with HCN test reagents for ≥3 hr to ensure detection of all *Ac* genotypes.

PCR amplification and DNA sequencing: DNA was extracted from fresh leaf tissue using a modified CTAB extraction protocol (Porebski *et al.* 1997). In *L. japonicus*, two independent cyanogenic glucoside biosynthetic pathways operate in roots and in aerial parts of the plant; each pathway uses a different CYP79D enzyme (Forslund *et al.* 2004). The genes encoding these two enzymes have been cloned (*CYP79D3*, *CYP79D4*; Forslund *et al.* 2004), and comparison of their cDNA sequences reveals that they are 96% identical. We designed degenerate primers from these *L. japonicus CYP79D* genes to PCR amplify any *T. repens* homologs. Initial PCR amplification was performed using the following primers: 5′-TGGACTTTTTTGCTTGTTGTGATATT-3′ and 5′-GCAGC CAATCTTGGTTTTGC-3′, located at the 5′ and 3′ ends, respectively, of the *L. japonicus* full-length cDNAs (see supplemental Tables S1 and S2). We subsequently designed several additional primers from *T. repens* sequences for use in PCR and DNA sequencing (supplemental Table S1).

PCR was performed with 20-μl volumes and standard reaction conditions, using GoTaq DNA polymerase (Promega, Madison, WI). Annealing temperatures were adjusted for primer combinations. PCR products were cloned into plasmids using TA cloning (Invitrogen, Carlsbad, CA), purified, and then sequenced. A minimum of eight clones were sequenced per PCR product to detect allelic variation and to allow detection of PCR artifacts. Singletons unique to a single sequenced clone were considered artifacts of polymerase error and disregarded. PCR using *CYP79D* primers consistently amplified a single product in *Ac* plants, which DNA sequencing confirmed to be a *T. repens CYP79D* ortholog. This *T. repens* gene is hereafter referred to as *CYP79D15*, in accordance with cytochrome P450 nomenclatural convention (http://drnelson.utmem.edu/cytochromeP450.html).

CYP79D15 was sequenced in 22 *Ac* plants. For outgroup comparison in DNA sequence analyses, orthologs of *CYP79D15* were also amplified and sequenced in two closely related cyanogenic clover species: *T. nigrescens* ssp. *petrisavii* and *T. isthmocarpum* (Table 1). All DNA sequencing was performed using BigDye terminators (Applied Biosystems, Foster City, CA), with reactions run on an ABI 3130 sequencer in the Biology Department of Washington University. No more than two sequence haplotypes (alleles) were detected per individual, an indication that paralogous or homeologous gene copies were not amplified or included in sequence analyses.

Determination of gene copy number: We performed Southern hybridizations to assess *CYP79D15* copy number in *Ac* and *ac* plants and to detect any other genes sharing sequence similarity with *CYP79D15*. Twenty-two *Ac* plants and 16 *ac* plants were sampled for Southern hybridizations (Table 1). Genomic DNA was purified as described above, and ~1 μg DNA was digested with *Afl*III or *Ase*I and run on a 0.6% Seakem (Fisher Scientific, Pittsburgh) agarose gel. The DNA was then transferred to a nitrocellulose membrane using standard methods. A probe corresponding to the middle portion of *CYP79D15* (892 bp, spanning 52% of the gene) was prepared using the DIG probe synthesis kit (Roche, Indianapolis); hybridization washes were performed at high stringency following the manufacturer's protocol. Primers used in PCR amplification of probes are listed in supplemental Table S1. On the basis of the DNA sequence of *CYP79D15*, *Ase*I is predicted to lack restriction sites within the gene; *Afl*III is predicted to have two restriction sites within the gene, one of which occurs within the probed region.

Gene-expression analysis: To assess *CYP79D15* expression in *Ac* and *ac* plants, total RNA was extracted from fresh young leaf tissue of four *Ac* and four *ac* plant samples (Table 1), using RNeasy kits (QIAGEN, Valencia, CA). Reverse transcription and RT–PCR were performed using a Thermoscript RT–PCR kit (Invitrogen). The *CYP79D15* primers that were used to amplify a probe for Southern hybridizations were also used in RT–PCR. These primers are predicted to amplify a 752-bp portion of cDNA corresponding to the 892-bp genomic DNA sequence. As a positive control for RT–PCR, primers were designed to amplify a 151-bp region of 5.8S ribosomal RNA. All RT–PCR primers are listed in supplemental Table S1. PCR conditions were similar to those for amplifying genomic DNA as described above, with annealing temperatures adjusted for primer combinations. The identities of all RT–PCR products were confirmed by DNA sequencing.

DNA sequence analyses: DNA sequences from *CYP79D15* were edited and aligned visually using Biolign software (Hall 2001). The position of a single intron within the gene was inferred by comparison to published *L. japonicus CYP79D* cDNA sequences (see supplemental Table S2). Molecular population genetic analyses were conducted using DnaSP 4.10 (Rozas *et al.* 2003). Levels of nucleotide diversity per silent site were estimated as π (Nei 1987) and θ_W (Watterson 1975). Tests of selection were performed using Tajima's (1989) *D* and Fay and Wu's (2000) *H*-tests, with statistical significance assessed by coalescent simulations (1000 replicates, using θ_W and levels of recombination estimated empirically from the data) and by the McDonald–Kreitman (McDonald and Kreitman 1991) test.

Diversity measures and tests of selection for *CYP79D15* were compared to previously published data sets for three other *T. repens* loci (*ACO1*, *ALDP*, *ZIP*; Olsen *et al.* 2007). These three genes have been sequenced in 18–20 white clover samples from across the species range; approximately half of the plant samples sequenced at each of the genes were also included in the present study (see Olsen *et al.* 2007). Sequences from all three previously published genes conform to neutral equilibrium expectations in tests of selection and are considered here to be representative of neutrally evolving white clover genes. Analyses of population structure using these three neutral loci have revealed no evidence of geographical structuring of genetic variation across the present species range (Olsen *et al.* 2007), a pattern consistent with the recent and rapid dispersal of this cosmopolitan, human-associated plant.

RESULTS

***CYP79D15* gene structure and copy number:** Degenerate PCR primers designed from the ends of *L. japonicus CYP79D* full-length cDNAs amplified a single 1.7-kb product in white clover plants that possess cyanogenic glucosides (*Ac* plants). The structure of the amplified gene is shown in Figure 1. Exon regions are 92–93% identical to two previously identified *L. japonicus CYP79* genes (*CYP79D3*, *CYP79D4*; Forslund *et al.* 2004), and the inferred amino acid sequence is 89–90% identical. Following the conventions of the cytochrome P450 nomenclature committee (http://drnelson.utmem.edu/cytochromeP450.html), the white clover gene has been designated *CYP79D15*. Whereas *L. japonicus* expresses two tissue-specific isoforms of CYP79D, with one (CYP79D3) expressed exclusively in aerial parts and the other (CYP79D4) expressed in roots (Forslund

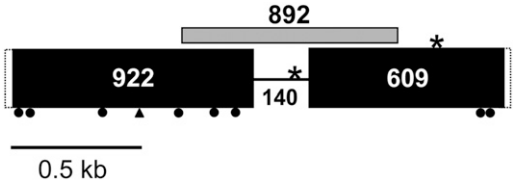

FIGURE 1.—Structure of the *CYP79D15* gene. Dotted rectangles at 5′ and 3′ ends of the gene indicate the boundaries of the full-length cDNA (start codon to stop codon) as inferred by alignment with *Lotus japonicus CYP79D* cDNAs (see supplemental Table S2); unshaded regions at the 5′ and 3′ ends were not included in sequence analyses. Solid rectangles represent sequenced exon regions, and the intervening line corresponds to the single intron. Sizes of exons and the intron are approximately to scale; numbers indicate aligned nucleotide lengths of each sequenced gene region. Solid dots indicate approximate positions of nonsynonymous substitutions; the triangle indicates the approximate location of a four- to five-codon indel (see Table 2). Asterisks indicate predicted *Afl*III restriction sites. The shaded bar indicates the gene region targeted for Southern hybridizations and RT–PCR.

et al. 2004), only a single gene encoding a CYP79D protein was detected in white clover by PCR; this result is also supported by Southern hybridizations (discussed below). White clover synthesizes cyanogenic glucosides exclusively in shoot growth (COLLINGE and HUGHES 1982b); therefore, the apparent absence of multiple tissue-specific CYP79D isoforms is not unexpected.

Whereas *CYP79D15* was consistently amplified in *Ac* plants (possessing cyanogenic glucosides), we were unable to amplify any portion of the gene in *ac* plants, despite repeated attempts using multiple primer combinations (see supplemental Table S1) in 16 different *ac* genotypes. We therefore performed Southern hybridizations to assess *CYP79D15* copy number and to detect any related genes, using a sample of 22 *Ac* and 16 *ac* plants (Table 1). Figure 2 shows hybridizations for a representative sample of *Ac* and *ac* plants digested with *Ase*I, which has no predicted restriction sites within *CYP79D15*. A single strong band is observed in *Ase*I digests of *Ac* plants, consistent with the occurrence of *CYP79D15* as a single-copy gene in *Ac* plants. In contrast, no bands are detectable for the *ac* samples, indicating the absence of sequences with close sequence similarity to the *CYP79D15* probe. Similarly, *Afl*III digests, which are predicted to generate two bands for a single-copy gene, reveal two bands in *Ac* plants and no bands in *ac* plants (supplemental Figure S1). These same banding patterns were observed consistently in all *Ac* and *ac* plants examined (data not shown). Thus, for this worldwide sample of white clover plants, the *Ac* phenotype is consistently associated with the occurrence of *CYP79D15* as a single-copy gene, and the *ac* phenotype is associated with the absence of much or all of the gene from the white clover genome. In addition, the lack of hybridization to more than one gene copy (Figure 2) indicates that there is no evidence for any additional *T. repens* genes with close sequence similarity to *CYP79D15*.

***CYP79D15* expression in *Ac* and *ac* plants:** RT–PCR primers were designed to amplify a 752-bp portion of *CYP79D15* cDNA, corresponding to a genomic DNA region of 892 bp in the center of the gene (see Figure 1). *CYP79D15* transcripts were detected in all four of the *Ac* plants examined and not in any of the four *ac* plants (Figure 3); DNA sequencing of the *Ac* RT–PCR products confirmed that they correspond to targeted exon portions of *CYP79D15*. The absence of *CYP79D15* expression in *ac* plants is consistent with the apparent absence of much or all of the gene from the genomes of *ac* plants (see above; Figure 2).

DNA sequence variation: *CYP79D15* was sequenced in 22 *Ac* plants representing a geographically widespread sample of the species range (Table 1). Previous analyses of neutral genetic variation have indicated no evidence of population substructure in white clover (OLSEN *et al.* 2007; see above). Fourteen substitution polymorphisms were observed at *CYP79D15* (5 synonymous; 8 nonsynonymous, encoding seven amino acid replacements; 1 noncoding); in addition, indels were observed both in the intron (seven regions of indels, ranging in size from 1 to 28 bp) and in exon 1, where an in-frame indel codes for insertions of four- or five-amino-acid residues downstream of Asp164 (Table 2). Observed polymorphisms characterize a total of 15 *CYP79D15* sequence haplotypes (Tables 1 and 2). No frameshift mutations, premature stop codons, or other polymorphisms were observed that would be predicted to lead to a loss of gene function. Tests of selection do not reveal any statistically significant deviations from neutral equilibrium (Tajima's $D = -0.8514$, $P > 0.1$; Fay and Wu's $H = 1.59$, $P > 0.1$; McDonald–Kreitman test, $P > 0.1$ by Fisher's exact test). This result is observed regardless of whether all *Ac* plants are treated as *AcAc* homozygotes (*i.e.*, possessing two *CYP79D15* sequences; test results reported here) or whether the accessions that possess only one *CYP79D15* haplotype (six accessions; see Table 1) are treated as hemizygotes (*i.e.*, contributing only one *CYP79D15* sequence; test results not shown).

CYP79D15 nucleotide diversity estimates are comparable between nonsynonymous sites ($\pi = 0.0013$, $\theta_W = 0.0016$) and silent sites ($\pi = 0.0013$, $\theta_W = 0.0029$). Comparison of these values with those of three previously examined neutral genes (*ACO1*, *ALDP*, *ZIP*; OLSEN *et al.* 2007) suggests at face value that *CYP79D15* has much lower levels of genetic diversity; mean silent-site nucleotide diversity for the three previously published loci is $\pi = 0.0142 \pm 0.0061$ and $\theta_W = 0.0113 \pm 0.0056$. The lower nucleotide diversity at *CYP79D15* is partly a reflection of overlapping indel and substitution polymorphisms within the intron, which creates an underestimate of nucleotide variation; recoding indels as single-character substitutions increases the estimates

TABLE 1

Trifolium samples used in analyses

Trifolium species	USDA accession	Country of origin	*Ac/ac* phenotype	*Li/li* phenotype	*CYP79D15* haplotype
T. repens	PI 100247[a,b]	New Zealand	*Ac*	*li*	2/14
T. repens	PI 200372[a,b,c]	Israel	*Ac*	*Li*	2/4
T. repens	PI 204930[a,b,c]	Turkey	*Ac*	*Li*	9/10
T. repens	PI 205062[a,b]	Turkey	*Ac*	*li*	11/12
T. repens	PI 214207[a,b]	Israel	*Ac*	*Li*	1/12
T. repens	PI 217444[a,b]	Italy	*Ac*	*Li*	1/12
T. repens	PI 221961[a,b]	Afghanistan	*Ac*	*li*	13/13
T. repens	PI 226996[a,b]	Uruguay	*Ac*	*Li*	1/3
T. repens	PI 230183[b]	Argentina	*Ac*	*Li*	14/15
T. repens	PI 234678[b]	France	*Ac*	*Li*	1/5
T. repens	PI 239977[a,b,c]	Portugal	*Ac*	*Li*	1/4
T. repens	PI 246751[a,b]	Spain	*Ac*	*Li*	1/1
T. repens	PI 260646[a,b]	Greece	*Ac*	*Li*	1/4
T. repens	PI 291828[a]	Chile	*Ac*	*Li*	—
T. repens	PI 294546[a,b]	France	*Ac*	*Li*	2/2
T. repens	PI 298485[a,b,c]	Israel	*Ac*	*Li*	1/1
T. repens	PI 302441[a,b]	Australia	*Ac*	*li*	1/12
T. repens	PI 311490[a]	Spain	*Ac*	*Li*	—
T. repens	PI 311494[a,b]	Spain	*Ac*	*Li*	1/12
T. repens	PI 315542[b]	Russia	*Ac*	*li*	5/5
T. repens	PI 345529[a,b]	Australia	*Ac*	*Li*	7/8
T. repens	PI 440745[a,b]	Russia	*Ac*	*li*	13/13
T. repens	PI 440746[a]	Russia	*Ac*	*li*	—
T. repens	PI 499685[a]	China	*Ac*	*li*	—
T. repens	PI 499688[a,b]	China	*Ac*	*li*	13/13
T. repens	PI 597530[b]	Lithuania	*Ac*	*li*	1/6
T. repens	PI 195534[a]	Italy	*ac*	*li*	—
T. repens	PI 208730[a]	Italy	*ac*	*li*	—
T. repens	PI 232109[a,c]	Germany	*ac*	*li*	—
T. repens	PI 251053[a]	Macedonia	*ac*	*li*	—
T. repens	PI 251190[a]	Serbia	*ac*	*li*	—
T. repens	PI 251191[a]	Yugoslavia	*ac*	*li*	—
T. repens	PI 251197[a,c]	Bosnia	*ac*	*li*	—
T. repens	PI 253323[a]	Slovenia	*ac*	*Li*	—
T. repens	PI 282378[a,c]	Italy	*ac*	*li*	—
T. repens	PI 419314[a]	Greece	*ac*	*li*	—
T. repens	PI 494747[a]	Romania	*ac*	*li*	—
T. repens	PI 516411[a]	Romania	*ac*	*li*	—
T. repens	PI 542904[a]	Croatia	*ac*	*li*	—
T. repens	PI 542905[a]	Croatia	*ac*	*Li*	—
T. repens	PI 542915[a]	Bosnia	*ac*	*li*	—
T. repens	PI 556991[a,c]	United States	*ac*	*li*	—
T. nigrescens ssp. *petrisavii*	PI 120103[b]	Turkey	*Ac*	*Li*	—
T. nigrescens ssp. *petrisavii*	PI 298478[b]	Israel	*Ac*	*Li*	—
T. isthmocarpum	PI 422595[b]	Morocco	*Ac*	*Li*	—

CYP79D15 haplotype labels correspond to those in Table 2.
[a] Analyzed by Southerns.
[b] Analyzed by *CYP79D15* DNA sequencing.
[c] Analyzed by RT–PCR.

of nucleotide diversity ($\pi = 0.0053$, $\theta_W = 0.0081$). In addition, because *CYP79D15* is absent in *ac* alleles (see above), the smaller effective population size of this gene must also be taken into account in comparisons with other autosomal nuclear loci. If one assumes that *Ac* and *ac* alleles occur at roughly equal frequencies within the species (see, *e.g.*, Daday 1958), such that the effective population size of *CYP79D15* is roughly half that of the other nuclear genes, then the recalibrated estimate of silent-site nucleotide diversity ($\pi = 0.01$) is comparable to that of the three previously examined neutral genes. Together with the nonsignificant tests of selection, this pattern suggests that *CYP79D15* is evolving in a manner consistent with neutral equilibrium expectations.

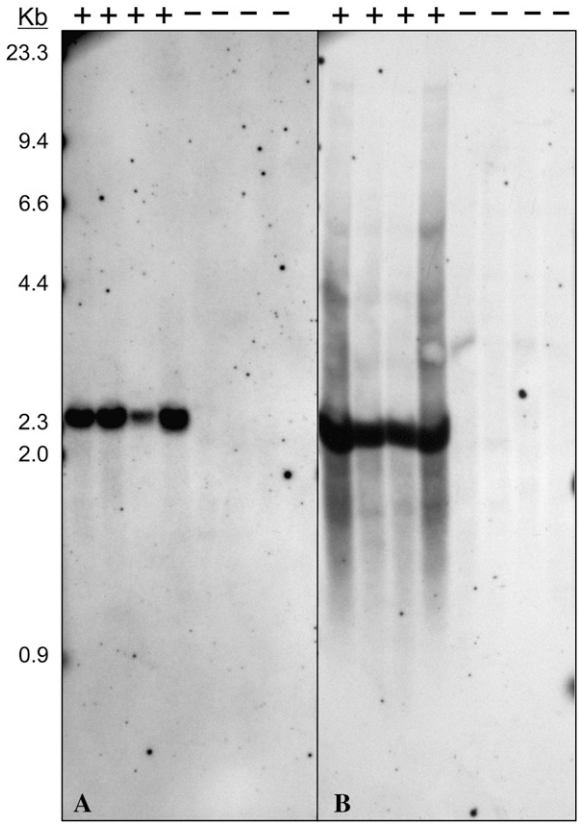

Figure 2.—Southern hybridizations of *Ac* and *ac* white clover samples digested with *Ase*I and hybridized with a *CYP79D15*-specific probe. *Ac* plants (possessing cyanogenic glucosides) and *ac* plants (lacking cyanogenic glucosides) are labeled + and −, respectively. The *CYP79D15* probe corresponds to an 892-bp portion of genomic DNA sequence spanning approximately half of the gene (see Figure 1). *Ase*I has no predicted restriction sites within *CYP79D15*. The two hybridizations shown (A and B) include partial replication of samples, as follows: (A) *Ac* accessions (left to right) 260646, 291828, 302441, 311494 and *ac* accessions 251053, 251190, 542915, 556991; (B) *Ac* accessions 246751, 260646, 291828, 302441 and *ac* accessions 195534, 232019, 516411, 556991.

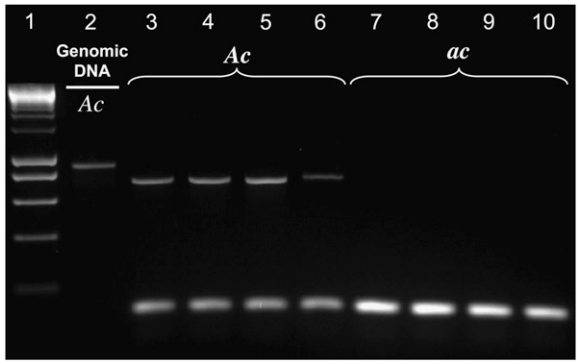

Figure 3.—Expression of *CYP79D15* in *Ac* and *ac* white clover plants. The targeted gene region is indicated in Figure 1. Lane 1 (ladder), first five bands from bottom: 200 bp, 400 bp, 600 bp, 800 bp, 1000 bp. Lane 2 (genomic DNA control): *CYP79D15* PCR from genomic DNA of *Ac* accession PI 239977; the predicted product size is 892 bp. Lanes 3–10: *CYP79D15* and 5.8S rDNA PCR products from cDNA of *Ac* accessions (lanes 3–6, accessions 239977, 202372, 298485, 204930) and *ac* accessions (lanes 7–10, accessions 556991, 232109, 251197, 282378); the predicted product sizes are 752 bp for *CYP79D15* (corresponding to 892 bp minus the 140-bp intron) and 151 bp for the 5.8S rDNA.

DISCUSSION

***CYP79D15* and the *Ac/ac* polymorphism:** The white clover cyanogenesis polymorphism was first identified >90 years ago (Armstrong *et al.* 1913; Ware 1925), and for the last half century this system has served as a model for understanding the ecological genetics of adaptive polymorphism in natural populations (reviewed by Hughes 1991; Hayden and Parker 2002). Although *Li* and its product have been well characterized at the nucleotide and protein levels (Oxtoby *et al.* 1991; Barrett *et al.* 1995; Olsen *et al.* 2007), the molecular identity of *Ac* has remained undetermined. Below we discuss evidence that *Ac* is very likely to be the cytochrome P450 gene *CYP79D15*.

1. *Ac/ac* is predicted to occur at the first step in cyanogenic glucoside biosynthesis. The biosynthesis of cyanogenic glucosides occurs in a highly channeled set of reactions catalyzed by three proteins, and interruption of the pathway at points other than the first dedicated step leads to the accumulation of reactive, toxic intermediates (reviewed by Bak *et al.* 2006). This has been demonstrated elegantly in metabolic engineering experiments, in which acyanogenic species (*A. thaliana*, *Nicotiana tabacum*) have been transformed to express components of the cyanogenic glucoside biosynthesis pathway from *Sorghum bicolor* (Bak *et al.* 1999, 2000; Kristensen *et al.* 2005). When *A. thaliana* is transformed to express the entire *S. bicolor* cyanogenic glucoside pathway, it is morphologically normal, while accumulating 4% dry-weight cyanogenic glucosides; in contrast, plants expressing only the first two steps in the pathway show severe stunting and metabolic evidence of detoxification reactions (Kristensen *et al.* 2005). Similarly, transgenic *N. tabacum* plants expressing only the first step of the *S. bicolor* pathway show evidence of intoxication from reactive intermediates (Bak *et al.* 2000). These transgenic studies strongly suggest that for white clover, the first dedicated step in the pathway would be most likely to account for the *Ac/ac* polymorphism.

Additional support for this hypothesis comes from *in vivo* and *in vitro* labeling experiments conducted in white clover prior to the development of the current metabolon model (Hughes and Conn 1976; Collinge and Hughes 1982a). These studies indicate that *ac* plants entirely lack the first step in the pathway while showing weak but measurable activity in subsequent steps (*e.g.*, Collinge and Hughes 1982a).

TABLE 2

Nucleotide substitution polymorphisms and coding-region indels at *CYP79D15*

		Exon 1									Intron:	Exon 2				
Haplotype	No.	29[a]	59[a]	280	319	384[a]	497[b]	746[a]	860[a]	903[a]	933	1385	1400	1517	1638[a]	1639[a]
1	(13)	T	C	T	T	G	—	G	A	T	A	A	C	T	C	C
2	(4)	.	.	.	.	.	—	.	.	A	.	.	.	.	.	.
3	(1)	.	.	.	A	.	—	.	.	.	.	.	.	.	.	.
4	(3)	.	.	.	A	.	—	.	.	A	.	.	.	.	.	.
5	(3)	.	.	.	A	.	—	.	.	A	.	.	T	.	A	T
6	(1)	.	.	.	A	.	—	.	.	A	.	.	T	.	.	.
7	(1)	.	.	.	A	A	5	A	.	A	.	G	.	.	.	.
8	(1)	A	T	.	A	A	5	.	.	A	.	G	.	.	.	.
9	(1)	.	.	.	A	A	4	.	.	A	.	.	.	.	A	T
10	(1)	.	.	.	A	A	4	.	.	A	T	.	.	.	A	T
11	(1)	.	.	.	A	A	4	.	.	A	T	.	.	C	A	T
12	(5)	.	.	.	A	.	4	.	.	A	.	.	.	.	.	.
13	(6)	.	.	.	A	.	4	.	.	A	T	.	.	.	A	T
14	(2)	.	.	.	A	.	4	.	.	.	.	.	.	.	.	.
15	(1)	.	.	C	A	.	4	.	G	.	.	.	.	.	.	.

Numbers in parentheses indicate the number of sequences observed per haplotype. Dots indicate that a nucleotide is identical to that of haplotype 1. Nucleotide positions 1638 and 1639 correspond to a single codon.

[a] Nonsynonymous polymorphisms.

[b] 4, four-codon insertion after Asp164 (Val-Asn-Asp-Asp); 5, five-codon insertion after Asp164 (Val-Asn-Asp-Asp-Asp).

2. *CYP79D15* catalyzes the first step in white clover cyanogenic glucoside biosynthesis. Our data strongly suggest that this first step in white clover cyanogenic glucoside biosynthesis is catalyzed by the gene product of *CYP79D15*. CYP79D proteins catalyze the first step in this pathway in both of the linamarin/lotaustralin-producing species that have been examined previously (cassava and *L. japonicus*) (Andersen *et al.* 2000; Forslund *et al.* 2004). In white clover, PCR using degenerate primers from *L. japonicus* amplifies a single gene belonging to the CYP79D protein subfamily (Figures 1 and 3). This gene, *CYP79D15*, shows very high sequence similarity to *L. japonicus* CYP79D orthologs, not only at the inferred amino acid level (≥89% identical), but also at the nucleotide level (≥92% identical) (see supplemental Table S2); thus, there is little question as to its placement within the CYP79D protein subfamily (see criteria described by Nelson 2006). Moreover, PCR amplifications and Southern hybridizations both indicate that there are no other genes present in the white clover genome with close sequence similarity to *CYP79D15* (Figures 1 and 2). Taken together, these observations suggest that *CYP79D15* is the only obvious candidate to catalyze the first step in white clover cyanogenic glucoside biosynthesis and as such would be the most likely candidate to underlie the *Ac/ac* cyanogenic glucoside polymorphism.
3. *CYP79D15* molecular variation corresponds to the *Ac/ac* phenotype. Additional evidence that *CYP79D15* corresponds to *Ac* comes from our analysis of *CYP79D15* molecular variation. In Southern hybridizations using a worldwide sample of white clover plants (26 *Ac* plants, 16 *ac* plants), we find a perfect correlation between the presence of *CYP79D15* in the white clover genome and the synthesis of cyanogenic glucosides (Figure 2; Table 1). Consistent with this finding, we also find a perfect correlation between the presence of cyanogenic glucosides and *CYP79D15* gene expression in leaf tissue (Figure 3). Given that *CYP79D15* is the obvious candidate to catalyze the first step in cyanogenic glucoside biosynthesis, these findings strongly suggest that the *Ac/ac* polymorphism arises through the presence/absence of *CYP79D15* in the white clover genome.

As a further examination of *CYP79D15* and the *Ac/ac* polymorphism, we assessed phenotypic and genetic variation in an independent sample of 48 white clover plants collected in the vicinity of Washington University (St. Louis). Plants were scored for the presence/absence of cyanogenic glucosides, the presence/absence of *CYP79D15* (as assayed by PCR), and SNP variation at two other loci (*ACO1*, *ZIP*; see Olsen *et al.* 2007). Whereas the *CYP79D15* presence/absence polymorphism again showed a perfect correlation with the *Ac/ac* phenotype, these polymorphisms were uncorrelated with genetic variation at either *ACO1* ($\chi^2 = 0.77$, $P > 0.1$) or *ZIP* ($\chi^2 = 0.16$, $P > 0.1$), suggesting that the *CYP79D15* correlation with *Ac/ac* is not an artifact of genomewide linkage disequilibrium. Definitive proof that *CYP79D15* accounts for the *Ac/ac* polymorphism would be provided by genetic transformations of *ac* plants with *CYP79D15* to generate the *Ac* phenotype.

White clover cyanogenesis and gene presence/absence polymorphisms: The apparent molecular genetic basis of the *Ac/ac* polymorphism mirrors our findings for the *Li/li* polymorphism, which also arises through the presence/absence of the underlying gene (Olsen *et al.* 2007). Adaptive presence/absence polymorphisms have been documented for several other genes that function in plant defense (Stahl *et al.* 1999; Tian *et al.* 2002; Shen *et al.* 2006; reviewed by Tiffin and Moeller 2006). To our knowledge, all such previously documented cases have occurred in the model plant *A. thaliana*, and all have involved R genes, which function in pathogen recognition and initiation of the defense response.

For white clover, the finding that both the *Ac/ac* and the *Li/li* polymorphisms apparently arise through gene presence/absence polymorphisms suggests that the cyanogenesis polymorphism may derive from the allotetraploid origin of the species. White clover is thought to have originated through the hybridization of a cyanogenic species (contributing *Ac* and *Li*) and an acyanogenic species (possessing neither gene) (Williams and Williamson 2001; Badr *et al.* 2002). One of the hallmarks of allopolyploidization is the rapid and widespread elimination of genomic sequences in the immediate aftermath of the hybridization event (*e.g.*, Feldman *et al.* 1997; Ozkan *et al.* 2001; reviewed by Ma and Gustafson 2005). Thus, if the cyanogenesis loci were deleted from the genomes of some plants and maintained in others following the hybridization event, and if selection acted to maintain these gene presence/absence polymorphisms as the genome became diploidized, then the *Ac/ac* and *Li/li* polymorphisms could date to the origin of the species.

An alternative hypothesis is that the *Ac/ac* and *Li/li* polymorphisms were already present in the cyanogenic diploid progenitor of white clover and that allotetraploidization occurred multiple times such that these polymorphisms were carried into white clover. The diploid species *T. nigrescens* has been reported to be polymorphic for the presence/absence of both cyanogenic glucosides and linamarase (Williams and Williamson 2001). *T. nigrescens* has also traditionally been considered a strong candidate to be a diploid progenitor of white clover, as it is one of few closely related cyanogenic species (Williams and Williamson 2001). Thus, the white clover cyanogenesis polymorphisms might represent ancient polymorphisms that predate the origin of the species.

Under both of these scenarios, the gene presence/absence polymorphisms are expected to represent longstanding balanced polymorphisms in white clover. This leads to the prediction that the genomic regions immediately flanking the cyanogenesis loci should bear molecular signatures of balancing selection. Such signatures have been detected for genomic regions flanking R gene presence/absence polymorphisms in *A. thaliana* (*e.g.*, Stahl *et al.* 1999; Tian *et al.* 2002). Testing this hypothesis in white clover must await the identification of the flanking genomic sequences, as the genomic locations of *Ac* and *Li* are not yet known.

Molecular evolution of *CYP79D15* sequences: Our analyses of nucleotide variation at *CYP79D15* do not reveal deviations from neutral equilibrium expectations. Because *CYP79D15* sequences correspond to only one-half of the *Ac/ac* balanced polymorphism, analysis of these sequences would not necessarily be expected to reveal a signature of balancing selection. Analyses of *Li* sequences have similarly revealed no signature of balancing selection on the functional allele class (Olsen *et al.* 2007). *Li* sequences differ from *CYP79D15*, however, in that they show evidence of a selective sweep following an episode of positive directional selection (Olsen *et al.* 2007).

While *CYP79D15* sequences do not show statistically significant deviations from neutrality, they do show a high level of protein-coding variation, with nearly double the number of nonsynonymous substitutions compared to synonymous substitutions, and a four- to five-codon indel polymorphism (Table 2). Some of this variation may have functional significance. Quantitative variation in cyanogenic glucosides among *Ac* plants has been commonly reported in white clover (*e.g.*, Hughes 1991; see also materials and methods), and segregation analyses have indicated that a portion of this variation is attributable to differences among functional *Ac* alleles (Corkill 1942; Hughes *et al.* 1984). This suggests that molecular variation at *CYP79D15* (or closely linked to it) may affect quantitative variation in cyanogenic glucosides. In this study we did not observe any obvious correlation between *CYP79D15* haplotypes and cyanogenic glucoside levels (K. Olsen, unpublished observation). However, association studies following the identification of the *CYP79D15* promoter may shed further light on the role of this gene in quantitative variation in cyanogenic glucosides.

Conclusions: A major goal of modern evolutionary biology is to understand the molecular underpinnings of adaptation. While genetic model organisms such as *A. thaliana* and crop species are well suited to molecular evolutionary analysis, they are not necessarily ideal for understanding the ecological relevance of natural genetic variation. As an alternative approach, in this study we have examined an ecological model system: the ecology of white clover cyanogenesis has been studied for nearly a century (*e.g.*, Armstrong *et al.* 1913). Here we have identified the gene that is likely to underlie the *Ac/ac* cyanogenic glucoside polymorphism; we have found that, like *Li/li*, this component of the cyanogenesis polymorphism apparently arises through a gene presence/absence polymorphism. With recent advances in the development of genetic and genomic tools for white clover (*e.g.*, Barrett *et al.* 2004; Cogan *et al.* 2007), this species may prove a useful resource for further studies on the molecular genetics of adaptation.

The authors thank Pierre-François Perroud for technical advice on Southern hybridizations, Faith Steffen for assistance in SNP genotyping, other members of the Olsen lab group for helpful discussions and comments on the manuscript, and Washington University, which provided funds for this research to K. Olsen.

LITERATURE CITED

Andersen, M. D., P. K. Busk, I. Svendsen and B. L. Møller, 2000 Cytochromes P-450 from cassava (*Manihot esculenta* Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic glucosides linamarin and lotaustralin. J. Biol. Chem. **275:** 1966–1975.

Armstrong, H. E., E. F. Armstrong and E. Horton, 1913 Herbage studies. II.Variation in *Lotus corniculatus* and *Trifolium repens*: (cyanophoric plants). Proc. R. Soc. Lond. Ser. B **86:** 262–269.

Badr, A., H. Sayed-Ahmed, A. El-Shanshouri and L. E. Watson, 2002 Ancestors of white clover (*Trifolium repens* L.), as revealed by isozyme polymorphisms. Theor. Appl. Genet. **106:** 143–148.

Bak, S., C. E. Olsen, B. L. Petersen, B. L. Møller and B. A. Halkier, 1999 Metabolic engineering of *p*-hydroxybenzylglucosinolate in *Arabidopsis* by expression of the cyanogenic CYP79A1 from *Sorghum bicolor*. Plant J. **20:** 663–671.

Bak, S., C. E. Olsen, B. A. Halkier and B. L. Møller, 2000 Transgenic tobacco and *Arabidopsis* plants expressing the two multifunctional *Sorghum* cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. Plant Physiol. **123:** 1437–1448.

Bak, S., S. M. Paquette, M. Morant, A. V. Morant, S. Saito *et al.*, 2006 Cyanogenic glycosides: a case study for evolution and application of cytochromes P450. Phytochem. Rev. **5:** 309–329.

Barrett, B., A. Griffiths, M. Schreiber, N. Ellison, C. Mercer *et al.*, 2004 A microsatellite map of white clover. Theor. Appl. Genet. **109:** 596–608.

Barrett, T., C. G. Suresh, S. P. Tolley, E. J. Dodson and M. A. Hughes, 1995 The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. Structure **3:** 951–960.

Butler, G. W., and B. G. Butler, 1960 Biosynthesis of linamarin and lotaustralin in white clover. Nature **187:** 780–781.

Cogan, N. O. I., M. C. Drayton, R. C. Ponting, A. C. Vecchies, N. R. Bannan *et al.*, 2007 Validation of in silico-predicted genic SNPs in white clover (*Trifolium repens* L.), an outbreeding allopolyploid species. Mol. Genet. Genomics **277:** 413–425.

Collinge, D. B., and M. A. Hughes, 1982a *In vitro* characterisation of the *Ac* locus in white clover (*Trifolium repens* L.). Arch. Biochem. Biophys. **218:** 38–45.

Collinge, D. B., and M. A. Hughes, 1982b Developmental and physiological studies on the cyanogenic glucosides of white clover, *Trifolium repens* L. J. Exp. Bot. **33:** 154–161.

Collinge, D. B., and M. A. Hughes, 1984 Evidence that linamarin and lotaustralin, the two cyanogenic glucosides of *Trifolium repens* L., are synthesised by a single set of microsomal enzymes controlled by the Ac/ac locus. Plant Sci. Lett. **34:** 119–125.

Coop, I. E., 1940 Cyanogenesis in white clover III. Study of linamarase. NZ J. Sci. Technol. B **22–23:** 71–83.

Corkill, L., 1942 Cyanogenesis in white clover (*Trifolium repens*). V. The inheritance of cyanogenesis. NZ J. Sci. Technol. B **23:** 178–193.

Daday, H., 1954a Gene frequencies in wild populations of *Trifolium repens*. I. Distribution by latitude. Heredity **8:** 61–78.

Daday, H., 1954b Gene frequencies in wild populations of *Trifolium repens*. II. Distribution by altitude. Heredity **8:** 377–384.

Daday, H., 1958 Gene frequencies in wild populations of *Trifolium repens*. III. World distribution. Heredity **12:** 169–184.

Daday, H., 1965 Gene frequencies in wild populations of *Trifolium repens*. IV. Mechanisms of natural selection. Heredity **20:** 355–365.

De Araujo, A. M., 1976 The relationship between altitude and cyanogenesis in white clover (*Trifolium repens* L.). Heredity **37:** 291–293.

Dirzo, R., and J. L. Harper, 1982a Experimental studies on slug–plant interactions. III. Differences in the acceptability of individual plants of *Trifolium repens* to slugs and snails. J. Ecol. **70:** 101–117.

Dirzo, R., and J. L. Harper, 1982b Experimental studies on slug–plant interactions. IV. The performance of cyanogenic and acyanogenic morphs of *Trifolium repens* in the field. J. Ecol. **70:** 119–138.

Dirzo, R., and J. Sarukhan, 1984 *Perspectives on Plant Population Ecology*. Sinauer Associates, Sunderland, MA.

Ennos, R. A., 1981 Detection of selection in populations of white clover (*Trifolium repens* L.). Biol. J. Linn. Soc. **15:** 75–82.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Feigl, F., and V. Anger, 1966 Replacement of benzidine by copper ethylacetoacetate and tetra base as spot-test reagent for hydrogen cyanide and cyanogen. Analyst **91:** 282–284.

Feldman, M., B. Liu, G. Segal, S. Abbo, A. A. Levy *et al.*, 1997 Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics **147:** 1381–1387.

Forslund, K., M. Morant, B. Jorgensen, C. E. Olsen, E. Asamizu *et al.*, 2004 Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus*. Plant Physiol. **135:** 71–84.

Foulds, W., and J. P. Grime, 1972 The response of cyanogenic and acyanogenic phenotypes of *Trifolium repens* to soil moisture supply. Heredity **28:** 181–187.

Ganders, F. R., 1990 Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. Heredity **64:** 387–390.

Hall, T., 2001 *Biolign* alignment and multiple contig editor. http://en.bio-soft.net/dna/BioLign.html.

Hayden, K. J., and I. M. Parker, 2002 Plasticity in cyanogenesis of *Trifolium repens* L.: inducibility, fitness costs and variable expression. Evol. Ecol. Res. **4:** 155–168.

Hughes, M. A., 1991 The cyanogenic polymorphism in *Trifolium repens* L. (white clover). Heredity **66:** 105–115.

Hughes, M. A., and E. E. Conn, 1976 Cyanoglucoside biosynthesis in white clover (*Trifolium repens* L.). Phytochemistry **15:** 697–701.

Hughes, M. A., J. D. Stirling and D. B. Collinge, 1984 The inheritance of cyanoglucoside content in *Trifolium repens* L. Biochem. Genet. **22:** 139–151.

Kakes, P., 1989 An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. Theor. Appl. Genet. **77:** 111–118.

Kristensen, C., M. Morant, C. E. Olsen, C. T. Ekstrom, D. W. Galbraith *et al.*, 2005 Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. Proc. Natl. Acad. Sci. USA **102:** 1779–1784.

Ma, X.-F., and J. P. Gustafson, 2005 Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet. Genome Res. **109:** 236–249.

Majumdar, S., K. K. De and S. Banerjee, 2004 Influence of two selective forces on cyanogenesis polymorphism of *Trifolium repens* L. in Darjeeling Himalaya. J. Plant Biol. **47:** 124–128.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. Nature **351:** 652–654.

Melville, J., and B. W. Doak, 1940 Cyanogenesis in white clover II. Isolation of glucoside constituents. NZ J. Sci. Technol. B **22:** 67–70.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nelson, D. R., 2006 Cytochrome P450 nomenclature, 2004, pp. 1–10 in *Cytochrome P450 Protocols*, Ed. 2 (Methods in Molecular Biology, Vol. 320), edited by I. R. Phillips and E. A. Shephard. Humana Press, Totowa, NJ.

Olsen, K. M., B. L. Sutherland and L. L. Small, 2007 Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.). Mol. Ecol. **16:** 4180–4193.

Oxtoby, E., M. A. Dunn, A. Pancoro and M. A. Hughes, 1991 Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (*Trifolium repens* L.). Plant Mol. Biol. **17:** 209–219.

Ozkan, H., A. A. Levy and M. Feldman, 2001 Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell **13:** 1735–1747.

Pederson, G. A., and G. E. Brink, 1998 Cyanogenesis effect on insect damage to seedling white clover in a bermudagrass sod. Agron. J. **90:** 208–210.

Porebski, S. L., G. Bailey and B. R. Baum, 1997 Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. **15:** 8–15.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DNAsp, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Saucy, F., J. Studer, V. Aerni and B. Schneiter, 1999 Preference for acyanogenic white clover (*Trifolium repens*) in the vole *Arvicola terrestris*: I. Experiments with two varieties. J. Chem. Ecol. **25:** 1441–1454.

Seigler, D. S., and A. M. Brinker, 1993 Characterisation of cyanogenic glucosides, cyanolipids, nitroglycosides, organic nitro compounds and nitrile glucosides from plants. Methods Plant Biochem. **8:** 51–131.

Shen, J. D., H. Araki, L. L. Chen, J. Q. Chen and D. C. Tian, 2006 Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. Genetics **172:** 1243–1250.

Silvertown, J., and D. Charlesworth, 2001 *Introduction to Plant Population Biology*, Ed. 4. Blackwell Science, Oxford.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson, 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. Nature **400:** 667–671.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tian, D. C., H. Araki, E. Stahl, J. Bergelson and M. Kreitman, 2002 Signature of balancing selection in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **99:** 11525–11530.

Tiffin, P., and D. A. Moeller, 2006 Molecular evolution of plant immune system genes. Trends Genet. **22:** 662–670.

Till-Bottraud, I., P. Kakes and B. Dommee, 1988 Variable phenotypes and stable distribution of the cyanotypes of *Trifolium repens* L. in southern France. Acta Oecol. **9:** 393–404.

Viette, M., C. Tettamanti and F. Saucy, 2000 Preference for acyanogenic white clover (*Trifolium repens*) in the vole *Arvicola terrestris*. II. Generalization and further investigations. J. Chem. Ecol. **26:** 101–122.

Ware, W. M., 1925 Experiments and observations on forms and strains of *Trifolium repens*. J. Agric. Sci. **15:** 47–67.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Williams, W. M., and M. L. Williamson, 2001 Genetic polymorphism for cyanogenesis and linkage at the linamarase locus in *Trifolium nigrescens* Viv. subsp. *nigrescens*. Theor. Appl. Genet. **103:** 1211–1215.

Communicating editor: E. S. Buckler