Molecular Ecology (2007) **16**, 4180–4193 doi: 10.1111/j.1365-294X.2007.03506.x

Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.)

K. M. OLSEN, B. L. SUTHERLAND and L. L. SMALL
Department of Biology, Campus Box 1137, Washington University, St. Louis, MO 63130-4899, USA

Abstract

White clover (*Trifolium repens*) is naturally polymorphic for cyanogenesis (hydrogen cyanide release following tissue damage). The ecological factors favouring cyanogenic and acyanogenic plants have been examined in numerous studies over the last half century, making this one of the best-documented examples of an adaptive polymorphism in plants. White clover cyanogenesis is controlled by two, independently segregating Mendelian genes: *Ac/ac* controls the presence/absence of cyanogenic glucosides; and *Li/li* controls the presence/absence of their hydrolysing enzyme, linamarase. In this study, we examine the molecular evolution and population genetics of *Li* as it relates to the cyanogenesis polymorphism. We report here that *Li* exists as a single-copy gene in plants possessing linamarase activity, and that the absence of enzyme activity in *li/li* plants is correlated with the absence of much or all of the gene from the white clover genome. Consistent with this finding, we confirm by reverse transcription–polymerase chain reaction that *Li* gene expression is absent in plants lacking enzyme activity. In a molecular population genetic analysis of *Li* and three unlinked genes using a worldwide sample of clover plants, we find an absence of nucleotide variation and statistically significant deviations from neutrality at *Li*; these findings are consistent with recent positive directional selection at this cyanogenesis locus.

Keywords: adaptive polymorphism, β-glucosidase, cline, cyanogenesis, linamarase, positive selection

Received 1 April 2007; revision received 26 June 2007; accepted 18 July 2007

Introduction

Understanding the genetic basis of adaptation is a major goal of evolutionary biology. In the last decade, molecular population genetic studies have begun to reveal how selection and other microevolutionary processes shape the molecular evolution of genes underlying adaptive traits. The nature of such studies requires well-resolved genetic systems with clearly defined candidate genes. As a consequence, the choice of study systems thus far has been dictated more by genetic tractability than by the availability of systems where the ecological significance and fitness consequences of trait variation are well documented. In plants, the vast majority of molecular population genetic studies have focused on *Arabidopsis thaliana*, crop species or their relatives (reviewed by Wright & Gaut 2005). Notable exceptions include studies of floral trait variation and associated reproductive ecology (e.g. Bradshaw & Schemske 2003; Clegg & Durbin 2003; Zufall & Rausher 2004).

White clover (*Trifolium repens* L., Fabaceae) features one of the longest-studied and best-documented examples of an adaptive polymorphism in plants. This species is polymorphic for cyanogenesis (cyanide release following tissue damage), with both cyanogenic and acyanogenic plants occurring in natural populations. For more than 50 years, this polymorphism has been the subject of extensive ecological genetic research, involving both natural and experimental populations (reviews in Hughes 1991; Hayden & Parker 2002); it is a textbook example of a chemical defence polymorphism maintained by opposing selective pressures (e.g. Dirzo & Sarukhan 1984; Silvertown & Charlesworth 2001). Here, we present findings on the molecular genetic basis and molecular evolution of this adaptive polymorphism, focusing specifically on one of the two genes responsible for presence/absence of cyanide production.

White clover was first identified as polymorphic for cyanogenesis in the early 20th century (Armstrong *et al*.

Correspondence: Kenneth M. Olsen, Fax: (314) 935-4432; E-mail: kolsen@wustl.edu

1913; Ware 1925). Inheritance of the polymorphism is controlled by two independently segregating Mendelian genes (Coop 1940; Melville & Doak 1940; Corkill 1942): *Ac/ac* controls the presence/absence of cyanogenic glucosides (linamarin and lotaustralin) in the vacuoles of leaf and stem tissue; *Li/li* controls the presence/absence of their hydrolysing enzyme, linamarase, in the cell wall (reviewed by Hughes 1991). In cyanogenic plants, cell rupture from tissue damage brings the enzyme into contact with its substrate and generates the cyanogenic response. Both *Ac* and *Li* show incomplete dominance at the biochemical level, so that at least one functional allele must be present at both loci for a plant to be cyanogenic; two nonfunctional (null) alleles at either gene (*ac/ac* or *li/li*) confers the acyanogenic phenotype. The two genes can be scored independently for the presence or absence of functional alleles using hydrogen cyanide (HCN)-detection assays with exogenous addition of either cyanogenic glucosides or linamarase (method described below).

Over the last half century, several dozen studies have investigated the distributions of cyanogenic and acyanogenic plants in wild populations and the ecological factors that maintain the polymorphism (see reviews by Hughes 1991; Hayden & Parker 2002; Richards & Fletcher 2002). It is now well established that cyanogenesis in white clover serves as a deterrent against herbivores, specifically against small, grazing generalists (e.g. mollusks, insects, voles), which are a major source of leaf damage (Dirzo & Harper 1982a; Pederson & Brink 1998; Saucy *et al*. 1999; Viette *et al*. 2000; see also references therein). The factors favouring acyanogenic clover morphs are less clearly understood but are closely related to success under cool growing conditions. Altitudinal and latitudinal clines in cyanogenesis, correlated with minimum winter temperatures, have been reported since the 1950s (e.g. Daday 1954a, b). Cyanogenesis has been proposed to be deleterious in colder climates, since freezing would lead to cell rupture and autotoxicity from HCN release (Daday 1954a, b, 1958); alternatively, if there are fewer herbivores in colder climates, plants that do not invest in the chemical defence may be at a competitive advantage (Kakes 1997).

The *Ac* gene has not been characterized at the molecular level. In contrast, *Li* has been cloned from a complementary DNA (cDNA) library and sequenced (Oxtoby *et al*. 1991); in addition, the crystal structure of white clover linamarase (a cyanogenic β-glucosidase) has been determined, and the protein's key catalytic sites have been identified (Barrett *et al*. 1995). In plants with functional *Li* alleles, linamarase constitutes up to 5% of the total soluble leaf protein, while the enzyme is entirely absent in *li/li* plants (Hughes & Dunn 1982). Northern blots suggest that in *li/li* plants, the gene is either not transcribed or there is rapid degradation of *li* pre-mRNA (Oxtoby *et al*. 1991; Hughes 1993).

Southern hybridizations, using *Li* cDNA as a probe, have indicated the presence/absence of linamarase cosegregates with restriction fragment variation (Oxtoby *et al*. 1991; Hughes 1993), which suggests that genetic variation linked to the *Li* locus is responsible for the *Li/li* enzyme polymorphism. However, southerns based on *Li* cDNA probes were difficult to interpret, as exon portions of the gene are conserved with noncyanogenic β-glucosidase genes, leading to complex restriction-fragment banding patterns (Hughes *et al*. 1990; Oxtoby *et al*. 1991; Hughes 1993). Southern hybridizations that incorporate intron portions of *Li* would be expected to reveal more gene-specific hybridization patterns.

In this study, we examine the genomic organization of *Li* and the molecular evolution of this gene as it relates to the linamarase enzyme polymorphism in white clover. Specifically, we examine: *Li* gene copy number in plants with and without linamarase activity, using Southern hybridizations that incorporate *Li* intron sequences; *Li* gene expression in plants with and without linamarase activity; and the molecular population genetics of *Li* in comparison to three unlinked loci in a worldwide sample of white clover accessions.

Materials and methods

Study system and samples

Trifolium repens is a common legume of fields and pastures of temperate Eurasia, where it is native, as well as temperate regions throughout the world, where it has been introduced as a forage crop and has spread as a weed. The species' current wide geographical range reflects a recent, rapid range expansion following the development of agriculture in Europe and subsequent migrations of European colonists around the world. It is a perennial and an insect-pollinated, obligate outcrosser; it also spreads vegetatively by stolons. Chromosome number and genetic map data indicate that the white clover is a diploidized allotetraploid (Barrett *et al*. 2004). While the genetic map locations of *Ac* and *Li* are currently unknown, both genes are believed to be present in only one of the two parental genomes (Williams & Williamson 2001).

Seeds from a geographically diverse sample of *T. repens* strains, representing the native and introduced species range, were obtained from the US Department of Agriculture (USDA) clover germplasm collection or from naturalized populations growing in Raleigh, North Carolina (Table 1). Twenty-two cyanogenic and 10 acyanogenic strains were examined; USDA accessions were selected from germplasm that had previously been characterized for cyanogenesis (http://www.ars-grin.gov/index.html). Seeds of USDA accessions were lightly scarified, germinated, and grown in standard greenhouse conditions at

Table 1 *Trifolium* accessions used in analyses. Accessions with prefix PI were obtained from the USDA clover germplasm collection. Accessions used in Southern hybridizations are indicated in bold; those used in RT–PCR are indicated by an asterisk. All *Li* accessions were sequenced at *Li*, *ACO1*, *ALDP*, and *ZIP* unless otherwise noted. *Li* haplotype labels correspond to Table 2

Trifolium species	Accession	Country of origin	*Li/li* pheno-type	*Ac/ac* pheno-type	*Li* haplotypes
T. repens	PI 200372	Israel	*Li*	*Ac*	1/1
T. repens	PI 204930	Turkey	*Li*	*Ac*	2/2
T. repens	**PI 214207**	**Israel**	*Li*	*Ac*	**1/3**
T. repens	PI 226996	Uruguay	*Li*	*Ac*	4/5
T. repens	PI 230183	Argentina	*Li*	*Ac*	6/6
T. repens	PI 234678	France	*Li*	*Ac*	3/3
T. repens	**PI 239977***	**Portugal**	*Li*	*Ac*	**7/7**
T. repens	PI 291828†	Chile	*Li*	*Ac*	7/7
T. repens	**PI 294546**	**France**	*Li*	*Ac*	**7/7**
T. repens	PI 298485	Israel	*Li*	*Ac*	5/5
T. repens	**PI 311494***	**Spain**	*Li*	*Ac*	**3/8**
T. repens	PI 350706	Australia	*Li*	*Ac*	5/9
T. repens	PI 384699	Morocco	*Li*	*Ac*	5/7
T. repens	PI 418905	Italy	*Li*	*Ac*	1/5
T. repens	PI 418911	Italy	*Li*	*Ac*	5/10
T. repens	PI 419316	Greece	*Li*	*Ac*	11/11
T. repens	PI 419401	Greece	*Li*	*Ac*	12/13
T. repens	PI 420001	Japan	*Li*	*Ac*	7/14
T. repens	PI 517126	Morocco	*Li*	*Ac*	3/5
T. repens	PI 517515†	Ethiopia	*Li*	*Ac*	5/15
T. repens	NC 47†‡§	USA (Raleigh, NC)	*Li*	*Ac*	7/7
T. repens	NC A28*‡§	USA (Raleigh, NC)	*Li*	*Ac*	7/16
T. repens	**PI 100247**	**New Zealand**	*li*	*Ac*	—
T. repens	**PI 205062***	**Turkey**	*li*	*Ac*	—
T. repens	**PI 251191**	**Macedonia**	*li*	*ac*	—
T. repens	**PI 440745***	**Russia**	*li*	*Ac*	—
T. repens	PI 516411	Romania	*li*	*ac*	—
T. repens	PI 542915	Bosnia-Herzegovina	*li*	*ac*	—
T. repens	NC F83	USA (Raleigh, NC)	*li*	*ac*	—
T. repens	NC A88	USA (Raleigh, NC)	*li*	*ac*	—
T. repens	NC I76	USA (Raleigh, NC)	*li*	*ac*	—
T. repens	NC I88	USA (Raleigh, NC)	*li*	*ac*	—
T. nigrescens ssp. *petrisavii*	PI 120103‡	Turkey	*Li*	*Ac*	—
T. nigrescens ssp. *petrisavii*	PI 298478	Israel	*Li*	*Ac*	—
T. isthmocarpum	PI 422595‡	Morocco	*Li*	*Ac*	—

†no sequence data for *ZIP*; ‡no sequence data for *ACO1*; §no sequence data for *ALDP*.

Washington University. A single plant was grown per accession. North Carolina accessions were grown from stem cuttings.

Cyanogenesis tests

A modified Feigl–Anger test (Feigl & Anger 1966) was used to confirm cyanotypes in USDA samples and to test for cyanogenesis in North Carolina samples. For each accession, leaf tissue was picked and placed in six wells of a 48-well polystyrene microtiter plate (with two to three leaves per well). Leaf samples were frozen at –80 °C for at least 30 min to rupture cells, then thawed for 10 min at 37 °C. For each set of six wells, two wells received 30 μL 10 mM linamarin solution (A.G. Scientific), two wells received 40 μL 0.2 EU/mL linamarase (Gallard-Schlesinger), and two samples were left unsupplemented. One well per plate was used as a positive control (linamarin and linamarase solutions with no leaf sample). Leaf samples were macerated using pipet tips, and a piece of Feigl–Anger test paper was then clamped between the microtiter plate and its cover. The plate was incubated for 2 h at 37 °C. HCN release in a well was detectable by the generation of a dark blue spot on the test paper over the well.

Cyanogenic plants (*Ac__, Li__*) produce a colour reaction in all six wells; acyanogenic plants lacking cyanogenic glucosides but possessing enzyme (*ac/ac, Li__*) produce a colour change only in wells with exogenous linamarin; acyanogenic plants possessing cyanogenic glucosides but lacking enzyme (*Ac__, li/li*) produce a colour reaction only in wells with exogenous linamarase; and acyanogenic plants lacking functional alleles at both loci do not produce a colour change in any well.

Polymerase chain reaction amplification and DNA sequencing

DNA was extracted from fresh leaf tissue using a modified cetyltrimethyl ammonium bromide (CTAB) extraction protocol (Porebski *et al*. 1997). Primers for *Li* polymerase chain reaction (PCR)-amplification were designed off the published cDNA sequence (GenBank Accession no. X56733), which corresponds to the complete mature peptide plus a portion of amino-end transit peptide. Most PCR amplifications were performed using forward primer 5′-TCCATCACTACTACTCATATCCATGC-3′ and reverse primer 5′-TGGGCTGGTCCATTTGATTTAAC-3′. A number of additional primers and primer combinations were used in preliminary PCR and as internal primers in DNA sequencing (see Table S1, Supplementary material). PCR was performed with 20-μL reactions and standard reaction conditions, using Ex *Taq* high fidelity DNA polymerase (TaKaRa). Annealing temperatures were adjusted for primer combinations. PCR products were cloned into plasmids using TOPO TA cloning (Invitrogen), purified and then sequenced; eight to nine clones were sequenced per accession to detect allelic variation and to allow detection of singletons arising through PCR error. The *Li* gene was sequenced in 22 cyanogenic white clover plants. For outgroup comparison in DNA sequence analyses, *Li* was also sequenced in two closely related cyanogenic species: *Trifolium nigrescens* ssp. *petrisavii* and *Trifolium isthmocarpum* (Table 1); eight to nine clones were sequenced per individual.

All DNA sequencing was performed using BigDye terminators (Applied Biosystems), with reactions run on an ABI 3130 sequencer in the Biology Department of Washington University. Initial PCR and sequencing revealed a second gene with close sequence similarity to *Li*, which subsequent sequencing and Southern hybridization revealed to be a paralogous gene copy (see below). Whereas we were unable to amplify any portion of *Li* in plants lacking enzyme activity (despite repeated attempts using multiple primer combinations in nine acyanogenic accessions), the putative *Li* paralogue was readily amplified in accessions both with and without enzyme activity.

For 20 cyanogenic accessions and the outgroups, we also sequenced ~0.5-kb portions of three other nuclear genes for comparison of nucleotide diversity levels with *Li*. These three genes are: *ACO1*, encoding 1-aminocyclopropane-1-carboxylate (ACC) oxidase; *ALDP*, encoding a plastidic aldolase; and *ZIP*, encoding a putative fatty acid desaturase. Primers were designed from published *Trifolium* gene sequences to amplify regions spanning one or more introns. PCR was performed using Ampli*Taq* DNA polymerase (Applied Biosystems) with standard reaction conditions. A minimum of eight clones per amplicon was sequenced to detect allelic variation and to control for PCR artefacts. No more than two sequence haplotypes (alleles) were detected per accession, an indication that paralogous or homeologous sequences were not included in sequence analyses (see below). DNA sequences are available on GenBank under accession nos EF990367–EF990516.

Determination of gene copy number

To test for *Li* gene copy number in plants with and without enzyme activity, we performed Southern hybridizations on genomic DNA digests from four linamarase-containing accessions and four accessions lacking the enzyme (Table 1). Genomic DNA was purified as described above, and ~10 μg DNA was digested with either *Afl*II or *Ase*I and run on a 0.6% Seakem (Fisher Scientific) agarose gel. The DNA was then transferred to a nitrocellulose membrane using standard methods.

Three probes were used in Southern hybridizations: probe L1, corresponding to a 775-bp region within a large intron near the 5′ end of the *Li* genomic sequence; probe P1, corresponding to the same intron region of the putative *Li* paralogue identified in initial sequencing (a sequence with 89% sequence similarity to probe L1); and probe L2, corresponding to a 827-bp portion of *Li* at the 3′ end of the gene (spanning 4 exons, 4 introns and 60 bp of the 3′UTR). Probe L2 shares approximately 93% sequence similarity with the corresponding portion of the putative *Li* paralogue. Probes were prepared using the DIG probe synthesis kit (Roche); hybridization washes were performed at high stringency following the manufacturer's protocol. Primers used in PCR amplification of probes are listed in the Supplementary material (Table S1). Based on the DNA sequences of *Li* and the *Li* paralogue, *Ase*I is predicted to cut once within the L1 region of *Li*, but not within the probe P1 region of the paralogue. For the probe L2 region, *Ase*I is predicted to have one detectable cut site in both *Li* and the *Li* paralogue (a second cut site occurs 6 bp from the 5′ end of the probe in both genes). *Afl*II is predicted to be a noncutter within the gene sequence of both *Li* and the *Li* paralogue.

Gene expression analysis

Total RNA was extracted from fresh young leaf tissue of two accessions with and without linamarase activity

(Table 1) using RNeasy kits (QIAGEN). Reverse transcription and reverse transcription (RT)–PCR was performed using a Thermoscript RT–PCR kit (Invitrogen). To test for *Li* gene expression, *Li*-specific PCR primers were designed to amplify a portion of the gene near the 3′ end corresponding to a 660-bp cDNA sequence and a 1048-bp genomic DNA sequence. To test for expression of the *Li* paralogue, primers specific to the paralogue were designed to amplify a region corresponding to 464 bp of cDNA and 946 bp of genomic DNA. As a positive control for RT–PCR, primers were designed to amplify a 151-bp region of 5.8S ribosomal RNA. RT–PCR primers are listed in 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.

Analysis of DNA sequence variation

DNA sequences from *Li*, *ACO1, ALDP*, and *ZIP* were edited and aligned visually using BIOLIGN software (http://www.maizegenetics.net/index.php?page=bioinformatics/index.html). Intron positions were inferred by comparison to published cDNAs. Most 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). Allele frequency-based tests of selection were performed using Tajima's (1989) *D* and Fay & Wu's (2000) *H* tests; statistical significance was assessed by coalescent simulations (1000 replicates), using θ_W and levels of recombination estimated empirically from the data. To test for significant differences in Tajima's *D* values between pairs of loci, the ΔD test of Hahn *et al*. (2002) was employed, as implemented for multilocus comparisons (see McDaniel & Shaw 2005); 1000 pairs of coalescent replicates were used per simulation. In addition, the McDonald & Kreitman (1991) test was performed to compare polymorphism and divergence at *Li* and the three other loci.

The composite likelihood ratio (CLR) test of Kim & Stephan (2002), as modified by Jensen *et al*. (2005), was employed as a further test of selection on *Li*. The CLR test calculates the likelihood of observing a given number of derived polymorphisms under a selective sweep model and compares this likelihood to the expectation under a neutral model of evolution. The maximum-likelihood ratio from the observed data was calculated using the program CLSW, with the identities of ancestral and derived polymorphisms determined by comparison of *Li* sequences to the outgroup species. The maximum-likelihood ratio (LR1) was then compared to the distribution of ratios obtained from 1000 neutral data sets simulated from the *Li* sequence data in the program SSW. As the CLR test rejected the neutral equilibrium model, the goodness-of-fit (GOF) test of Jensen *et al*. (2005) was then performed to assess whether the observed distribution was consistent with a null hypothesis of a selective sweep; the GOF test is designed to discriminate between deviations from neutral equilibrium caused by selection and deviations arising through historical demographic effects.

Assessments of population differentiation

Genome-wide genetic differentiation among individuals in a sample (population structure) can potentially confound inferences about selection at the molecular level. To test for evidence of population structure in the sampled clover plants, DNA sequences from *ACO1*, *ALDP*, and *ZIP* were analysed using three methods. First, since population structure would be expected to result in linkage disequilibrium (LD) between loci, pairwise tests of LD between the three genes were performed using GENEPOP (http://wbiomed.curtin.edu.au/genepop/index.html), with haplotypes in each accession coded as diploid genotypic data. Second, neighbour-joining distance trees were constructed from DNA sequences for each gene using PAUP* 4.0b (Swofford 2002). The three trees were compared to each other to test for evidence of congruent patterns of genetic differentiation among accessions; in addition, each tree was assessed for any patterns of geographical differentiation across the species range. Third, a Mantel test was used to compare matrices of geographical vs. genetic distances among sampled accessions to test for genetic isolation by distance across the species range. Genetic distances among pairs of individuals were calculated in GENEPOP with 10 000 permutations; geographical distances were calculated as great circle distances from geographical coordinates.

Results

Cyanogenesis assays

Biochemical assays were performed with a worldwide sample of 32 white clover plants to test for the presence/absence of cyanogenesis and to determine the genetic basis of acyanogenesis in those plants lacking cyanide production. Twenty-two plants were found to be cyanogenic; these plants, by definition, carry functional alleles at both *Ac* (accounting for the presence of cyanogenic glucosides) and *Li* (accounting for the presence of their hydrolysing enzyme linamarase) (Table 1). Out of the 10 acyanogenic accessions, seven were found to lack both components of the cyanogenic response (*ac li* phenotype); three acyanogenic accessions were found to contain cyanogenic glucosides, indicating the presence of a functional *Ac* allele (Table 1). The occurrence of nonfunctional alleles is

Fig. 1 Structure of the *Li* gene. Black rectangles represent exons and correspond to the previously published cDNA sequence; intervening lines represent introns. Lengths of exons and introns are approximately to scale. Grey bars indicate portions of the gene used to design probes for Southern hybridizations. The cross-hatched bar indicates the region targeted in gene expression analyses. Black dots indicate positions of *Ase*I restriction sites, and asterisks indicate approximate locations of amino acid replacement polymorphisms.

often correlated between *Ac* and *Li* in wild populations (Daday 1954a; Ennos 1982; Till-Bottraud *et al*. 1988; discussed below). Thus, the observation that most of the acyanogenic accessions lack functional alleles at both genes is not unexpected.

Structure of the Li *gene*

Li was sequenced in the cyanogenic accessions of our sample. The structure of the sequenced region is shown in Fig. 1. The region is approximately 3.9 kb and includes 12 complete exons, 12 introns, a portion of the exon at the 5′ end of the gene sequence, and portion of the 3′-UTR. The sequenced region includes all codons corresponding to the 490-amino acid mature linamarase peptide (Barrett *et al*. 1995), plus 3 codons corresponding to the amino-end signal peptide (Oxtoby *et al*. 1991).

Repeated attempts to PCR-amplify *Li* in samples lacking enzyme activity (*li/li* accessions) failed to amplify any portion of the *Li* gene. Instead, a closely related sequence was amplified in these accessions. We were able to consistently amplify this same sequence in cyanogenic accessions as well, leading us to suspect that it was a paralogue of *Li*, rather than allelic variation. The putative *Li* paralogue shows conservation of intron positions and splice sequences with *Li*. In addition, there are no frame-shift mutations or premature stop codons in sequences from nine acyanogenic and five cyanogenic accessions, suggesting that the gene encodes a functional protein. The inferred amino acid sequence is approximately 92% identical to linamarase, and amino acids are conserved at eight key sites that characterize the particular β-glucosidase family to which linamarase belongs (family 1 glycosyl hydrolases) (Barrett *et al*. 1995). This conservation at both the nucleotide and amino acid levels suggests that the putative *Li* paralogue encodes a functional protein, which is present in both cyanogenic and acyanogenic plants.

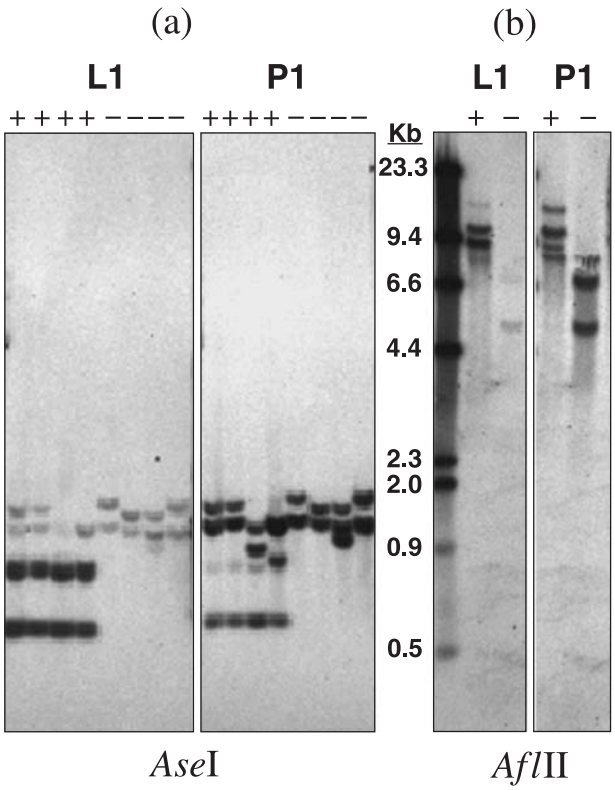

Fig. 2 Southern hybridizations of accessions with and without linamarase activity (+ and –, respectively). (a) *Ase*I digests probed with *Li* probe 1 (L1) and the *Li* paralogue probe 1 (P1). *Ase*I is predicted to have a single cut site within the L1 probe region of *Li*, and no cut sites within the P1 probe region of the *Li* paralogue. Accessions are identified in Table 1. (b) *Afl*II digests of a linamarase-containing accession (PI 311494) and a linamarase-lacking accession (PI 440745) probed with *Li* probe 1 (L1) and the *Li* paralogue probe 1 (P1). *Afl*II is predicted to have no cut sites within the probed regions of either *Li* or the *Li* paralogue.

Li *copy number*

To assess gene copy number for *Li* and related genes in cyanogenic plants and plants lacking linamarase, we performed Southern hybridizations using (i) a probe specific to a large intron near the 5′ end of *Li* (probe L1); (ii) a probe specific to the corresponding intron in the putative *Li* paralogue (probe P1); and (iii) a probe spanning the 3′ end of *Li* (probe L2). The locations of the probes within *Li* are indicated in Fig. 1.

Figure 2(a) shows the results of probe L1 and probe P1 hybridizations for cyanogenic and acyanogenic accessions digested with *Ase*I. This enzyme is predicted to cut within probe L1 once, and not within the paralogue probe P1. Two strong bands are observed in each of the four cyanogenic accessions (~0.6 kb and ~0.8 kb; Fig. 2a), a pattern consistent with *Li* occurring as a single-copy

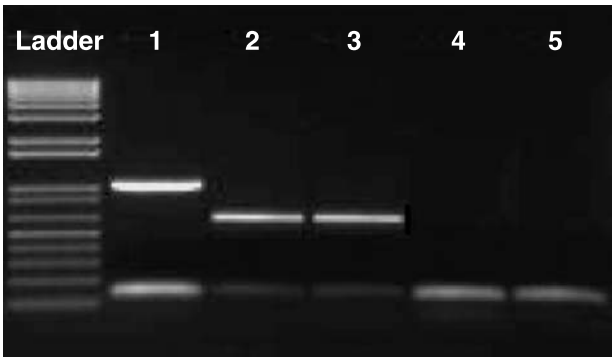

Fig. 3 Expression of *Li* in white clover accessions with and without linamarase activity, as assessed by RT–PCR. Lane 1 (genomic DNA control): PCR from genomic DNA of linamarase-containing accession PI 311494 using *Li* RT–PCR primers (top band) and 5.8S rDNA primers (bottom band). Lanes 2 and 3: PCR from cDNA of linamarase-containing accessions PI 311494 and PI 239977, respectively, using *Li* RT–PCR primers (top band) and 5.8S rDNA primers (bottom band). Lanes 4 and 5: PCR from cDNA of *li/li* accessions PI 205062 and PI 440745, respectively, using *Li* RT–PCR primers (no product) and 5.8S rDNA primers (bottom band). *Li* RT–PCR primers are expected to amplify a 660-bp cDNA region corresponding to a 1048-bp genomic DNA region; the 5.8S rDNA primers are expected to amplify a 155-bp product in both genomic cDNA and genomic DNA. Ladder (first 10 bands from bottom): 100, 200, 300, 400, 500, 650, 850, 1000, 1650, and 2000 bp.

gene in cyanogenic plants. Additional weak bands occur in these cyanogenic accessions; hybridization with probe P1 confirms that these weak bands are more closely related to the *Li* paralogue than to *Li* (Fig. 2a). In the four accessions lacking linamarase activity, probe L1 reveals weak bands only, all of which hybridize strongly to the paralogue probe P1 (Fig. 2a). These patterns suggest that in these acyanogenic accessions, there is no genomic sequence with close homology to *Li* probe L1. In addition, both the cyanogenic and acyanogenic accessions appear to carry at least two gene copies with greater similarity to the sequenced *Li* paralogue than to *Li*.

*Afl*II digests corroborate inferences based on *Ase*I. Figure 2(b) shows a representative cyanogenic accession and acyanogenic accession digested with this enzyme, which is predicted not to cut within either *Li* or the putative *Li* paralogue. When probed with L1, the linamarase-containing accession shows a strongest band slightly below 9.4 kb, a second strong band slightly above 9.4 kb, and two faint bands (~8 kb and ~12 kb). When this accession is probed with the paralogue-specific probe P1, the band that hybridized most strongly to probe L1 is weaker, while the other three bands are all stronger (Fig. 2b). This pattern suggests that the *Li* gene corresponds to the band below 9.4 kb, while the other bands are more closely related to the *Li* paralogue than to *Li*. For the accession-lacking linamarase, the L1 probe shows faint bands only (~6.6 kb and ~5 kb); these same bands hybridize strongly to the paralogue-specific probe P1. These patterns are again consistent with the occurrence of *Li* in linamarase-containing accessions only, and the occurrence of other genes (which are more closely related to the *Li* paralogue than to *Li*) in both the cyanogenic and acyanogenic clover genomes.

As with probe L1, probe L2 hybridizes strongly to bands only in cyanogenic accessions, both for *Afl*II and *Ase*I digests (Fig. S1, Supplementary material). Taken together, results of the Southern hybridizations suggest the following: *Li* occurs as a single-copy gene in cyanogenic accessions; much of the *Li* gene sequence, including both 5′ and 3′ portions of the gene, is apparently absent from the genome of plants lacking linamarase activity; and the *Li* paralogue and related genes are present in the genomes of both cyanogenic and acyanogenic plants.

Li expression

RT–PCR primers for *Li* were designed to amplify a 660-bp portion of cDNA, corresponding to a 1048-bp genomic sequence at the 3′ end of the gene (see Fig. 1). Gene expression is detected only in plants that contain linamarase activity (Fig. 3). This finding is consistent with the apparent absence of much of the *Li* gene from the genomes of acyanogenic plants (Fig. 2; see also Fig. S1), as well as with a previous study using Northern blots, which found high *Li* expression in plants with linamarase and little or none in *li/li* plants (Oxtoby *et al*. 1991). RT–PCR using primers specific to the *Li* paralogue failed to generate PCR products of any size in either cyanogenic or acyanogenic plants (data not shown). This result suggests that expression of the *Li* paralogue is low or lacking in leaf tissue, or that the transcript is unstable.

Molecular population genetics

We sequenced the *Li* gene in a sample of 22 cyanogenic accessions representing the species range of *Trifolium repens*. Nucleotide variation is extremely low across both coding and noncoding portions of the gene. A total of 18 substitution polymorphisms was observed across the 3.9-kb region (Table 2); 13 of these occur in introns, and 5 encode amino acid replacements (Leu-87 > Ser; Ile-179 > Thr; Lys-288 > Asn; Pro-418 > Thr; Arg-435 > Gly); no synonymous substitutions were observed. Nucleotide diversity at *Li* is dramatically lower than that of the three other loci examined (Table 3). Whereas the average pairwise nucleotide diversity at *Li* is less than 0.1% ($\pi = 0.00067$), values for the other three genes are approximately 10–25 times greater (Table 3).

Table 2 Nucleotide polymorphisms at *Li*. Numbers in parentheses indicate the number of sequences observed per haplotype. Asterisks indicate nonsynonymous polymorphisms and correspond to positions indicated in Fig. 1. Dots indicate that a nucleotide is identical to that of haplotype 1

Haplotype	No.	348 (Intron 2)	505 (Intron 2)	583 (Intron 2)	859 (Intron 2)	1011 (Intron 2)	1071 (Intron 2)	1184 (Intron 2)	1188 (Intron 2)	1841 (Intron 2)	2673 (Intron 8)	2793* (Exon 9)	2816* (Exon 9)	3098 (Intron 9)	3272* (Exon 11)	3490* (Exon 12)	3541* (Exon 12)	3578 (Intron 12)	3604 (Intron 12)
1	(4)	T	T	A	T	C	A	C	G	T	C	A	G	A	A	C	A	C	G
2	(2)	A	C	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
3	(5)	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
8	(1)	A	.	.	.	.	.	.	.	C	.	.	.	.	.	.	G	.	.
4	(1)	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	G	.	.
7	(11)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
14	(1)	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.
9	(1)	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
5	(9)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
10	(1)	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.
11	(2)	.	.	T	.	.	.	A	A	.	T	.	.	.	T	.	.	T	.
12	(1)	.	.	T	.	.	.	.	.	.	T	.	A	.	T	.	.	T	.
13	(1)	.	.	T	.	.	.	.	.	.	T	.	.	.	T	.	.	T	.
15	(1)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
16	(1)	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
6	(2)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A

Tests of selection based on allele frequencies reveal deviations from neutral equilibrium specifically at *Li*. Values for both Tajima's (1989) test ($D = -1.2440$) and Fay & Wu's (2000) test ($H = -2.926$) are statistically significant at $P = 0.05$ (Table 3). In contrast, none of the other three genes show any deviations from neutral expectations in these tests ($P > 0.1$ in all cases; Table 3); in addition, the ΔD test (Hahn *et al*. 2002) indicates that the value of Tajima's *D* at *Li* is significantly different from that of *ACO1*, the locus with the most similar *D* value ($P = 0.05$). These patterns suggest that the observed deviations from neutrality are unique to *Li*, and that they are not due to demographic factors (e.g. population expansion), which would be expected to affect all loci examined. Rather, the significantly negative values of *D* and *H* are consistent with the action of positive directional selection at *Li*.

Similarly, LD-based tests of selection reveal deviations from neutral equilibrium at *Li* that are consistent with positive selection at this cyanogenesis locus. The CLR test (Kim & Stephan 2002) rejects a neutral equilibrium model for the *Li* data set (LR1 = 4.241, $P = 0.050$), and the goodness-of-fit test of Jensen *et al*. (2005) indicates that these data are consistent with the null hypothesis of a selective sweep model (GOF1 = 321.160, $P = 0.120$) rather than deviations arising through demographic effects.

Assessments of polymorphism and divergence by the McDonald–Kreitman (MK) test also provide evidence for positive selection acting on *Li* in *T. repens*, albeit with weaker statistical support than for frequency spectrum and LD-based tests. When the MK test is performed using *Trifolium isthmocarpum* as the outgroup species, there is a statistically significant excess of nonsynonymous substitutions fixed between species (*G* test, $P = 0.0328$). On the other hand, the test is not significant at $P = 0.05$ when *Trifolium nigrescens* is used as the outgroup, although it does approach statistical significance ($P = 0.0604$). By comparison, the MK test for *ALDP* does not approach statistical significance for either outgroup comparison ($P > 0.1$ for both tests); MK tests could not be performed for *ACO1* and *ZIP* because of a lack of nonsynonymous variation. Together these findings are generally consistent with an inference of positive directional selection acting to fix nonsynonymous substitutions in *Li* alleles of *T. repens*.

Population structure

If present, population structure in a sample set can potentially confound molecular population genetic inferences about selection. There is no evidence of population structure among the sampled clover plants. Neighbour-joining trees constructed from the three neutrally evolving genes (*ACO1*, *ALDP*, and *ZIP*) are incongruent with each other and with the geographical locations of sampled

Table 3 Nucleotide variation and tests of selection for *Li* and three other nuclear genes in a worldwide sample of cyanogenic *Trifolium repens* accessions. Statistically significant values ($P < 0.05$) are indicated in bold

Gene	Region sequenced	No. of sequences	Approx. length	No. of segregating sites	π*	θ_W*	Tajima's *D*	Fay and Wu's *H*†
Li	exons 1–12	44	3.9 kb	18	0.0007	0.0011	**–1.2440 (*P* = 0.0470)**	**–2.926 (*P* = 0.0420)**
ACO1	intron 3	40	0.56 kb	36	0.0190	0.0158	0.6964 ($P > 0.1$)	4.1180 ($P > 0.1$)
ALDP	exons 4–5‡	40	0.57 kb	17	0.0164	0.0130	0.8462 ($P > 0.1$)	0.9692 ($P > 0.1$)
ZIP	exons 1–2‡	36	0.53 kb	8	0.0073	0.0050	1.1148 ($P > 0.1$)	–0.4222 ($P > 0.1$)

*Calculated for silent sites; †values are shown for *Trifolium nigrescens* ssp. *petrisavii* as the outgroup; these values differ negligibly from those with *Trifolium isthmocarpum* as the outgroup; ‡exon numbers based on blasts of *Trifolium* cDNAs against genomic sequences of other legume species.

haplotypes (data not shown). Moreover, on each tree there are numerous instances where the two haplotypes within a single individual occur on distant branches, and where geographically distant accessions share identical haplotypes [e.g. a shared *ACO1* haplotype in plants originating from Turkey (PI 204930) and Portugal (PI 239977); Table 1]. Explicitly testing for genetic isolation by distance with a Mantel permutation test confirms that there is no correlation between geographical and genetic distances among the sampled accessions ($P = 0.6$). In addition, pairwise tests of linkage disequilibrium between loci are nonsignificant in all comparisons (Fisher's exact tests, $P > 0.2$); population structure would be expected to generate LD among the three neutrally evolving genes in the sequenced accessions.

The apparent lack of population structure in *T. repens* is consistent with its weedy life history and very recent human-associated range expansion in both its native and non-native range. Since *Li* is the only locus out of four examined to show any deviations from neutral equilibrium, the inferred selection at this gene seems unlikely to be an artefact of population structure, as deviations from neutral equilibrium caused by population structure would be expected to affect all four loci examined.

Discussion

Genetic basis of the Li/li *enzyme polymorphism in* Trifolium repens

While the ecology of the white clover cyanogenesis polymorphism has been a subject of intensive research over the last half-century, the molecular genetic basis of this polymorphism has remained unresolved. In this study, we have examined the genetic basis and molecular evolution of the *Li/li* polymorphism, one of two loci (along with *Ac/ac*) responsible for the presence/absence of cyanide production in *Trifolium repens*. Southern hybridizations and sequencing of PCR products indicate that *Li* occurs as a single-copy gene in plants possessing linamarase activity, and that much or all of the gene is absent from the genomes of *li/li* plants. This finding suggests that the *Li/li* component of the white clover cyanogenesis polymorphism arises by the presence/absence of the entire *Li* gene, rather than by variation in levels of gene expression or by variation in functionality of the *Li* gene product.

Previous attempts to identify the genetic basis of the white clover *Li/li* polymorphism relied on Southerns that employed the *Li* cDNA sequence as a probe (Hughes *et al*. 1990; Oxtoby *et al*. 1991; Hughes 1993). Since exon portions of *Li* are conserved with other β-glucosidase genes (Hughes 1993; see also Fig. 2 and Fig. S1), the resulting blots revealed complex banding patterns that were difficult to interpret. Nonetheless, an association between restriction fragment variation and the enzyme polymorphism was observed, suggesting that the enzyme polymorphism results from genetic variation linked to *Li*. That observation is consistent with our own findings of a gene deletion polymorphism at *Li*. The apparent absence of *Li* from *li/li* plants is also consistent with Northern blots (Oxtoby *et al*. 1991; Hughes 1993) and our own RT–PCR analyses (Fig. 3), both of which indicate that *Li* gene expression occurs only in plants with linamarase activity.

The evolutionary origin of the *Li* gene deletion polymorphism remains unclear. *Trifolium repens* is believed to be a diploidized allotetraploid derived from two diploid *Trifolium* species (Badr *et al*. 2002; Barrett *et al*. 2004). A recent molecular systematic study by Ellison *et al*. (2006), based on nuclear rDNA and cpDNA sequences, has identified two species, *Trifolium pallescens* and *Trifolium occidentale*, as the most likely parents of *T. repens*. Neither of these species has been reported to be cyanogenic, and cyanogenesis assays conducted on three *T. pallescens* accessions (PI 516376, PI 631776, PI 634109) and one *T. occidentale* accession (PI 369723) indicate that both species are acyanogenic (K. Olsen, personal observation). This observation suggests that neither of these proposed progenitor species

is a clear candidate to have contributed the *Li* gene to *T. repens*. Alternatively, the cyanogenic species *Trifolium nigrescens* has previously been proposed to be one of the diploid progenitors of white clover (Williams & Williamson 2001; Badr *et al*. 2002); *T. nigrescens* ssp. *nigrescens* has been found to be polymorphic at both *Ac* and *Li* (Williams & Williamson 2001). A clear answer on the origin of the *Li/li* polymorphism in white clover will require definitive information on the identities of its diploid progenitors.

Genes related to Li

Our results indicate the occurrence of at least two gene copies in the *T. repens* genome that share close sequence similarity with *Li* (Fig. 2; see also Fig. S1). Two lines of evidence suggest that these genes do not function in cyanogenesis. First, unlike *Li*, they are present in the genomes of both linamarase-containing plants and *li/li* plants (Fig. 2). Second, our RT–PCR results suggest that the *Li*-paralogue sequenced in our study is apparently not expressed in leaf tissue, the primary cyanogenic tissue of *T. repens*. An additional noncyanogenic β-glucosidase has previously been cloned from a *T. repens* cDNA library and sequenced (Oxtoby *et al*. 1991). Whereas the *Li* paralogue sequenced in the present study shares 94% DNA sequence identity and 92% amino acid sequence similarity with *Li*, the previously identified gene is more distantly related (76% cDNA sequence identity, 60% amino acid sequence identity). *Li* and these related gene copies encode β-glucosidases belonging to the family 1 group of glycosyl hydrolases; these enzymes occur in almost all organisms and serve diverse biological functions (Barrett *et al*. 1995). Thus, the observation of multiple gene copies related to *Li* is not surprising. This is particularly true for *T. repens*, given its apparent origin as an allotetraploid.

Natural selection and the cyanogenesis polymorphism

Alternatives to selection at Li. Our molecular population genetic analyses are consistent with an inference of positive selection and a selective sweep at or near the *Li* locus (Table 3). Since *Li* is the only locus out of four examined to show these deviations from neutral equilibrium, and since there is no evidence for population structure in the sample set, these patterns are unlikely to be an artefact of population structure or other demographic factors, which would be expected to have genome-wide effects. Nonetheless, other alternatives to selection should be considered as potential determinants of the patterns of nucleotide diversity observed at *Li.*

First, as an allotetraploid, white clover likely possesses many pairs of homeologous genes. If sequences from more than one gene were mistakenly incorporated into diversity estimates at the three other sequenced loci (*ACO1*, *ALDP, ZIP*), then estimates of nucleotide diversity at these loci would be artificially inflated above that of *Li*. We do not believe that this has occurred. For all three of these genes (and *Li*), no more than two sequence haplotypes (alleles) were detected per accession from among the pool of sequenced clones (≥ 8 per accession), a strong indication that no more than one locus was sequenced per gene. In addition, for two of the genes (*ALDP* and *ZIP*), we were able to successfully amplify the putative homeologues using modified PCR conditions; in both cases, the homeologue PCR product was distinguishable by its size, and DNA sequencing revealed divergence in intron size and DNA sequence in comparison to our targeted loci.

Second, the *Li* sequence analyses were performed with the assumption that all *Li*__ plants carry two copies of the gene. If some *Li*_ plants were actually *Li/li* heterozygotes (i.e. hemizygotes for the presence of the gene), then sample sizes used in analyses of nucleotide diversity could be artificially inflated, since we assumed $2 \times 22 = 44$ *Li* haplotypes. To test whether undetected hemizygosity has affected our inferences, we repeated the nucleotide diversity-based tests of selection using the conservative assumption that only accessions possessing two different *Li* haplotypes are *Li/Li* homozygotes, and that those possessing one *Li* haplotype are hemizygotes. Twelve accessions possess two *Li* haplotypes, and 10 possess one haplotype (see Table 1). Using this reduced sample size of $N = 34$ sequences, we find that Tajima's D and Fay & Wu's H are both still significant at $P < 0.05$. Thus, our inference of selection at *Li* does not appear to be an artefact of an inflated sample set.

Recombination is another factor that could potentially affect our nucleotide diversity estimates. Levels of nucleotide diversity are often correlated with recombination rates (e.g. Roselius *et al*. 2005 and references therein). Since *Li* is apparently absent from the genomes of plants lacking linamarase (Fig. 2), the effective population size of this gene is lower than that of other loci, potentially lowering the effective recombination rate and associated levels of nucleotide diversity. We do not believe that this effect can account for the deviations from neutral expectations that we observed. First, Fay & Wu's (2000) H test, which uses outgroup comparisons to determine ancestral vs. derived nucleotide states at polymorphic sites, reveals a significant excess of high-frequency derived substitutions at *Li*; an excess of derived substitutions is inconsistent with suppressed nucleotide diversity. In addition, many of the cyanogenic plants in this study originate from geographical regions with mild winter climates, where populations are often monomorphic for the *Li* allele (e.g. Daday 1958); thus, hemizygosity rates for *Li* may not be high enough across the species range to substantially affect the effective recombination rate at this locus.

Adaptive evolution of Li. The inferred positive selection at *Li* is consistent with a scenario whereby a mutation arose which conferred a selectively favoured cyanogenic phenotype. One possibility would be the occurrence of a mutation that improved the enzyme's efficiency, thereby making it less costly for the plant to produce it. In plants with functional *Li* alleles, the linamarase enzyme constitutes up to 5% of the soluble leaf protein, which represents large energetic investment. The reproductive output of *li/li* plants (as measured by floral mass) has been shown to be greater than that of *Li/li* plants (Kakes 1989), a pattern suggesting clear energetic costs associated with linamarase production. Thus, any changes that could increase the protein's efficiency would likely be strongly favoured.

None of the amino acid replacement polymorphisms observed in the sampled clover plants is a clear candidate to have been the target of this inferred positive selection. When selection favours a new mutation, the derived (rather than ancestral) amino acid residue is expected to rise to high frequency in the population. However, outgroup comparisons indicate that for four of five polymorphic sites, it is the ancestral residue that occurs at high frequency; at the fifth site, the two residues occur at approximately equal frequencies. Moreover, none of the amino acid replacement polymorphisms observed in *Li* correspond to sites identified as key positions for protein function, based on the crystal structure of the linamarase protein (Barrett *et al*. 1995).

Beyond the *Li* coding region itself, the target of selection could also have involved *cis*-regulatory regions (or other nonprotein-coding portions of *Li* that were not sequenced), or possibly a different locus altogether that is in strong linkage disequilibrium with *Li*. Of these alternatives, the latter may be the least likely. As a self-incompatible, obligate outcrosser, *T. repens* is characterized by high genetic diversity and high recombination rates (Barrett *et al*. 2004); LD would therefore be expected to decline rapidly around the target of selection in this species (Stephan *et al*. 2006).

Balancing selection for the Li/li *polymorphism*

Ecological context. Our inference of positive selection associated with functional *Li* alleles is distinct from, but not incompatible with, the long-standing ecological evidence that balancing selection acts to maintain the *Li/li* polymorphism. There is an abundance of evidence that under conditions where generalist herbivores are present, functional *Li* alleles are strongly favoured in clover populations (reviewed by Hughes 1991; Hayden & Parker 2002). For example, in a mixed planting experiment using individuals with or without functional *Li* alleles (nearly all having functional *Ac* alleles), Ennos (1981) documented decreased survivorship of plants lacking linamarase and calculated a selection coefficient of $s = 0.3$ against the individuals lacking functional *Li* alleles. In a separate study of naturally polymorphic populations, Ennos (1982) demonstrated that functional alleles of *Ac* and *Li* are in linkage disequilibrium, even though the genes are independently segregating and populations show random mating; this finding again suggests that there is selection favouring the occurrence of cyanogenic plants, since a functional allele is required at both loci for HCN release. Feeding experiments have further demonstrated that cyanogenesis is selectively favoured in the presence of herbivores. For example, in choice experiments, both slugs and voles avoid cyanogenic white clover plants; if forced to feed on them they reduce their overall food consumption (Dirzo & Harper 1982a; Saucy *et al*. 1999). In mixed plantings with herbivores present, cyanogenic plants show greater seedling survivorship than acyanogenic morphs as a consequence of reduced herbivore damage (Dirzo & Harper 1982b).

Evidence that selection favours acyanogenic plants comes largely from observations of clinal variation in cyanogenesis in natural populations. Frequencies of acyanogenic plants are positively correlated with lower minimum winter temperatures, leading to altitudinal and latitudinal clines that occur both on a continental scale (e.g. Europe, North America; Daday 1958) and more localized geographical scales in mountainous areas (e.g. Daday 1954b; Majumdar *et al*. 2004). These clines occur not only in native Eurasian populations (e.g. De Araujo 1976; Till-Bottraud *et al*. 1988; Majumdar *et al*. 2004), but also where the species has been introduced and naturalized within the last 500 years (e.g. Daday 1958; Fraser 1986; Ganders 1990). This rapid evolution of clines following introduction suggests that balancing selection on the cyanogenesis polymorphism is quite strong. While the factors favouring acyanogenic morphs in colder climates remain unclear, they may be related to cyanogenic autotoxicity in areas with frequent frost-induced cell rupture (e.g. Daday 1965; Dirzo & Harper 1982b), or resource allocation trade-offs in areas of low herbivore abundance (e.g. Foulds & Grime 1972; Kakes 1989). The high energetic costs associated with linamarase synthesis (see above) suggest that the latter explanation may be especially likely.

Molecular evolutionary context. Balancing selection acts to maintain two or more divergent classes of alleles, and at the molecular level, the signature of balancing selection is predicted to be an excess of intermediate-frequency polymorphisms over neutral equilibrium expectations. Since *li* alleles correspond to an apparent gene deletion (Fig. 2), balancing selection for the *Li/li* polymorphism cannot be documented by examining nucleotide variation within the gene. An alternative approach would be to identify and sequence the genomic regions flanking *Li* in plants with and without the gene, to look for the molecular signature of a balanced polymorphism in these flanking

regions. This approach has been employed in *Arabidopsis thaliana*, and signatures of balancing selection have been detected at several presence/absence loci examined to date (Stahl *et al*. 1999; Tian *et al*. 2002; Shen *et al*. 2006; reviewed by Tiffin & Moeller 2006). Interestingly, all such instances of balancing selection have involved *Arabidopsis* R genes, which function in pathogen recognition and initiation of plant defence responses; thus, as with *Li/li*, these previously documented presence/absence balanced polymorphisms have evolved in response to natural selection acting on plant defence mechanisms. Beyond presence/absence polymorphisms, signatures of balancing selection have also been detected in other *A. thaliana* defence genes, including both R genes (e.g. *RPS2*; Caicedo *et al*. 1999), and the glucosinolate biosynthesis gene *MAM2* (Kroymann *et al*. 2003).

Conclusions

Over the last decade, molecular population genetics has begun to reveal important insights into the mechanisms of adaptive evolution. The full potential of this approach will be realized as research moves beyond model genetic systems into species that are ecologically well characterized. This study offers a step in that direction. We have examined the genetic basis and molecular evolution of a textbook example of an adaptive polymorphism. Our findings indicate that the *Li/li* polymorphism in white clover likely reflects the presence/absence of the *Li* gene, and that recent selection on *Li* has left the molecular signature of a selective sweep at this locus. It will be interesting to compare the molecular evolution of *Li* with that of other candidate genes in this ecological model system.

Acknowledgements

The authors thank Pierre-François Perroud and Young-Sook You for technical advice with Southern hybridizations; Stuart McDaniel and Richard Moore for helpful discussions; Katie Hyma for computational assistance with the CLR test; and Washington University, which funded this work through funds to K. Olsen.

References

Armstrong HE, Armstrong EF, Horton E (1913) Herbage Studies. II.-Variation in *Lotus corniculatus* and *Trifolium repens*: (cyanophoric plants). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **86**, 262–269.

Badr A, Sayed-Ahmed H, El-Shanshouri A, Watson LE (2002) Ancestors of white clover (*Trifolium repens* L.), as revealed by isozyme polymorphisms. *Theoretical and Applied Genetics*, **106**, 143–148.

Barrett B, Griffiths A, Schreiber M *et al*. (2004) A microsatellite map of white clover. *Theoretical and Applied Genetics*, **109**, 596–608.

Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA (1995) The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure*, **3**, 951–960.

Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.

Caicedo AL, Schaal BA, Kunkel BN (1999) Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **96**, 302–306.

Clegg MT, Durbin ML (2003) Tracing floral adaptations from-ecology to molecules. *Nature Reviews Genetics*, **4**, 206–215.

Coop IE (1940) Cyanogenesis in white clover III. Study of Linamarase. *New Zealand Journal of Science and Technology*, **B22–23**, 71–83.

Corkill L (1942) Cyanogenesis in white clover (*Trifolium repens*) V. The inheritance of cyanogenesis. *New Zealand Journal of Science and Technology*, **B23**, 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 AM (1976) The relationship between altitude and cyanogenesis in white clover (*Trifolium repens* L.). *Heredity*, **37**, 291–293.

Dirzo R, Harper JL (1982a) Experimental studies on slug–plant interactions. III. Differences in the acceptability of individual plants of *Trifolium repens* to slugs and snails. *Journal of Ecology*, **70**, 101–117.

Dirzo R, Harper JL (1982b) Experimental studies on slug–plant interactions. IV. The performance of cyanogenic and acyanogenic morphs of *Trifolium repens* in the field. *Journal of Ecology*, **70**, 119–138.

Dirzo R, Sarukhan J (1984) *Perspectives on Plant Population Ecology*. Sinauer & Associates, Sunderland, Massachusetts.

Ellison NW, Liston A, Steiner JJ, Williams WM, Taylor NL (2006) Molecular phylogenetics of the clover genus (*Trifolium* — Leguminosae). *Molecular Phylogenetics and Evolution*, **39**, 688–705.

Ennos RA (1981) Detection of selection in populations of white clover (*Trifolium repens* L.). *Biological Journal of the Linnean Society*, **15**, 75–82.

Ennos RA (1982) Association of the cyanogenic loci in white clover. *Genetical Research, Cambridge*, **40**, 65–72.

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.

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

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

Fraser J (1986) Cyanogenesis in naturalised white clover (*Trifolium repens* L.) populations in Atlantic Canada. *Progress Report, Clovers and Special Purpose Legumes Research*, **19**, 12–14.

Ganders FR (1990) Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. *Heredity*, **64**, 387–390.

Hahn MW, Rausher MD, Cunningham CW (2002) Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics*, **161**, 11–20.

Hayden KJ, Parker IM (2002) Plasticity in cyanogenesis of *Trifolium repens* L. inducibility, fitness costs and variable expression. *Evolutionary Ecology Research*, **4**, 155–168.

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

Hughes MA (1993) Molecular genetics of plant cyanogenic β-glucosidases. *ACS Symposium Series*, **533**, 153–169.

Hughes MA, Dunn MA (1982) Biochemical characterization of the *Li* locus, which controls the activity of the cyanogenic β-glucosidase in *Trifolium repens* L. *Plant Molecular Biology*, **1**, 169–181.

Hughes MA, Sharif AL, Dunn MA, Oxtoby E, Pancoro A (1990) Restriction fragment length polymorphism analysis of the *Li* locus in *Trifolium repens*. *Plant Molecular Biology*, **14**, 407–414.

Jensen JD, Kim Y, DuMont VB *et al*. (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, **170**, 1401–1410.

Kakes P (1989) An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theoretical and Applied Genetics*, **77**, 111–118.

Kakes P (1997) Difference between the male and female components of fitness associated with the gene *Ac*. *Trifolium repens*. *ACTA Botanica Neerlandica*, **46**, 219–223.

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765–777.

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proceedings of the National Academy of Sciences, USA*, **100** (Suppl. 2), 14587–14592.

Majumdar S, De KK, Banjeree S (2004) Influence of two selective forces on cyanogenesis polymorphism of *Trifolium repens* L. in Darjeeling Himalaya. *Journal of Plant Biology*, **47**, 124–128.

McDaniel SF, Shaw AJ (2005) Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. *Molecular Ecology*, **14**, 1121–1132.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.

Melville J, Doak BW (1940) Cyanogenesis in white clover II. Isolation of glucoside constituents. *New Zealand Journal of Science and Technology*, **B22**, 67–70.

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

Oxtoby E, Dunn MA, Pancoro A, Hughes MA (1991) Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (*Trifolium Repens* L.) *Plant Molecular Biology*, **17**, 209–219.

Pederson GA, Brink GE (1998) Cyanogenesis effect on insect damage to seedling white clover in a bermudagrass sod. *Agronomy Journal*, **90**, 208–210.

Porebski SL, Bailey G, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, **15**, 8–15.

Richards AJ, Fletcher A (2002) The effects of altitude, aspect, grazing and time on the proportion of cyanogenics in neighbouring populations of *Trifolium repens* L. (white clover). *Heredity*, **88**, 432–436.

Roselius K, Stephan W, Städler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics*, **171**, 753–763.

Rozas J, Sánchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

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

Shen JD, Araki H, Chen LL, Chen JQ, Tian DC (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics*, **172**, 1243–1250.

Silvertown J, Charlesworth D (2001) *Introduction to Plant Population Biology*, 4th edn. Blackwell Science, Oxford, UK.

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

Stephan W, Song YS, Langley CH (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, **172**, 2647–2663.

Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0b10. Sinauer & Associates, Sunderland, Massachusetts.

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

Tian DC, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA*, **99**, 11525–11530.

Tiffin P, Moeller DA (2006) Molecular evolution of plant immune system genes. *Trends in Genetics*, **22**, 662–670.

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

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

Ware WM (1925) Experiments and observations on forms and strains of *Trifolium repens*. *Journal of Agricultural Science*, **15**, 47–67.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Williams WM, Williamson ML (2001) Genetic polymorphism for cyanogenesis and linkage at the linamarase locus in *Trifolium nigrescens* Viv. subsp. *nigrescens*. *Theoretical and Applied Genetics*, **103**, 1211–1215.

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution*, **22**, 506–519.

Zufall RA, Rausher MD (2004) Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature*, **428**, 847–850.

Kenneth Olsen is an assistant professor in the Biology Department of Washington University in St. Louis. His research focuses on plant evolutionary biology, including the genetics of crops domestication and adaptive evolution in wild species. Brittany Sutherland performed RT-PCR analyses for this study while a first-year graduate student at Washington University. Linda Small is a research technician in the Olsen lab who performed DNA sequencing and Southern hybridisations for this project.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Southern hybridizations of accessions with and without linamarase activity (+ and – respectively) probed with *Li* probe L2. Accessions are identified in Table 1. *Afl*II is predicted to have no cut sites within the probed regions of either *Li* or the *Li* paralogue. *Ase*I is predicted to have one detectable cut site in both *Li* and the *Li* paralogue within the probed region (a second cut site occurs 6 bp from the 5′ end of the probe in both genes).

Table S1 Primers used in PCR, DNA sequencing, PCR of probes for Southerns, and RT–PCR.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03506.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.