JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12466

SHORT COMMUNICATION

Adaptive cyanogenesis clines evolve recurrently through geographical sorting of existing gene deletions

N. J. KOOYERS[1] & K. M. OLSEN

Department of Biology, Washington University in St. Louis, St. Louis, MO, USA

Keywords:

adaptive cline;
copy number variation;
cyanogenesis;
parallel evolution;
standing variation.

Abstract

Identifying the genetic basis of parallel phenotypic evolution provides insight into the process of adaptation and evolutionary constraint. White clover (*Trifolium repens*) has evolved climate-associated adaptive clines in cyanogenesis (the ability to produce hydrogen cyanide upon tissue damage) in several world regions where it has been introduced. Gene-deletion polymorphisms at the *CYP79D15* and *Li* loci underlie the presence/absence of the cyanogenic phenotype. Both loci have undergone multiple independent gene-deletion events, which are identifiable through molecular signatures in flanking regions. To investigate whether cyanogenesis clines in introduced populations have evolved through the sorting of standing genetic variation or *de novo* gene deletions, we examined cyanogenesis gene-flanking regions in three world regions. In comparison with native Eurasian populations, we find no evidence for novel gene deletion events in any introduced region, which suggests that these adaptive clines have evolved through the geographical sorting of pre-existing genetic variation.

Introduction

Parallel phenotypic evolution has long been used to study the process of adaptation (e.g. Oakeshott *et al.*, 1982; Losos *et al.*, 1998; Rundle *et al.*, 2000; Elmer *et al.*, 2010). Until recently, studies were rarely able to identify the molecular basis of phenotypic evolution; however, the advent of modern sequencing technology and association mapping techniques has begun to improve our understanding of the genetic variation that encodes ecologically important traits (e.g. Lowry & Willis, 2010; Jones *et al.*, 2012; Prasad *et al.*, 2012). This new information resurrects several interesting evolutionary questions that were posed before data existed to address them. For instance, are parallel phenotypes in different populations of a species always a product of the same genetic mechanisms? And do parallel phenotypic patterns in different geographical regions evolve from standing genetic variation in a common ancestor or from novel mutations in each area?

Several recent studies have attempted to address these questions (reviewed in Arendt & Reznick, 2008; Manceau *et al.*, 2010; Elmer & Meyer, 2011). Identical mutational mechanisms frequently have been found to underlie parallel phenotypic evolution (Cooper *et al.*, 2003; Nachman *et al.*, 2003; Turner *et al.*, 2008; Paaby *et al.*, 2010). However, there are additional cases where shared phenotypes in geographically disparate populations arise through different mutations at a given locus (e.g. Gross *et al.*, 2009; Chan *et al.*, 2010; reviewed by Martin & Orgogozo, 2013) or through mutations at different underlying loci (Steiner *et al.*, 2009; Samis *et al.*, 2012). Characterizing the genomic and geographical context of parallel phenotypic evolution can provide clues about the nature of evolutionary constraint and the mechanisms underlying adaptation.

Cyanogenesis, the ability to produce hydrogen cyanide after tissue damage, occurs as a discrete polymorphism in white clover (Armstrong *et al.*, 1913; Ware, 1925). Cyanogenic plants deter herbivores by causing reduced cellular respiration in herbivores that ingest cyanogenic plants. The cyanogenic phenotype requires

Correspondence: Kenneth M. Olsen, Biology Department, Washington University in St. Louis, One Brookings Drive, Campus Box 1137, Saint Louis, MO 63130, USA.
Tel.: +1-314-935-7013; fax: +1-314-935-4432;
e-mail: kolsen@wustl.edu
[1]Present address: Department of Biology, University of Virginia, Charlottesville, VA 22904, USA

two biochemical components, cyanogenic glucosides and their hydrolyzing enzyme (linamarase). Each of these components may be either present or absent in a given plant, and their inheritance can be described by two independently segregating Mendelian polymorphisms; *Ac/ac* and *Li/li* correspond, respectively, to the genes controlling the presence/absence of cyanogenic glucosides and linamarase, where the dominant allele corresponds to the presence of each component (Coop, 1940; Melville & Doak, 1940; Hughes, 1991). At the molecular level, the *Ac/ac* polymorphism corresponds to a gene-deletion polymorphism at *CYP79D15*, which encodes the cytochrome P450 responsible for the first step in cyanogenic glucoside biosynthesis (Olsen *et al.*, 2008, 2013). The *Li/li* polymorphism reflects a gene-deletion polymorphism at the *Li* locus, which encodes the linamarase protein (Olsen *et al.*, 2007, 2013). In a survey of USDA white clover accessions, derived mostly from native Eurasian populations, we have found that the genomic span of the cyanogenesis gene deletions varies in size at both the *CYP79D15* and *Li* loci; this pattern suggests that there have been multiple mutational origins of the gene-deletion alleles at both loci (Olsen *et al.*, 2013).

The cyanogenesis polymorphism is spatially distributed as adaptive clines, with higher frequencies of cyanogenic plants occurring in warmer and more arid climates (e.g. Daday, 1954a; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Kooyers *et al.*, 2014). Cyanogenesis clines occur not only in white clover's native Eurasian range (e.g. Daday, 1954a,b), but also in several world regions where it has been introduced and naturalized within the last 500 years (Daday, 1958; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013). For example, we have documented a latitudinal cline across the central USA, with >85% cyanogenic plants in southern Louisiana and a steady decline to 11% in northern Wisconsin (Kooyers & Olsen, 2012). The rapid evolution of adaptive clines in these introduced populations indicates that there are strong selective forces maintaining the cyanogenesis polymorphism. Several factors have been implicated; these include evidence that regional variation in herbivore pressure and/or drought stress, and energetic tradeoffs associated with cyanogenic component production can maintain cyanogenesis clines (reviewed in Hughes, 1991; Kooyers & Olsen, 2013).

As we know that the cyanogenesis polymorphism has evolved through multiple gene deletions at *CYP79D15* and *Li* (Olsen *et al.*, 2013), the evolution of cyanogenesis clines in introduced regions could potentially occur through two mechanisms: (i) geographical sorting of pre-existing gene-deletion alleles that were introduced with the species introduction or (ii) *de novo* gene deletions that have occurred after species introduction. There are clear predictions for each mechanism. If there has been evolution from pre-existing variation, one would expect to find the same gene-deletion sizes shared throughout the range of the species. If clines in introduced regions have instead evolved through selection on newly arisen gene-deletion alleles, then one would expect to find some novel deletion sizes that are unique to a given region.

Materials and methods

To test for evidence of novel gene deletions in populations where white clover has been introduced, we sampled plants from the central USA (n = 30), the South Island of New Zealand (n = 31) and the US Pacific Northwest (n = 35; Tables S1 and S2; see Kooyers & Olsen, 2013 for details on sampling and cyanogenesis

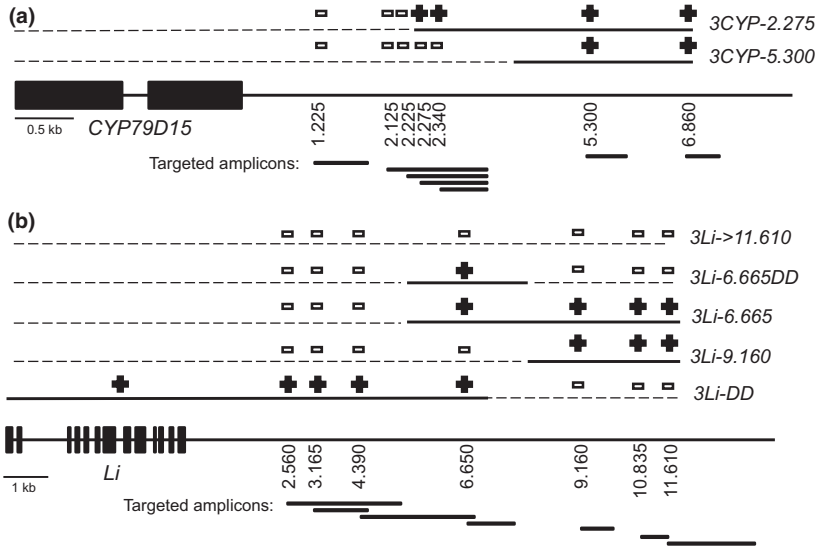

Fig. 1 Schematic depiction of the 3′ flanking regions of the (a) *CYP79D15* and (b) *Li* cyanogenesis genes. Numbers downstream of each gene indicate locations of 5′ ends of forward primers used in screening for amplicon presence. Targeted amplicon sizes and locations are indicated by horizontal bars and are approximately to scale. Inferred gene-deletion alleles are indicated above each gene, with allele names indicated to the right. Plus and minus signs indicate the presence or absence of each targeted amplicon.

Table 1 Numbers of individuals in each *CYP79D15* gene-deletion class.

	Ac gene present (control)	*3CYP-2.275*	*3CYP-5.300*	Total
Central USA	7	22	0	29
New Zealand	3	24	4	31
US Pacific Northwest	7	25	3	35
Eurasia (native range)*	45	14	2	61

*United States Department of Agriculture accessions from Olsen *et al.* (2013).

phenotyping). Adaptive clines have been documented in all three regions (Kooyers & Olsen, 2012; Kooyers & Olsen, 2013). To characterize and distinguish among different gene-deletion alleles, we used sequences downstream of each cyanogenesis gene to design primer pairs for PCR assays at increasing distances from each gene; the presence or absence of amplicons was used to assess the 3′ span of cyanogenesis gene deletions. The upstream boundaries of *CYP79D15* and *Li* gene deletions were not assessed due to a lack of information on the upstream flanking sequences (see Olsen *et al.*, 2013). For *CYP79D15*, forward primers of each primer pair were located approximately 1.225 kb, 2.125 kb, 2.225 kb, 2.275 kb, 2.340 kb and 5.300 kb downstream of the *CYP79D15* stop codon (Fig. 1a; primer and PCR information is indicated in Table S3). Attempted amplifications of additional primer pairs at 2.405 kb and 4.060 kb were unsuccessful due to nonspecific amplification; thus, our PCR assays do not provide resolution for determining whether there might be novel-deletion sizes within the window between 2.340 kb and 5.300 kb downstream of *CYP79D15*. For *Li*, amplicon forward primers were located approximately 2.560 kb, 3.165 kb, 4.390 kb, 6.665 kb, 9.160 kb, 10.835 kb and 11.610 kb downstream of the *Li* stop codon (Fig. 1b). We attempted to PCR-amplify every targeted amplicon in each sample except for the *Li* 11.610 kb downstream region (which matched results for 10.835 kb exactly in a subsample of individuals). Failure to amplify a product after three attempts was taken as evidence that the genomic region was not present in an individual. Amplification of a region was taken as evidence that this sequence was not within the boundaries of a gene deletion. This method parallels that used to detect gene-deletion allele sizes in Olsen *et al.* (2013) with clearer resolution due to the addition of new primer pairs.

Results and discussion

Among *ac* plants (i.e. plants lacking the *CYP79D15* gene), we identified two gene-deletion alleles (*3CYP-2.275* and *3CYP-5.300*), which correspond to the same two gene-deletion alleles identified in a previous survey of native Eurasian plants (Olsen *et al.*, 2013; Fig. 1a). The *3CYP-2.275*-deletion allele was present in 71 of the 78 *ac* plants in introduced regions (Table 1; Table S1). This allele corresponds to the *3CYP-2.34*-deletion allele described by Olsen *et al.* (2013) (here, we have narrowed the range of the potential deletion sizes with additional PCR sites; see Table S3). The *3CYP-5.300*-deletion allele was present in seven plants assayed; it corresponds to the second deletion allele described by Olsen *et al.* (2013) (with additional amplicon targeting in the present study reducing the range of gene-deletion boundary by 1.5 kb). Both alleles were found in New Zealand and the Pacific Northwest, whereas only the *3CYP-2.275* deletion was found in the central USA (Table 1; see also Table S1). The fact that we detected no novel gene-deletion sizes at *CYP79D15* suggests that the *ac* gene deletions occurred prior to population introductions into the non-native species range.

Among *li* plants (i.e. plants lacking the *Li* gene), we identified three gene-deletion alleles (*3Li-6.665*, *3Li-9.160*, *3Li->11.160*), two of which were detected previously in Eurasian plants (Olsen *et al.*, 2013; Fig. 1b). The *3Li-6.665* allele is present in 55 of 84 *li* plants in introduced populations (Table S2); it corresponds to the allele of the same name reported by Olsen *et al.* (2013). A variant of this allele with an additional downstream deletion was detected in five plants (*3Li-6.665DD*; Fig. 1, Table S2). In 19 *li* plants, the genomic deletion apparently extends more than 11.61 kb downstream of the *Li* stop codon (the limit

Table 2 Numbers of individuals in each *Li* gene-deletion class.

	Li gene present (control)	*3Li-6.665*	*3Li-6.665DD*	*3Li-9.160*	*3Li->11.610*	Total
Central USA	6†	20	0	1	3	30
New Zealand	5†	16	3	0	6	30
US Pacific Northwest	1	19	2	2	10	34
Eurasia (native range)*	38‡	13	4	0	10	65

*United States Department of Agriculture accessions from Olsen *et al.* (2013).
†Two plants carry *3Li-DD* downstream-deletion allele.
‡Seven plants carry *3Li-DD* downstream-deletion allele.

of our genome walking sequence data; *3Li->11.610*; Fig. 1b); this allele was also observed by Olsen *et al.* (2013). The *3Li*-9.160-deletion allele, which was not previously reported, was present in three *li* plants. Like *CYP79D15*, the *li* gene-deletion alleles are present in multiple disparate geographical areas (Table 2). The *3Li-6.665* and *3Li->11.160* alleles are both common in all areas, whereas the 3*Li-9.160* deletion is rare but present in both the central USA and US Pacific Northwest. These patterns suggest that, as with *ac* alleles, *li*-deletion alleles arose prior to population introductions into the non-native range. Interestingly, we also found a downstream deletion in some plants that possess the *Li* gene (*3Li-DD* allele; Fig. 1b). This allele was observed in both the central USA and New Zealand (Table 2, Table S2) and could plausibly reflect recombination between gene-presence and gene-deletion alleles.

The sharing of *ac* and *li* gene-deletion alleles across the native and introduced species range strongly suggests that cyanogenesis clines have evolved in introduced regions through the geographical sorting of pre-existing variation. However, it is important to note two caveats in this inference. First, our ability to detect novel-deletion alleles is limited by the number and location of regions targeted in PCR assays. As noted above, our PCR survey of the *CYP79D15* downstream region does not provide resolution that would allow the detection of any novel-deletion alleles in the 2.340–5.300 kb size range. Thus, region-specific *ac* alleles within this size range would remain undetected. Secondly, if there were mutational biases that favoured the repeated evolution of specific deletion sizes, any convergently evolved, region-specific alleles would be misinferred in our analysis to represent standing genetic variation. The probability of biases in gene-deletion sizes increases if there are regions of genomic instability such as those created by repetitive DNA (e.g. Jelesko *et al.*, 1999; Kuo *et al.*, 2006). Whereas no repeats were detected in the *Li* downstream sequence, BLAST analysis of the *CYP79D15* downstream sequence revealed a series of 10 tandem 180–200 bp repeats located 2.275–5.300 kb downstream of the *CYP79D15* stop codon. These repeats show a trend of increasing genetic divergence across the array (Table S4), a pattern potentially consistent with unequal crossing-over between adjacent repeats; this mutational bias could be a factor in the convergent evolution of *ac*-deletion alleles of the same size.

Although the possibility of undetected novel alleles cannot be excluded, our inference of cyanogenesis allele sharing among geographical regions is highly compatible with the known introduction history of *Trifolium repens*. White clover was intentionally and repeatedly introduced widely across North America, New Zealand and other temperate regions for use as a forage crop and lawn plant. This population history would be expected to maximize the probability of introducing standing genetic variation, both at cyanogenesis genes and across the genome. Indeed, population genetic surveys at neutral loci have revealed no detectable genetic bottlenecks associated with population introductions into the non-native range (Kooyers & Olsen, 2012, Kooyers & Olsen, 2013). The relatively recent introduction of the species (<500 years, with European colonization) would also provide relatively little time for the evolution of new *ac*- and *li*-deletion alleles through mutation. For this study system and others, we suggest that if standing genetic variation is abundantly represented in an introduced population, novel mechanisms of producing the same adaptive phenotypes may be irrelevant and lost in the process of local adaptation.

Acknowledgments

We appreciate the guidance of Keith Widdup and Brent Barrett while collecting plants in New Zealand. We thank the Washington University greenhouse for maintaining plant populations and Kate Waselkov and the Olsen Lab for improving this manuscript through their feedback. Funding was provided through an NSF CAREER award to KMO (DEB-0845497) and NSF Doctoral Dissertation Improvement Grant to NJK (DEB-1110588).

References

Arendt, J. & Reznick, D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**: 26–32.

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

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D. *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**: 302–305.

Coop, I.E. 1940. Cyanogenesis in white clover III. Study of linamarase. *N. Z. J. Sci. Technol.* **B22–23**: 71–83.

Cooper, T.F., Rozen, D.E. & Lenski, R.E. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **100**: 1072–1077.

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

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

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

Elmer, K.R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**: 298–306.

Elmer, K.R., Fan, S., Gunter, H.M., Jones, J.C., Boekhoff, S., Kuraku, S. *et al.* 2010. Rapid evolution and selection

inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Mol. Ecol.* **19**: 197–211.
Gross, J.B., Borowsky, R. & Tabin, C.J. 2009. A Novel Role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet.* **5**: e1000326.
Hughes, M.A. 1991. The cyanogenic polymorphism in *Trifolium repens* L (white clover). *Heredity* **66**: 105–115.
Jelesko, J.G., Harper, R., Furuya, M. & Gruissem, W. 1999. Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA.* **96**: 10302–10307.
Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J. *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.
Kooyers, N.J. & Olsen, K.M. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Mol. Ecol.* **21**: 2455–2468.
Kooyers, N.J. & Olsen, K.M. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover *(Trifolium repens L.)*. *Heredity* **111**: 495–504.
Kooyers, N.J., Gage, L.R., Al-Lozi, A. & Olsen, K.M. 2014. Aridity shapes cyanogenesis cline evolution in white clover *(Trifolium repens L.)*. *Mol. Ecol.* **23**: 1053–1070.
Kuo, H.F., Olsen, K.M. & Richards, E.J. 2006. Natural variation in a subtelomeric region of *Arabidopsis*: implications for the genomic dynamics of a chromosome end. *Genetics* **173**: 401–417.
Losos, J.B., Jackman, T.R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**: 2115–2118.
Lowry, D.B. & Willis, J.H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**: 1–14.
Manceau, M., Domingues, V.S., Linnen, C.R., Rosenblum, E.B. & Hoekstra, H.E. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **365**: 2439–2450.
Martin, A. & Orgogozo, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235–1250.
Melville, J. & Doak, B.W. 1940. Cyanogenesis in white clover II. Isolation of glucoside constituents. *N. Z. J. Sci. Technol.* **B22**: 67–70.
Nachman, M.W., Hoekstra, H.E. & D'Agostino, S.L. 2003. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. USA.* **100**: 5268–5273.
Oakeshott, J.G., Gibson, J.B., Anderson, P.R., Knibb, W.R., Anderson, D.G. & Chambers, G.K. 1982. Alcohol-dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution* **36**: 86–96.
Olsen, K.M., Sutherland, B.L. & Small, L.L. 2007. Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Mol. Ecol.* **16**: 4180–4193.
Olsen, K.M., Hsu, S.C. & Small, L.L. 2008. Evidence on the molecular basis of the *Ac/ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L). *Genetics* **179**: 517–526.
Olsen, K.M., Kooyers, N.J. & Small, L.L. 2013. Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Mol. Ecol.* **22**: 724–738.
Paaby, A.B., Blacket, M.J., Hoffmann, A.A. & Schmidt, P.S. 2010. Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents. *Mol. Ecol.* **19**: 760–774.
Prasad, K.V.S.K., Song, B.H., Olson-Manning, C., Anderson, J.T., Lee, C.R., Schranz, M.E. *et al.* 2012. A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* **337**: 1081–1084.
Rundle, H.D., Nagel, L., Boughman, J.W. & Schluter, D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**: 306–308.
Samis, K.E., Murren, C.J., Bossdorf, O., Donohue, K., Fenster, C.B., Malmberg, R.L. *et al.* 2012. Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. *Ecol. Evol.* **2**: 1162–1180.
Steiner, C.C., Römpler, H., Boettger, L.M., Schöneberg, T. & Hoekstra, H.E. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol. Biol. Evol.* **26**: 35–45.
Turner, T.L., Levine, M.T., Eckert, M.L. & Begun, D.J. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* **179**: 455–473.
Ware, W.M. 1925. Experiments and observations on forms and strains of *Trifolium repens*. *J. Agric. Sci.* **15**: 47–67.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Results from PCR amplifications of markers in the downstream flanking region of *CYP79D15*.

Table S2 Results from PCR amplifications of markers in the downstream flanking region of *Li*.

Table S3 Primers used for amplifying flanking regions of *CYP79D15* and *Li*.

Table S4 Summary of tandem repetitive sequences found between 2340 and 5300 bp downstream of *CYP79D15*.

Received 13 March 2014; revised 15 June 2014; accepted 22 July 2014