The evolutionary appearance of non-cyanogenic hydroxynitrile glucosides in the *Lotus* genus is accompanied by the substrate specialisation of paralogous β-glucosidases resulting from a crucial amino acid substitution

Daniela Lai[1], Maher Abou Hachem[2], Fran Robson[3], Carl Erik Olsen[1], Trevor L. Wang[3], Birger Lindberg Møller[1], Adam M. Takos[1], and Fred Rook[1*]

[1]Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg, Denmark,
[2]Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark,
[3]Department of Metabolic Biology, John Innes Centre, NR4 7UH Norwich, United Kingdom

[*]For correspondence: Fred Rook, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark. Tel: +45 3533 3343, Fax: +45 3533 3300, Email: frro@plen.ku.dk

Email addresses:
Daniela Lai: dal@plen.ku.dk
Maher Abou Hachem: maha@bio.dtu.dk
Fran Robson: fran.robson@jic.ac.uk
Carl Erik Olsen: ceo@plen.ku.dk
Trevor L. Wang: trevor.wang@jic.ac.uk
Birger Lindberg Møller: blm@plen.ku.dk
Adam M. Takos: adta@plen.ku.dk

Running title: *L. japonicus* β-glucosidase specialisation

Keywords: *Lotus japonicus*, β-glucosidase, enzyme evolution, cyanogenesis, hydroxynitrile glucosides, rhodiocyanosides, *Lotus burttii*, *Lotus krylovii*, *Lotus filicaulis*, *Lotus corniculatus*

Word count: Total 7219, Summary 205, Introduction 904, Results 3108, Discussion 1293, Experimental Procedures 959, Acknowledgements 82, Table titles 21, Figure Legends 647.

SUMMARY

Lotus japonicus, like several other legumes, biosynthesizes the cyanogenic α-hydroxynitrile glucosides lotaustralin and linamarin. Upon tissue disruption these compounds are hydrolysed by a specific β-glucosidase, resulting in the release of hydrogen cyanide. *L. japonicus* also produces the non-cyanogenic γ- and β-hydroxynitrile glucosides rhodiocyanoside A and D using a biosynthetic pathway that branches off from lotaustralin biosynthesis. We previously established that BGD2 is the only β-glucosidase responsible for cyanogenesis in leaves. Here we show that the paralogous BGD4 has the dominant physiological role in rhodiocyanoside degradation. Structural modelling, site-directed mutagenesis, and activity assays establish that a glycine residue (G211) in the aglycone binding site of BGD2 is essential for its ability to hydrolyse the endogenous cyanogenic glucosides. The corresponding valine (V211) in BGD4 narrows the active site pocket, resulting in the exclusion of non-flat substrates such as lotaustralin and linamarin but not of the more planar rhodiocyanosides. Rhodiocyanosides and the *BGD4* gene only occur in *L. japonicus* and a few closely related species associated with the *L. corniculatus* clade within the *Lotus* genus. This suggests the evolutionary scenario that substrate specialization for rhodiocyanosides evolved from a promiscuous activity of a progenitor cyanogenic β-glucosidase resembling BGD2 and required no more than a single amino acid substitution.

INTRODUCTION

A wide chemical diversity of bioactive compounds has evolved in plants and one of their functions is to provide protection against herbivores or microbial pathogens. The amino acid-derived hydroxynitrile glucosides are a large class of such specialized metabolites found in many plant species (Nielsen *et al.*, 2002; Bjarnholt and Møller, 2008). The α-hydroxynitrile glucosides in this class are known as cyanogenic glucosides. In the plant, cyanogenic glucosides and their activating β-glucosidases are stored separately, constituting a two-component defence system. Tissue disruption, by for instance feeding herbivores, brings the hydroxynitrile glucosides into contact with their activating β-glucosidases. Hydrolysis of the *O*-β-glucosidic bond results in the release of the hydroxynitrile aglycone, which in the case of an α-hydroxynitrile will dissociate with the release of hydrogen cyanide (HCN).

In the legume model *Lotus japonicus*, the two most abundant hydroxynitrile glucosides are the isoleucine derived cyanogenic glucoside lotaustralin and the non-cyanogenic γ-hydroxynitrile glucoside rhodiocyanoside A (Forslund *et al.*, 2004; Figure 1). Also present are linamarin, a valine derived cyanogenic glucoside, and the isoleucine derived β-hydroxynitrile glucoside rhodiocyanoside D. Although the precise role of rhodiocyanosides in plant chemical defence remains to be established, the aglycone of rhodiocyanoside A is able to cyclize and form a furanone with potentially antimicrobial properties (Bjarnholt and Møller, 2008; Saito *et al.*, 2012). In *L. japonicus*, a gene cluster on chromosome 3 contains most of the biosynthetic genes required for the production of hydroxynitrile glucosides, and includes members of two cytochrome P450

and one glucosyltransferase gene family (Forslund *et al.*, 2004; Takos *et al.*, 2011; Takos and Rook, 2012). Linamarin and lotaustralin are the main cyanogenic glucosides found in other legume species such as white clover (*Trifolium repens*) and lima bean (*Phaseolus lunatus*) (Butler, 1965). In contrast, rhodiocyanosides have only been reported to occur in two *Lotus* species, apart from their presence in the unrelated genera *Ribes* and *Rhodiola*, and are thought to have evolved more recently from the *Lotus* cyanogenic glucoside pathway as some of the biosynthetic enzymes are shared (Bjarnholt *et al.*, 2008; Takos *et al.*, 2011).

The β-glucosidases (β-D-glucoside glucohydrolase, EC 3.2.1.21) that hydrolyse *O*-linked β-glucosidic bonds belong to the glucoside hydrolase family 1 (GH1, www.cazy.org). The three dimensional structure of these enzymes consists of a typical $(\beta/\alpha)_8$ TIM barrel fold with an active site pocket that contains two invariant glutamic acid residues (Barrett *et al.*, 1995; Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004; Sue *et al.*, 2006; Barleben *et al.*, 2007). Hydrolysis of the glucosidic bond involves a sequential two-step reaction with one glutamic acid serving as an acid/base catalyst and the second glutamic acid acting as a nucleophile that displaces the aglycone and forms a glycosyl enzyme intermediate, which is subsequently hydrolysed by water. The first crystal structure of a GH1 enzyme to be determined was from the β-glucosidase responsible for cyanogenesis in white clover (*T. repens*), and three-dimensional structures have now been solved for a number of β-glucosidases involved in the activation of plant chemical defence compounds (Barrett *et al.*, 1995, Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004; Sue *et al.*, 2006; Barleben *et al.*, 2007). It is proposed that substrate specificity of β-glucosidases depends on the recognition of the aglycone moiety, as binding of the

glucose moiety involves highly conserved amino acids and is essentially invariant. Non-conserved amino acid residues that form the aglycone-binding site and are important for substrate specificity have been reported for the maize β-glucosidase ZmGlu1, cassava linamarase, and sorghum SbDhr1 (Czjzek *et al.*, 2000; Keresztessy *et al.*, 2001; Verdoucq *et al.*, 2003; Verdoucq *et al.*, 2004). Two closely related β-glucosidase enzymes, BGD2 and BGD4 (85% amino acid identity), were implicated in hydroxynitrile glucoside catabolism in *L. japonicus*, and a genetic study revealed that only BGD2 was involved in cyanogenesis (Morant *et al.*, 2008; Takos *et al.*, 2010). While rhodiocyanoside A was an *in vitro* substrate for both β-glucosidases, BGD4 showed no activity on the cyanogenic glucosides lotaustralin and linamarin.

Evolution of metabolic pathways for new chemical defence compounds is thought to occur by recruitment of existing enzymes, either from general or specialized metabolism, and often involves a process of gene duplication and the selection for new functionalities (Pichersky and Gang, 2000; Ober *et al.*, 2010). Divergent evolution of enzyme function within the same family has been proposed to find its basis in a promiscuous activity of the progenitor enzyme and the selection for more specialized enzymes (Jensen 1976; Khersonsky and Tawfik, 2010). The evidence for this is considered to be largely circumstantial and identification of the mutational paths that lead to enzyme divergence presents a major challenge (Khersonsky and Tawfik, 2010; Weng *et al.*, 2012). In this study we provide experimental evidence that the ability to biosynthesize rhodiocyanosides is restricted to a group of closely related species within the *Lotus* genus, with the *BGD4* transcript detected in only three of the species. Mutants defective in the *BGD4* gene, obtained by TILLING, demonstrate that BGD4 has a

dominant role in the bio-activation of rhodiocyanosides in *L. japonicus*. While BGD2 is able to catalyse the hydrolysis of all endogenous hydroxynitrile glucosides, BGD4 activity is restricted to the non-cyanogenic rhodiocyanosides. The presence of a valine residue (V211) in the aglycone binding pocket of BGD4, corresponding to a glycine residue (G211) in BGD2, largely determines the narrower substrate range of this enzyme. These observations suggest a mutational path for the functional divergence of β-glucosidases during the evolution of hydroxynitrile glucoside metabolism in *Lotus* spp.

RESULTS

Genetic evidence for BGD4 mediated catabolism of non-cyanogenic hydroxynitrile glucosides

A genetic approach previously established that the *BGD2* gene encodes the cyanogenic β-glucosidase in the leaves of *L. japonicus* (Takos *et al.*, 2010). The paralogous *BGD4* gene was not involved in cyanogenesis, but rhodiocyanoside A was efficiently hydrolysed by BGD4 *in vitro*. To confirm that BGD4 is involved in the catabolism of non-cyanogenic hydroxynitrile glucosides in *L. japonicus*, we used the reverse genetics approach of TILLING (targeting induced local lesions in genomes; McCallum *et al.*, 2000) to identify mutations in *BGD4*.

The mutagenized *L. japonicus* accession MG-20 population used in the forward genetic screen for *cyanogenesis deficient* mutants was also used as the basis for a second *L. japonicus* TILLING population in addition to the one available in the Gifu accession (Perry *et al.*, 2003; Takos *et al.*, 2010; http://revgenuk.jic.ac.uk/). The gene structures of *BGD2* and *BGD4* consist of 13 exons dispersed over 7 kb of genomic sequence. A 1074 bp region covering both exon 7 and 8 was used as target fragment to identify mutations in the coding region of *BGD4*. In a TILLING screen of 3072 plants, a total of 15 lines containing point mutations in this region were identified (Figure S1), of which 7 lines contained amino acid substitutions. Viable seeds were recovered for 5 of the 7 lines (Table 1) and these lines were analysed further. The relative hydroxynitrile glucoside degrading activity in leaves of the *BGD4* mutant lines was studied following disruption of the tissue by grinding in a buffer, and correlated with predicted effects on protein stability using a BGD4 structural model based on the crystal structure of the cyanogenic

β-glucosidase from white clover (Barrett *et al*., 1995). For a detailed description of this BGD4 structural model, the reader is referred to the section on comparative structural modelling.

Line JI7175 is a likely loss-of-function mutant of *BGD4*, a mutation of the splice donor site of intron 8 likely results in an incorrectly processed transcript and simultaneously introduces an in frame stop-codon at this position that would truncate the BGD4 protein at amino acid residue 306. The additional G238D mutation in this line is the same amino acid substitution that conferred a partial loss of BGD2 function to the *cyd2-7* allele (Table 1, Takos *et al.*, 2010). Line JI7175 showed a reduced degradation of rhodiocyanosides that was more pronounced than in *cyd2-2*, a *BGD2* loss-of-function mutant, but was not impaired in lotaustralin break-down (Figure 2a,b). Line JI10179 showed a similar reduction in rhodiocyanoside degradation (Figure 2c). The G272D mutation in line JI10179 introduces a negative charge in the hydrophobic core of BGD4 and results in a steric clash with the Y262 and W203 residues, and is therefore likely to destabilize the enzyme. These two mutant lines, JI7175 and JI10179, provide genetic proof that BGD4 has a dominant role in rhodiocyanoside catabolism.

Of the remaining *BGD4* mutant lines, a reduction in rhodiocyanoside degradation was also observed for line JI8554. The G197 residue mutated in this line is located at the C-terminus of α-helix three, and in the structural model of BGD4 the G197E mutation leads to major steric clashes with F193, W203 and Q270, and is likely to be destabilizing. Lines JI8251 and JI7270 contained amino acid substitutions that were not expected to affect BGD4 enzymatic activity and both showed efficient degradation of all hydroxynitrile glucosides (Table 1, Figure 2c). The BGD4 structural model showed that the D184 residue is surface exposed and the D184N mutation in line JI8251 is

unlikely to have a significant effect on protein stability. The A189 residue on the same α-helix is buried and the A189T mutation in line JI7270 results in a slightly more bulky and polar side chain, but this change is relatively subtle.

To establish that BGD2 and BGD4 were jointly and exclusively responsible for the degradation of hydroxynitrile glucosides in leaves of *L. japonicus*, we crossed lines *cyd2-2* and JI7175 and selected for a double mutant in the F2 population. Following leaf disruption, no detectable hydrolysis of either lotaustralin or rhodiocyanoside A was observed in the double mutant line (Figure 2c), confirming the proposed physiological role of both enzymes.

A single amino acid is the main determinant of the substrate specificity difference between BGD2 and BGD4

Previous structural analyses of GH1 β-glucosidases show a cleft leading to an active site pocket (Barrett *et al.* 1995; Czjzek *et al.* 2000; Verdoucq *et al.*, 2004; Sue *et al.*, 2006; Barleben *et al.*, 2007). Variable residues in the upper part of the active-site form the aglycone binding region, and highly conserved residues make up the glycone-binding site in the bottom part. Two highly conserved peptide motifs T(L/F)NEP and (I/V)TENG, containing the catalytic glutamic acids, are located at the bottom of the active-site pocket. We previously generated a structural model of BGD2 based on the crystal structure of TrCBG, the cyanogenic β-glucosidase of white clover (Takos *et al.*, 2010). This model aided the present identification of the amino acids that could potentially explain the difference in substrate specificity between BGD2 and BGD4. Due to the high sequence identity of BGD2 and BGD4 (Figure S2), there are only a small number of amino acid differences associated with the active site. Several of these

are located on the α-helix immediately following the conserved TLNEP motif containing the acid/base catalyst residue E208 at the end of the fourth β-strand (Figure 3a). Between BGD2 and BGD4, these differences are G/V211, F/Y212, S/T213 and L/V219, with the first amino acid indicating the residue present in BGD2. In the model, the glycine at position 211 in BGD2, a valine in BGD4, is located in the active site cleft at the interface between the glycone and aglycone binding region at approximately 8 Å from the catalytic nucleophile, making it a likely candidate to affect enzyme specificity. Other residues of interest are L/V219 and T/D391 as these occupy positions previously implicated in the specificity of amygdalin hydrolase for its diglucoside substrate (Zhou *et al.*, 2002).

To assess experimentally the contribution these amino acids made to substrate specificity we mutagenized the coding regions of both BGD2 and BGD4 to reflect their counterpart. Single or multiple amino acid changes were introduced and the wild-type BGD2 and BGD4 and the mutated versions were transiently expressed in *N. benthamiana* under control of the CaMV 35S promoter. Tobacco leaf-discs excised from transfected areas were homogenized in MES buffer and supplied with various cyanogenic glucosides. Cyanogenesis was visualized using semi-quantitative HCN-sensitive Feigl-Anger paper (Figure 3b). In a separate experiment, the degradation of lotaustralin and rhodiocyanoside A was analysed by LC-MS (Figure 3c). The non-endogenous cyanogenic glucoside prunasin (Figure 1) was used as a control to demonstrate β-glucosidase activity in this transient expression system as all *L. japonicus* enzyme versions accepted this substrate. As expected BGD2 efficiently hydrolysed the endogenous cyanogenic glucosides lotaustralin and linamarin, but no hydrogen cyanide

was released when these substrates were combined with BGD4, nor was any BGD4-mediated lotaustralin degradation observed in the LC-MS experiment after 60 min. The essential role of the G211 residue in BGD2 was demonstrated by its mutation to a valine, which resulted in the loss of detectable HCN production (BGD2-G211V, Figure 3b). Furthermore, no lotaustralin degradation was observed by LC-MS after a 60 min. incubation period, while prunasin and rhodiocyanoside A degradation were near completion (Figure 3c). Conversely, mutating V211 in BGD4 to a glycine did result in detectable hydrogen cyanide production from both endogenous cyanogenic glucosides and in substantial lotaustralin degradation in the LC-MS experiment (BGD4-V211G, Figure 3b,c). The BGD4 mutant version containing G211/F212/S213 (BGD4-GFS) and the one containing G211/F212/S213/L219 (BGD4-GFS-L) consistently showed higher HCN releasing activity with the Feigl-Anger paper than the BGD4-V211G construct. The LC-MS analysis (Figure 3c) also suggested that in contrast with what was observed for BGD2-G211V, the BGD4-V211G mutant version was less able to maintain its activity on prunasin. The single BGD4 mutants containing either V219L or D391T amino acid substitutions, showed no detectable HCN releasing activity, indicating that they played no major role (Figure 3b).

Comparative structural modelling highlights the significance of the G/V211 residue

Previously we modelled the BGD2 protein structure to the crystal structure of the cyanogenic β-glucosidase TrCBG from white clover (Takos *et al.* 2010). To expand on this model we additionally considered the recently published crystal structure of the rice β-glucosidase Os4BGlu12 in complex with 2-deoxy-2-fluoroglucose (G2F) covalently

bound to its catalytic nucleophile E393 (Sansenya *et al.*, 2011). This served to identify the conserved amino acids that bind the glucosyl group in the lower part of the active site. We also modelled the protein structure of BGD4 using both templates and compared it with the structural models of BGD2 (Figure 4, Figure S3). The position and orientation of G2F in both models is shown in Figure 4 as it indicates the side of the active site pocket that interacts with the aglycone moiety.

High sequence identity between the *L. japonicus* BGDs and both the TrCBG and Os4BGlu12 templates (67-70%) resulted in very reliable models as reflected by high ProQ server scores (see Experimental procedures), and by excellent overlays of the main chains and of the catalytic and substrate binding residues at the glycone part of the active site cleft (Figure S3). The models clearly showed that the change from G to V at position 211, narrows the active site directly through its relatively bulky side chain (Figure 4). An *in silico* analysis of the effects of a G to V substitution performed on the white clover template showed in addition that the introduced V clashes with the catalytic acid/base. Such an effect of the G211V mutation is also conceivable in BGD2 as the amide nitrogen of residue 211 is at hydrogen bonding distance from the main chain carbonyl of the catalytic acid/base E208 (Figure 4a). Similarly, *in silico* mutation of G211 in the model of BGD2 to a V results in steric clashes with both W163 and F222, suggesting that the G211V change could further narrow the active site indirectly by pushing the side chains of W163 and F222 outwards into the active site pocket above the bound glucosyl. Such a movement is observed in the Os4BGlu12 template as this enzyme possesses a threonine at position 211. The F212 and S213 residues in BGD2 are substituted by the slightly larger Y212 and T213 in BGD4, and these changes are

unlikely to have a major impact but may also affect the flexibility or spaciousness of the active site.

In conclusion, the more restricted specificity of BGD4 seems to be associated with a narrower binding cleft (Figure 4b,c). The more spacious pocket present in BGD2 allows the accommodation of the bulky methyl-ethyl moiety of lotaustralin at the aglycone subsite. The longer but flat aglycone moieties of the rhodiocyanosides, and of the non-endogenous prunasin, would not be affected to the same extent by a narrower binding cleft, and the reduced flexibility in the presence of V211 could even offer better steric complementarity and stabilize productive binding of these substrates.

The occurrence of rhodiocyanosides and BGD4 in the *Lotus* genus

The occurrence of non-cyanogenic γ- and β-hydroxynitrile glucosides in *L. japonicus* was first reported by Forslund *et al.* (2004). To establish if rhodiocyanosides occurred more widely in the genus *Lotus*, we analysed 25 publically available accessions, representing 16 additional *Lotus* species, for their content of hydroxynitrile glucosides (Table 2, Figure S4). Adaptive polymorphisms for cyanogenesis are well-known to occur in natural populations of legumes, and acyanogenic individuals may result from a loss of the biosynthetic pathway, the β-glucosidase activity, or both (Olsen *et al*. 2013). Variation for the presence of rhodiocyanosides also exists in natural accessions of *L. japonicus* (Bjarnholt *et al*., 2008). Given this natural variation in biosynthetic ability and catabolism, we also tested if the *Lotus* accessions were cyanogenic and able to hydrolyse rhodiocyanosides. For accessions not containing endogenous cyanogenic glucosides or rhodiocyanosides, the ability of leaf extracts to hydrolyse exogenously added lotaustralin or rhodiocyanoside A was determined. In addition two sets of primer

pairs amplifying the catalytic region of the *BGD2* and *BGD4* genes were used to determine if the accessions expressed functional transcripts of these genes in leaves. The primer sets were designed against nucleotide sequences conserved in both genes and the PCR products obtained were sequenced to determine which *BGD* genes were expressed. Both primer sets amplified parts of the same transcript in each species for which a product was obtained. Relatively subtle variations in codon usage and predicted amino acid sequence were observed between the fragments amplified from the different *Lotus* species and accessions (Figure S5).

The two species most closely related to *L. japonicus* are *L. krylovii* and *L. burttii*, and these also originate from Asia (Degtjareva *et al.*, 2006; Degtjareva *et al.*, 2008). Both species contained cyanogenic glucosides, but rhodiocyanosides were the dominant compounds present (Figure S4). *L. krylovii* leaves were poorly cyanogenic, but the addition of exogenous lotaustralin to leaf homogenates resulted in efficient HCN production. Only a *BGD2*-type transcript, showing 100% identity at the amino acid level to BGD2 from *L. japonicus*, could be amplified from this species. *L. burttii* was acyanogenic, even following the addition of exogenous lotaustralin. LC-MS analysis showed very poor hydrolysis of rhodiocyanosides in a degradation assay, with over 90% of rhodiocyanosides remaining after 60 min. No *BGD2* or *BGD4* transcripts could be amplified in *L. burttii*, but a transcript for a BGD with a S at position 211 could be amplified with difficulty. *L. filicaulis* resembled *L. japonicus* in its hydroxynitrile glucoside profile, but was acyanogenic and no *BGD2* transcript could be amplified. This possibly reflects a naturally occurring polymorphism for cyanogenesis in this species represented by a single accession. *L. filicaulis* did however express a *BGD4*-type sequence (98% amino acid identity) and was able to hydrolyse rhodiocyanosides,

demonstrating that *BGD4* is not unique to *L. japonicus*. *L. preslii* lacked hydroxynitrile glucosides and the ability to hydrolyse them. Only a low abundant *BGD4*-like transcript could be amplified from this species, but a point mutation in the splice acceptor site of intron 8 resulted in a non-functional transcript. Rhodiocyanosides were also present in one out of three accessions of *L. tenuis* (accession LOT44), all of which were cyanogenic and contained *BGD2*-like transcripts (typically 95% amino acid identity).

All the five species containing rhodiocyanosides are within the so-called *L. corniculatus* clade of the *Lotus* genus. The order in which the species are presented in Table 2 reflects the well-established phylogeny of the *Lotus* genus, with only *L. drepanocarpus* not previously assigned a position (Degtjareva *et al.*, 2006; Degtjareva *et al.*, 2008). The *L. corniculatus* clade is indicated by the boxed group of species. *L. corniculatus*, a tetraploid species used as forage legume, is itself not known to contain rhodiocyanosides (Forslund *et al.*, 2004), and their absence was confirmed in five additional *L. corniculatus* accessions. *L. conimbricensis* is the only *Lotus* species previously reported to contain rhodiocyanosides (Bjarnholt *et al.*, 2008), although they were not present in the accession used in this study. This species falls just outside the *L. corniculatus* clade. No rhodiocyanosides were present in the tested *Lotus* species not closely associated with the clade. Although species outside the *L. corniculatus* clade contained no rhodiocyanosides, all those that showed cyanogenesis were able to degrade rhodiocyanoside A when this substrate was exogenously supplied (Table 2). This hydrolysing ability correlated with the presence of BGD sequences with a G amino acid residue at the position corresponding to G211 in BGD2 (Figure S5). These BGDs typically showed about 80% amino acid identity with BGD2. This again suggests that the wider active site pocket required for activity on lotaustralin and linamarin enables it

to also accommodate rhodiocyanosides and that this additional functionality represents a promiscuous catalytic activity of the cyanogenic β-glucosidases in these species.

Recruitment and evolution of β-glucosidases in hydroxynitrile glucoside metabolism

Cyanogenic glucosides are a class of chemical defence compounds present in a wide variety of plant species. We previously revised the view of a single evolutionary origin of this defence pathway by demonstrating the independent recruitment of the biosynthetic enzymes for the pathway in three plant species: *L. japonicus*, *Manihot esculenta* (cassava), and *Sorghum bicolor* (Takos *et al.*, 2011). Using the available genome sequences, we now examined the phylogenetic relationship of the cyanogenic β-glucosidases in these three species. This analysis, based on predicted amino acid sequences, showed that the BGD2 and BGD4 paralogs most closely grouped with cassava and sorghum β-glucosidases that were unrelated to linamarase and dhurinase, the cyanogenic β-glucosidases from cassava (Hughes *et al*., 1992) and sorghum (Cicek and Esen, 1998) respectively (Figure 5a). The β-glucosidase function in cyanogenic glucoside catabolism was therefore most likely also independently recruited in these different plant lineages.

We further investigated the evolutionary relationship between BGD2 and BGD4, and the newly identified β-glucosidases from the various *Lotus* species described above. For this analysis between closely related enzymes, an alignment of the DNA sequences was chosen. For clarity, each *Lotus* species was only represented by a single accession as there is little or no sequence variation between accessions of the same species. This

DNA-based tree of the molecular relatedness of β-glucosidases showed that the *BGD2*-like sequences, all containing the crucial glycine codon, followed the established phylogeny of the *Lotus* genus (Figure 5b) (Degtjareva *et al*., 2006; Degtjareva *et al*., 2008). In addition to this known phylogeny, *L. drepanocarpus* was found to be most closely related to *L. ornithopodioides*. The observed close relation between the β-glucosidase genes from *L. tenuis* and *L. corniculatus* is consistent with the hypothesis that *L. tenuis* was one of the diploid parents of the allotetraploid *L. corniculatus* (Grant and Small, 1996). As is apparent from the low bootstrap values at the relevant branch points, the position of the branch containing the *BGD4*-type sequences is inconclusive with respect to timing their divergence from the *BGD2*-type sequences to either before or after *L. conimbricensis* and the *L. corniculatus* group diverged. Using the codeml program of the PAML package we determined that an evolutionary selection model with varying dN/dS, the ratio of nonsynonymous/synonymous substitution rates, for the different branches of the phylogenetic tree was significantly more likely than a model based on a uniform dN/dS ratio ($2\Delta\ln L = 89.2$, $p < 0.001$). For example, a history of positive selection was suggested for the branch leading to the *BGD4*-type sequences (dN/dS = 1.43), in contrast to purifying selection for the branch containing *BGD2* from *L. japonicus* (dN/dS = 0.23, Figure S6). Although the dN/dS ratio's for either branch were not found to be significantly different from a dN/dS ratio of 1.0, different modes of selection acting on the branches could have influenced the level of sequence divergence observed.

DISCUSSION

Assigning a physiologically relevant activity to enzymes such as β-glucosidases based on *in vitro* assays is often complicated by the observation of a broad substrate specificity or promiscuity that does not necessarily reflect the physiological role. Here we report on a reverse genetics approach that identified *L. japonicus* mutants in the *BGD4* gene, and confirmed the previously proposed physiological role of BGD4 in the bio-activation of rhodiocyanosides (Takos *et al.*, 2010). A double mutant line conclusively showed that BDG2 and BGD4 are jointly responsible for hydroxynitrile glucoside catabolism in the leaves of *L. japonicus*. Structural homology modelling and experimental analysis identified the crucial amino acid explaining the difference in substrate specificity observed between BGD2 and BGD4, and suggested an evolutionary scenario for functional divergence based on the promiscuous activity of an ancestral cyanogenic β-glucosidase.

In the maize β-glucosidase ZmGlu1, which has DIMBOAGlc as its natural substrate, the aglycone only occupies the narrowest part of the aglycone binding pocket and is sandwiched between two hydrophobic walls (Czjzek *et al.*, 2000). The glycone binding site is located in the lower half of the active site slot, and contains highly conserved amino acids involved in hydrogen binding with the glycone moiety. The binding of the aglycone and glycone moieties is such that the glycosidic bond is correctly positioned with respect to the two catalytic glutamic acid residues at the bottom of the active site pocket (Czjzek *et al.*, 2000). In the structural models of the *Lotus* BGDs, the position of the glycone moiety is predicted based on its interactions with the highly conserved

amino acids, but modelling the binding of the aglycone moiety is more speculative. However, judging from the position of the covalently bound C1 atom of the glycone moiety, the aglycone subsite is most likely positioned near the G/V211 residue at the narrowest point of the binding pocket (Figure 4b and Figure 4c). The branching of the lotaustralin and linamarin aglycone moieties at their *O*-linked carbon atom (Figure 1) would require the wider pocket present in BGD2. In contrast, the aglycone moieties of rhodiocyanoside A and the even larger aromatic one of prunasin, are less branched at their *O*-linked carbon and are relatively flat molecules, which therefore can be accommodated in the narrower active site pocket of BGD4.

Directed laboratory enzyme evolution experiments suggest that new enzyme functions may arise by mutations that have little effect on the original enzyme function but substantial effects on enzyme promiscuity (Aharoni *et al.,* 2005). Following gene duplication, such promiscuous activities may be selected for with the multifunctional enzymes serving as progenitors for more specialized enzymes (Jensen, 1976; Khersonsky and Twafik, 2010). Due to the progressive accumulation of mutational changes between diverging enzymes, the mutational path that originally led to the distinct enzyme functions observed in nature is often difficult to reconstruct, although experimental strategies include the systematic exploration of the sequence and catalytic landscape separating divergent enzymes (O'Maille *et al.*, 2008), and the resurrection and functional characterization of ancestral proteins (Huang *et al.*, 2012). Alternatively, plausible mutational paths may be deduced when gene duplication and functional divergence are relatively recent evolutionary events, as is the case for the BGD2 and BGD4 β-glucosidases in *L. japonicus*.

A few examples of naturally occurring single amino acid differences that alter the specificity of closely related enzymes in plant specialized metabolism have been reported. A single amino acid difference between two *O*-methyltransferases from *Thalictrum tuberosum* altered their ability to methylate substrates such as the alkaloid norcoclaurine (Frick and Kutchan, 1999). In sorghum, the sesquiterpene synthases *SbTPS1*, *SbTPS2*, and *SbTPS3* are recently duplicated genes organized in a tandem repeat. Mutational studies showed that variation at single amino acids in the active site largely determined the relative proportions of the individual sesquiterpene volatiles produced from the common farnesyl diphosphate substrate (Zhuang *et al.*, 2012; Garms *et al.*, 2012). In basil (*Ocimum basilicum*), chavicol *O*-methyltransferase (CVOMT1) and eugenol *O*-methyltransferase (EOMT1) are 90% identical and catalyse the methylation of chavicol with equal efficiency, but EOMT1 was much more efficient in methylating eugenol (Gang *et al.*, 2002). Molecular modelling and site-directed mutagenesis showed that a single amino acid difference in the active site explained the difference in substrate preference. It was not clear however, whether CVOMT1 evolved from EOMT1, suggesting a scenario of enzyme specialization, or whether a new substrate specificity originated with EOMT1.

Also in the case of BGD2 and BGD4, two alternative scenarios could potentially explain their evolutionary relationship. In the first scenario, BGD2 acquired the ability to degrade lotaustralin and linamarin and the ancestral gene to both paralogs was not involved in cyanogenesis but able to hydrolyse rhodiocyanosides. However, the available data on the wider occurrence of cyanogenesis in legumes, the recently evolved ability to biosynthesize rhodiocyanosides within the *Lotus* genus, the restricted

occurrence of BGD4-type sequences to a few species in the *L. corniculatus* clade, as well as a more plausible mutational path for the functional divergence of the two enzymes suggested by the site-directed mutagenesis (Figure 3), all support the second scenario that BGD4 originated by substrate specialization from a promiscuous cyanogenic β-glucosidase.

The first line of support for this scenario of enzyme specialization is that the ability to synthesize rhodiocyanosides was most likely derived from the existing biosynthetic pathway for the cyanogenic glucoside lotaustralin. This is suggested by the wider occurrence of lotaustralin and linamarin as defence compounds in legumes, and the structure of the hydroxynitrile glucoside biosynthetic pathway in *L. japonicus*. The production of an amino acid derived oxime by CYP79D3 is the shared first enzymatic step in the synthesis of all hydroxynitrile glucoside compounds. The subsequent α-hydroxynitrile producing CYP736A2 enzyme is specific for the biosynthesis of cyanogenic glucosides, but biochemical and genetic evidence indicate the existence of a functionally related cytochrome P450 enzyme primarily involved in rhodiocyanoside biosynthesis but also contributing to cyanogenic glucoside production (Takos *et al.*, 2011; Saito *et al.*, 2012). With the evolutionary appearance of rhodiocyanosides, shown in this study to be associated with the *L. corniculatus* clade of the *Lotus* genus, selection for a more specialized β-glucosidase for this type of substrates became possible.

The molecular data on β-glucosidase substrate specificity provides additional support for the proposed evolutionary scenario based on the most plausible mutational path for enzyme divergence. The G211 amino acid residue in BGD2 allows it to accommodate substrates that are branched at the *O*-linked carbon atom of the aglycone moiety, such as linamarin and lotaustralin. A planar substrate, such as rhodiocyanoside A, also fits the

BGD2 active site and this latent ability may have supported the evolution of rhodiocyanoside biosynthesis since a bioactivating enzyme was already present. The proposed promiscuity of the ancestral enzyme is supported by the fact that cyanogenic *Lotus* species not containing rhodiocyanoside A are able to degrade this non-endogenous substrate. Following gene duplication of a cyanogenic β-glucosidase, a single nucleotide mutation that changed the glycine codon (GGT) to a valine codon (GTT) would have been sufficient to result instantly in enzyme specialization towards rhodiocyanosides. This is demonstrated by the BGD2-G211V mutant enzyme, which remains fully functional on prunasin and rhodiocyanosides but has lost the ability to hydrolyse lotaustralin and linamarin.

In general it can be suggested that the evolution of new glucosylated compounds would be favoured by the pre-existence of glycosylating and deglycosylating enzymes. Their presence would serve to detoxify and stabilize a new aglycone, and facilitate bio-activation of the glycosylated compound when it is beneficial. Such pre-existing enzymatic activities are likely provided by promiscuous glycosyltransferases and β-glucosidases, while more specialized enzymes may evolve as the novel pathway becomes established. The recent bifurcation of hydroxynitrile glucoside metabolism in the *Lotus* genus provides an interesting model system to study the molecular evolution of enzyme functions and the emergence of a new metabolic pathway in plant specialised metabolism.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

L. japonicus wild-type (accession MG-20) and mutant lines were germinated from seed on filter paper and grown on soil under a 16 h light cycle. *L. burttii* (Kawaguchi *et al.*, 2005) and *L. filicaulis* seeds (Sandal *et al.*, 2002) were kindly provided by Dr. Niels Sandal (University of Aarhus, Denmark). Species from the *Lotus* genus other than *L. japonicus* were kindly provided by the gene bank of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany, http://gbis.ipk-gatersleben.de/gbis_i/). These additional *Lotus* spp. were grown in a greenhouse at 22 °C. *Nicotiana benthamiana* were directly sown on soil and grown in a greenhouse at 22 °C.

Chemicals

The substrates lotaustralin and rhodiocyanoside A were purified by preparative HPLC (Bjarnholt *et al.*, 2008). Linamarin was purchased from AG Scientific (http://www.agscientific.com/). Amygdalin was purchased from Sigma-Aldrich (http://www.sigmaaldrich.com). Prunasin was chemically synthesized in our lab by Dr. M.S. Motawia.

TILLING and isolation of BGD4 mutants

TILLING for BGD4 was conducted by RevGen UK (http://revgenuk.jic.ac.uk/) at the John Innes Centre (UK) following the method of Le Signor *et al.*, 2009. Specific primers for the region covering both exon 7 and 8 of BGD4 were designed using

CODDLE (codons optimized to discover deleterious lesions; http://www.proweb.org/coddle) and the primer sequences are: forward, 5´-CGCTGATGAGGCAAGAATAA-3´, and reverse 5´-CTAAGGGATTGTCGGGAGTG-3´.

Analysis of hydroxynitrile glucoside content and degradation

Apical leaves of Lotus plants were assayed for their ability to hydrolyse hydroxynitrile glucosides by grinding leaves in 300 µl of 20 mM MES buffer (pH 6.0) and incubation at 30^{o}C. 25µl aliquots were collected at selected time points and further hydrolysis prevented by addition of 140 µl of methanol as described in Takos *et al*. 2010. Leaf discs from infiltrated tobacco leaves were assayed as described above, except that incubation mixtures were spiked with exogenous hydoxynitrile glucosides to a final concentration of 200 µM. Samples were boiled in a water bath for 2 min, and cooled on ice. To correct for volume losses during sample preparation, the methanol added to prevent further hydrolysis of the hydroxynitrile glucosides was spiked with amygdalin (200 µM as final concentration). The extracts were filtered and analytical LC-MS was carried out as described in Takos *et al.*, 2011. Semi-quantitative detection of cyanogenesis was by Fiegl-Anger paper as described in Takos *et al.*, 2010.

RNA extraction and cloning of BGDs

Total RNA was prepared from 100 mg of leaf tissue using the Spectrum Plant Total RNA kit (SIGMA), digestion of DNA was by RNase-Free DNase Set (Qiagen) using on-column DNase I digestion. For cDNA synthesis, 1–2 µg of total RNA was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen). To amplify the

BGD2 and *BGD4* cDNA fragments from the different *Lotus* spp. two primers sets were used. The sequences of the primer pairs used were Forward1: 5'-ATATCACCGCTACAAGGAAGATG-3' and Reverse1: 5'-ATGCCTGATATTTCTTCTTGTAC-3' giving a 518 bp PCR product, and Forward2: 5'-AATATTACAACAATCTCATCAAC-3' and Reverse2: 5'-AGTGTCTATAAGAGCTTCCTCTAG-3' giving a 902 bp fragment.

Site-directed mutagenesis

Mutations were introduced into BGD2 and BGD4 coding sequences following the QuikChange® Site-Directed Mutagenesis protocol (Stratagene). For each mutation, two complementary primers were synthesized containing the corresponding nucleotide changes (see Supplemental Table 1 online). Each primer pair was used in a PCR reaction with pDONOR207 containing *BGD2* or *BGD4* wild type sequences as template. Mutations introduced into the β-glucosidases coding sequences were confirmed by nucleotide sequencing.

Transient expression in leaves of *N. benthamiana* plants

Expression constructs were prepared by cloning wild-type and mutant forms of *BGD2* and *BGD4* by Gateway recombination into the pJAM1502 expression vector as described (Takos *et al.*, 2010). All constructs were transformed into *A. tumefaciens* (AGL1) by electroporation. Transient expression in leaves of 4-week old *N. benthamiana* plants was as previously described (Takos *et al.*, 2010), and the strains containing the expression constructs were co-infiltrated with a strain expressing the gene-silencing inhibitor protein 19 (Voinnet *et al.*, 2003). After 5 days, leaf discs (1 cm

diameter) were cut from infiltrated leaves of each plant and homogenized separately in 300 µl of 20 mM MES buffer, pH 6.5.

Structural modelling

The BGD2 and BGD4 structural models were generated using MODELLER (Sali *et al.*, 1995) available at the HH-pred server bioinformatics toolkit, and the crystal structures of the cyanogenic β-glucosidase from white clover (PDB code: 1CBG; Barrett *et al.*, 1995) and the rice β-glucosidase Os4BGlu12 (PDB code: 3PTM; Sansenya *et al.*, 2011) as templates. Peptides lacking the predicted N-terminal targeting sequences were submitted for modelling. Molecular rendering and structural alignments and measurements, and *in silico* mutations were performed using Pymol v1.0 software (http://www.pymol.org/) and Siwss Pdb viewer 3.7 (http://www.expasy.org/spdbv/, Guex and Peitsch, 1997). The model quality was checked with ProQ5 (http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi; Cristobal *et al.*, 2001), which resulted in LG and MaxSub scores of 5.524,0.544 and 5.368,0.539 for BGD2 and BGD4, respectively. This corresponds to very good to extremely good models for both proteins.

Phylogenetic analysis

The Molecular Evolutionary Genetics Analysis version 5 (MEGA5) software was used for the construction of sequence alignments and phylogenetic trees (Tamura *et al.* 2011). Cassava and sorghum β-glucosidase amino acid sequences related to BGD2 and BGD4 were identified by BLAST searches in the genome sequences available at www.phytozome.net. The partial β-glucosidase cDNA sequences from *Lotus spp.* were experimentally obtained as described above. Sequence alignments were constructed in

MEGA5 using ClustalW and manually adjusted, cDNA sequences were aligned as codons. Evolutionary trees were inferred using maximum likelihood analysis methods based on either the Jones-Taylor-Thornton model for amino acid substitutions, or the Tamura 3-parameter model for nucleotide substitutions. Bootstrap values were calculated using 1000 replications. Signatures of selection were assessed by estimating the dN/dS ratios using codeml in the PAML package (version 4.6; Yang, 2007). Different models and the statistical significance of dN/dS ratios for specific branches were tested by comparing log likelihood values.

SUPPORTING INFORMATION

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

Figure S1. Position of all nucleotide changes identified by TILLING of the *L. japonicus BGD4* gene.

Figure S2. Full amino-acid sequence alignment of BGD2, BGD4 and TrCBG.

Figure S3. Overlay of β-glucosidase structural models and the catalytic site residues.

Figure S4. Cyanogenesis and LC-MS analysis of hydroxynitrile glucoside content in the selected *Lotus* species.

Figure S5. Amino acid alignment of the BGDs amplified from *Lotus* spp.

Figure S6. Phylogenetic tree with dN/dS ratios.

Table S1. Sequences of primers used in this study.

ACKNOWLEDGEMENTS

We thank Bethany McCullagh (*RevGen*UK) for technical assistance with TILLING, and Mika Zagrobelny and Søren Bak for advice on phylogenetic analysis. The work was financially supported by the Danish National Research Foundation (Danmarks Grundforskningsfond), and the Danish Council for Independent Research| Natural Sciences. B.L.M. acknowledges financial support from the VILLUM Research Center 'Plant Plasticity' and from the Danish Council for Independent Research| Technology and Production Sciences. T.L.W. acknowledges support from the Biotechnology and Biological Sciences Research Council under grant number BB/I025891/1.

REFERENCES

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C. and Tawfik, D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37,** 73-76.

Barleben, L., Panjikar, S., Ruppert, M., Koepke, J. and Stöckigt, J. (2007) Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell*, **19,** 2886-2897.

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

Bjarnholt, N. and Møller, B.L. (2008) Hydroxynitrile glucosides. *Phytochemistry*, **69,** 1947-1961.

Bjarnholt, N., Rook, F., Motawia, M.S., Cornett, C., Jørgensen, C., Olsen, C.E., Jaroszewski, J.W., Bak, S. and Møller, B.L. (2008) Diversification of an ancient theme: hydroxynitrile glucosides. *Phytochemistry*, **69,** 1507-1516.

Butler, G.W. (1965) The distribution of the cyanoglucosides linamarin and lotaustralin in higher plants. *Phytochemistry*, **4,** 127-131.

Cicek, M. and Esen, A. (1998) Structure and expression of a dhurrinase (*β*-glucosidase) from sorghum. *Plant Physiol.* **116,** 1469-1478.

Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L. and Elofsson, A. (2001) A study of quality measures for protein threading models. *BMC Bioinformatics*, **2,** 5.

Czjzek, M., Cicek, M., Zamboni, V., Bevan, D.R., Henrissat, B. and Esen, A. (2000) The mechanism of substrate (aglycone) specificity in β-glucosidases is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA, -DIMBOAGlc, and –dhurrin complexes. *Proc. Natl. Acad. Sci. USA*, **97,** 13555-13560.

Degtjareva, G.V., Kramina, T.E., Sokoloff, D.D., Samigullin, T.H., Valiejo-Roman, C.M. and Antonov, A.S. (2006) Phylogeny of the genus *Lotus* (Leguminosae, Loteae): evidence from nrITS sequences and morphology. *Can. J. Bot.* **84,** 813-830.

Degtjareva, G.V., Kramina, T.E., Sokoloff, D.D., Samigullin, T.H., Sandral, G. and Valiejo-Roman, C.M. (2008) New data on nrITS phylogeny of *Lotus* (Leguminosae, Loteae). *Wulfenia*, **15,** 35-49.

Forslund, K., Morant, M., Jørgensen, B., Olsen, C.E., Asamizu, E., Sato, S., Tabata, S. and Bak, S. (2004) Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus*. *Plant Physiol.* **135,** 71-84.

Frick, S. and Kutchan, T.M. (1999) Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. *Plant J.* **17,** 329-339.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P. and Pichersky, E. (2002) Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. *Plant Cell*, **14,** 505-519.

Garms, S., Chen, F., Boland, W., Gershenzon, J. and Köllner, T.G. (2012) A single amino acid determines the site of deprotonation in the active center of sesquiterpene synthases SbTPS1 and SbTPS2 from *Sorghum bicolor*. *Phytochemistry*, **75,** 6-13.

Grant, W.F. and Small, E. (1996) The origin of the *Lotus corniculatus* (Fabaceae) complex: a synthesis of diverse evidence. *Can. J. Bot.* **74,** 975-989.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18,** 2714-2723.

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B. and Barkman, T.J. (2012) Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc. Natl. Acad. Sci. USA*, **109,** 2966-2971.

Hughes, M.A., Brown, K., Pancoro, A., Murray, B.S., Oxtoby, E. and Hughes, J. (1992) A molecular and biochemical analysis of the structure of the cyanogenic *β*-glucosidase (linamarase) from cassava (*Manihot esculenta* Cranz). *Arch. Biochem. Biophys.* **295,** 273-279.

Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30,** 409-425.

Kawaguchi, M., Pedrosa-Harand, A., Yano, K., Hayashi, M., Murooka, Y., Saito, K., Nagata, T., Namai, K., Nishida, H., Shibata, D., Sato, S., Tabata, S., Hayashi, M., Harada, K., Sandal, N., Stougaard, J., Bachmair, A. and Grant, W.F. (2005) *Lotus burtii* takes a position of the third corner in the *Lotus* molecular genetics triangle. *DNA Res.* **12,** 69-77.

Keresztessy, Z., Brown, K., Dunn, M.A. and Hughes, M.A. (2001) Identification of essential active-site residues in the cyanogenic β-glucosidase (linamarase) from cassava (*Manihot esculenta* Crantz) by site-directed mutagenesis. *Biochem. J.* **353,** 199-205.

Khersonsky, O. and Tawfik, D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79,** 471-505.

Le Signor, C., Savois, V., Aubert, G., Verdier, J., Nicolas, M., Pagny, G., Moussy, F., Sanchez, M., Baker, D., Clarke, J. and Thompson, R. (2009) Optimizing

TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnol. J.* **7,** 430-441.

McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiol*. **123,** 439-442.

Morant, A.V., Bjarnholt, N., Kragh, M.E., Kjærgaard, C.H., Jørgensen, K., Paquette, S.M., Piotrowski, M., Imberty, A., Olsen, C.E., Møller, B.L. and Bak, S. (2008) The β-glucosidases responsible for bioactivation of hydroxynitrile glucosides in *Lotus japonicus*. *Plant Physiol.* **147,** 1072-1091.

Nielsen, K.A., Olsen, C.E., Pontoppidan, K. and Møller, B.L. (2002) Leucine-derived cyano glucosides in barley. *Plant Physiol.* **129,** 1066-1075.

Ober, D. (2010) Gene duplications and the time thereafter – examples from plant secondary metabolism. *Plant Biology,* **12,** 570-577.

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.

O'Maille, P.E., Malone, A., Dellas, N., Hess, B.A., Smentek, L., Sheehan, I., Greenhagen, B.T., Chappell, J., Manning, G. and Noel, J.P. (2008) Quantitative

exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4,** 617-623.

Perry, J.A., Wang, T.L., Welham, T.J., Gardner, S., Pike, J.M., Yoshida, S. and Parniske, M. (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* **131,** 866-871.

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **5,** 439-445.

Saito, S., Motawia, M.S., Olsen, C.E., Møller, B.L. and Bak, S. (2012) Biosynthesis of rhodiocyanosides in *Lotus japonicus*: Rhodiocyanoside A is synthesized from (Z)-2-methylbutanaloxime via 2-methyl-2-butenenitrile. *Phytochemistry*, **77,** 260-267.

Sali, A., Potterton, L., Yuan, F., van Vlijmen H. and Karplus, M. (1995) Evaluation of comparative protein modelling by MODELLER. *Proteins*, **23,** 318-326.

Sandal, N., Krusell, L., Radutoiu, S., Olbryt, M., Pedrosa, A., Stracke, S., Sato, S., Kato, T., Tabata, S., Parniske, M., Bachmair, A., Ketelsen, T. and Stougaard, J. (2002) A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics*, **161,** 1673-1683.

Sansenya, S., Opassiri, R., Kuaprasert, B., Chen, C.-J. and Ketudat Cairns, J.R. (2011) The crystal structure of rice (*Oryza sativa* L.) Os4BGlu12, an oligosaccharide

and tuberonic acid glucoside-hydrolyzing β-glucosidase with significant thioglucohydrolase activity. *Arch. Biochem. Biophys.* **510,** 62-72.

Sue, M., Yamazaki, K., Yajima, S., Nomura, T., Matsukawa, T., Iwamura, H. and Miyamoto, T. (2006) Molecular and structural characterization of hexameric β-D-glucosidases in wheat and rye. *Plant Physiol.* **141,** 1237-1247.

Takos, A., Lai, D., Mikkelsen, L., Abou Hachem, M., Shelton, D., Motawia, M.S., Olsen, C.E., Wang, T.L., Martin, C. and Rook, F. (2010) Genetic screening identifies cyanogenesis-deficient mutants of *Lotus japonicus* and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. *Plant Cell,* **22,** 1605-1619.

Takos, A.M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., Motawia, M.S., Olsen, C.E., Sato, S., Tabata, S., Jørgensen, K., Møller, B.L. and Rook, F. (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant J.* **68,** 273-286.

Takos, A.M. and Rook, F. (2012) Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci.* **17,** 383-388.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731-2739.

Verdoucq, L., Czjzek, M., Moriniere, J., Bevan, D.R. and Esen, A. (2003) Mutational and structural analysis of aglycone specificity in maize and sorghum β-glucosidases. *J. Biol. Chem.* **278,** 25055-25062.

Verdoucq, L., Moriniere, J., Bevan, D.R., Esen, A., Vasella, A., Henrissat, B. and Czjzek, M. (2004) Structural determinants of substrate specificity in family 1 β-glucosidases. *J. Biol. Chem.* **279,** 31796-31803.

Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33,** 949-956.

Weng, J.-K., Philippe, R.N. and Noel, J.P. (2012) The rise of chemodiversity in plants. *Science*, **336,** 1667-1670.

Yang, Z. (2007) PAML4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586-1591.

Zhou, J., Hartmann, S., Shepherd, B.K. and Poulton, J.E. (2002) Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. *Plant Physiol.* **129,** 1252-1264.

Zhuang, X., Köllner, T.G., Zhao, N., Li, G., Jiang, Y., Zhu, L., Ma, J., Degenhardt, J. and Chen, F. (2012) Dynamic evolution of herbivore-induced sesquiterpene biosynthesis in sorghum and related grass crops. *Plant J*. **69,** 70-80.

Accession numbers

The sequences of the partial BGD cDNAs identified in this study are available from the GenBank/ EMBL data libraries under the following accession numbers: *L. drepanocarpus* (KF411421), *L. arenarius* (KF411420), *L. weilleri* (KF419278), *L. ornithopodioides* (KF411418), *L. edulis* (KF419282), *L. krylovii* (KF411422), *L. burttii* (KF411424), *L. tenuis* (LOT30 KF419279, LOT37 KF411423, LOT44 KF419280), *L. corniculatus* (LOT18 KF432203, LOT25 KF432205, LOT27 KF432204, LOT70 KF432206), *L. filicaulis* (KF411419), *L. preslii* (KF432202), *L. conimbricensis* (KF419281).

Table 1 Identity of the BGD4 mutant lines obtained by TILLING

Line	Wild type codon	Mutated codon	Amino acid position
JI7270	GCG (*Ala*, A)	ACG (*Thr*, T)	189
JI7175	GGT (*Gly*, G)	GAT (*Asp*, D)	238
	TGg*	TGA (Stop)	306
JI8554	GGA (*Gly*, G)	GAA (*Glu*, E)	197
JI8251	GAT (*Asp*, D)	AAT (*Asn*, N)	184
JI10179	GGT (*Gly*, G)	GAT (*Asp*, D)	272

The position of the mutated nucleotide is underlined. * exon/intron boundary.

Table 2 Occurrence of hydroxynitrile glucosides and corresponding β-glucosidase activities within the *Lotus* genus

Species[1]	Accession[2] Nr.	Country of origin	LC-MS[3] Lot.&Lin.	LC-MS[3] Rho.	Hydrol. Assay Cyanog[4].	Hydrol. Assay Rhodio[5].	BGD[6]
L. arenarius	LOT50	-	yes	no	yes	60%	BGD with G211
L. drepanocarpus	LOT53	France	yes	no	yes	80%	BGD with G211
L. ornithopodioides	LOT5	Greece	yes	no	yes	100%	BGD with G211
L. ornithopodioides	LOT6	Italy	yes	no	yes	80%	-
L. halophilus	LOT54	Italy	trace	no	no	0%	no product
L. weilleri	LOT16	Hungary	yes	no	yes	90%	BGD with G211
L. edulis	LOT45	France	yes	no	yes	75%	BGD with G211
L. japonicus	MG20	Japan	yes	yes	yes	100%	BGD2 + BGD4
L. krylovii	LOT15	China	yes	yes	yes	90%	BGD2
L. burttii	-	Pakistan	yes	yes	no	<10%	BGD with S211
L. tenuis	LOT30	USA	yes	no	yes	95%	BGD2
L. tenuis	LOT37	Hungary	yes	no	yes	95%	BGD2
L. tenuis	LOT44	Argentina	yes	yes	yes	100%	BGD2
L. corniculatus	LOT18	Czechoslov.*	yes	no	yes	90%	BGD2
L. corniculatus	LOT25	DDR*	yes	no	yes	80%	BGD2
L. corniculatus	LOT27	USSR*	yes	no	yes	90%	BGD2
L. corniculatus	LOT70	Canada	yes	no	yes	90%	BGD2
L. corniculatus	LOT72	Poland	yes	no	yes	90%	-
L. filicaulis	-	Algeria	yes	yes	no	100%	BGD4
L. preslii	LOT43	-	no	no	no	0%	non- functional BGD4
L. conimbricensis	LOT46	Italy	yes	no	yes	80%	BGD with G211
L. uliginosus	LOT34	France	no	no	no	0%	no product
L. uliginosus	LOT60	Spain	trace	no	no	0%	no product
L. pedunculatus	LOT29	Portugal	trace	no	no	<10%	no product
L. subbiflorus	LOT39	Portugal	no	no	no	0%	no product
L. subbiflorus	LOT41	France	no	no	no	0%	no product

1) Species are ordered according to their phylogenetic relationship with only *L. drepanocarpus* not previously assigned a position (Degtjareva *et al.*, 2006; Degtjareva *et al.* 2008). The *Lotus corniculatus* clade is boxed. 2) Accession numbers are the stock codes of the IPK unless otherwise indicated. 3) Presence of the cyanogenic glucosides lotaustralin and linamarin (Lot.&Lin.), or rhodiocyanosides (Rho.) determined by LC-MS. 4) The ability to hydrolyse endogenous cyanogenic glucosides or exogenously added lotaustralin determined by Feigl-Anger paper. 5) The ability to hydrolyse endogenous rhodiocyanosides or exogenously added rhodiocyanoside A determined by LC-MS. 6) The presence of the various BGD transcripts. The *BGD2*-type sequences within the *L. corniculatus* clade typically show about 98% nucleotide sequence identity with BGD2 from *L. japonicus*. The BGD sequences outside the *L. corniculatus* clade typically show between 85-90% nucleotide sequence identity with BGD2 from *L. japonicus* and are designated as BGD with G211. - not determined or not applicable. * Country of origin formerly known as.

Figure Legends:

Figure 1. Chemical structures of the hydroxynitrile glucosides used in this study.

In *L. japonicus*, the endogenous hydroxynitrile glucosides are the cyanogenic glucosides lotaustralin and linamarin, the γ-hydroxynitrile glucoside rhodiocyanoside A, and the β-hydroxynitrile glucoside rhodiocyanoside D. The cyanogenic glucoside prunasin is a non-endogenous substrate for BGD2 and BGD4.

Figure 2. BGD4 is involved in rhodiocyanoside A hydrolysis in *L. japonicus*.

Time-course of the hydrolysis of endogenous lotaustralin (a) and rhodiocyanoside A (b) following tissue disruption. Lines used are the parental wild-type line *L. japonicus* MG-20, the *cyd2-2* mutant defective in *BGD2*, and the JI7175 stop-codon mutant line in *BGD4*. (c) Relative levels of lotaustralin and rhodiocyanoside A remaining after 60 minutes in the lines indicated. Line *cyd2-2/JI7175* is the double mutant. Values are the mean of three biological replicates ± SE.

Figure 3: Amino acids responsible for substrate specificity of BGD2 and BGD4.

(a) Partial alignment of the BGD2 and BGD4 amino-acid sequences showing residues implicated in specificity. The conserved TLNEP motif with one of the two catalytic glutamic acids (E) is marked in yellow. Amino acids residues identified by modelling as candidates to confer specificity are marked in red, and blue.

(b) The ability of the various β-glucosidase enzyme mutants obtained by site-directed mutagenesis to hydrolyse the cyanogenic glucosides indicated following their transient

expression in *N. benthamiana*. HCN release following tissue disruption of leaf disks was visualized with Feigl-Anger paper.

(c) LC-MS analysis of the extent of prunasin, lotaustralin, and rhodiocyanoside A hydrolysis by the wild-type and mutant β-glucosidase enzymes indicated, following a 60 minute incubation with substrate. Enzymes were transiently expressed in *N. benthamiana*. Infiltration with only the vector containing the p19 gene served as control. Values are the mean of three biological replicates ± SE.

Figure 4. Comparative structural modelling of the active site pocket in BGD2 and BGD4.

(a) An overlay of amino acid residues in the protein structure homology models of BGD2 (cyan) and BGD4 (green) based on the white clover template. The structural differences around residue G/V211 at the aglycone-interacting side of the active site pocket are shown. Residue E208 is the catalytic acid/base close to G/V211, and the position of 2-deoxy-2-fluoro-α-D-glucopyranose is shown bound to the catalytic nucleophile E420 to visualize the location of the glycone binding site (based on the Os4BGlu12 model). The fluor atom is coloured yellow, and the C1 atom of the glucosyl moiety is labelled to indicate the relative position of the glycosidic bond.

(b) Semi-transparent surface representation showing the active site pocket of BGD2. The surface area contributed by G211 is indicated in red.

(c) Semi-transparent surface representation showing the narrower active site pocket of BGD4. The surface area contributed by V211 is indicated in red.

Figure 5. Evolutionary relationships between hydroxynitrile glucoside catabolising β-glucosidases.

(a) A maximum likelihood phylogenetic tree obtained as described in the experimental procedures. The tree is based on an amino acid alignment. The β-glucosidases most similar to BGD2 and BGD4 from *L. japonicus* were identified in the cassava and sorghum genome sequences. Gene codes correspond to those in the respective genome sequences. The β-glucosidases responsible for cyanogenesis in white clover (1CBG, *T. repens*), cassava (linamarase, *M. esculenta*), and sorghum (SbDhur1, *S. bicolor*) are included.

(b) A maximum likelihood phylogenetic tree obtained as described in the experimental procedures based on an alignment of the experimentally obtained partial cDNA sequences of β-glucosidases from the *Lotus* species indicated. LOT codes indicate the *Lotus spp.* accessions from which the cDNA sequence was obtained. The white clover sequence (1CBG, *T. repens*) is used as an outgroup. Branch lengths are measured in the number of substitutions per site, and bootstrap values were performed with 1000 replication and values are indicated at the branch points. Symbols in front of the species name indicate the type of β-glucosidase sequence isolated as indicated in Table 2: ● BGD2-type sequence, ○ BGD4-type sequence, ♦ BGD with G211.