Received Date : 08-Jul-2012

Revised Date : 30-Aug-2012

Accepted Date : 03-Sep-2012

Article type : Original Article

Novel seed coat lignins in the Cactaceae: structure, distribution and implications for the evolution of lignin diversity

Fang Chen,[1,3*] Yuki Tobimatsu,[2,] Lisa Jackson,[1] John Ralph,[2,4] and Richard A. Dixon[1,3*]

[1]Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA; [2]Department of Biochemistry, University of Wisconsin-Madison, Enzyme Institute, 1710 University Avenue, Madison, WI 53726, USA; [3]DOE Bioenergy Sciences Center, Oak Ridge, TN, USA; [4]DOE Great Lakes Bioenergy Research Center, Madison, WI, and Wisconsin Bioenergy Initiative, USA.

*For correspondence (fax +1 580 224 6692; e-mail radixon@noble.org or fchen@noble.org)

SUMMARY

We have recently described a hitherto unsuspected catechyl lignin polymer (C-lignin) in the seed coats of vanilla orchid and in cacti of one genus, *Melocactus* (Chen et al., PNAS 109: 1772-1777, 2012). We have now determined the lignin types in the seed coats of 130 different cactus species. Lignin in the vegetative tissues of cacti is of the normal guaiacyl/syringyl (G/S) type, but members of most genera within the subfamily Cactoidae possess seed coat lignin of the novel C-type only, which we show is a homopolymer formed by endwise β–O–4-coupling of caffeyl alcohol monomers onto the growing polymer, resulting in benzodioxane units. However, the species examined within the genera *Coryphantha, Cumarinia*, *Escobaria* and *Mammillaria* (Cactoideae) mostly had normal G/S lignin in their seeds, as did all six species in the subfamily Opuntioidae that were examined. Seed coat lignin composition is still evolving in the Cactaceae, as seeds of one *Mammillaria* species (*M. lasiacantha*)

possess only C-lignin, three *Escobaria* species (*E. dasyacantha*, *E. lloydii* and *E. zilziana*) contain an unusual lignin composed of 5-hydroxy-guaiacyl units, the first report of such a polymer occurring naturally in plants, and seeds of some species contain no lignin at all. We discuss the implications of these findings for the mechanisms underlying the biosynthesis of these newly discovered lignin types.

Keywords: Cactaceae, caffeyl alcohol, lignin composition, nuclear magnetic resonance spectroscopy, seed coat, taxonomy.

INTRODUCTION

The biosynthetic pathways leading to the formation of the cell wall polymer lignin have been extensively studied in recent years (Humphreys and Chapple, 2002; Boerjan *et al.*, 2003; Abramson *et al.*, 2010; Zhou *et al.*, 2010). The monolignols, a series of hydroxycinnamyl alcohols, are formed by a sequence of aromatic hydroxylation and *O*-methylation reactions (as well as successive side-chain reductions) to generate lignin precursors differing in their aromatic substitution patterns (Boerjan *et al.*, 2003). Natural lignins are generally composed of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, that are biosynthesized by polymerization of the three primary monolignols, *p*-coumaryl, coniferyl, and sinapyl alcohols, respectively (Figure 1a); natural angiosperm lignins have only low levels (<~2%) of H-units. Catechyl (C) and 5-hydroxyguaiacyl (5H) units that may derive from polymerization of the corresponding caffeyl and 5-hydroxyconiferyl alcohols (Figure 1a) are not found in 'normal' lignins. Until recently, these rare units were only found in the lignin of transgenic plants in which the first or second *O*-methylation reactions had been blocked (Marita *et al.*, 1999; Ralph *et al.*, 2001; Wagner *et al.*, 2011). In the case of catechyl units, the only report concerned the appearance of a small proportion of C-units in the lignin of transgenic gymnosperm (radiata pine) tracheary element cultures, probably as part of a co-polymer with dominant G units (Wagner *et al.*, 2011).

We recently made the unexpected observation that the seed coats of some monocots (*Vanilla planifolia*) and dicots (cacti of the genus *Melocactus*) contain large amounts of a catechyl homopolymer (Chen *et al.*, 2012). Nuclear magnetic resonance spectroscopy revealed that the C-lignin polymer from *V. planifolia* is derived from caffeyl alcohol monomers linked head to tail into benzodioxane chains via the 'endwise' radical coupling reactions that typify lignification (Chen *et al.*, 2012) (Figure 1b). The structure of the lignin from cacti has yet to be described.

Two obvious questions arise when considering a novel lignin polymer. What is its distribution in the plant kingdom, and how is it made *in planta*? To date, we have only found this polymer in *Vanilla*

species and members of the Cactaceae, although it is more than likely that it occurs elsewhere. Understanding its distribution may shed light on the mechanisms underlying its biosynthesis. Theoretically, the presumed caffeyl alcohol precursor of C-lignin would be formed if all routes to the *O*-methylated monolignols were blocked. The fact that C-lignin had not been discovered earlier is likely due to the fact that blocking lignin methylation is problematic for the plant if allowed to occur in the vascular tissue of the stem and leaves. For example, double knock-out or knock-down of the two monolignol *O*-methyltransferases (caffeic acid *O*-methyltransferase and caffeoyl CoA-3-*O*-methyltransferase) in Arabidopsis or Medicago leads to a dwarf phenotype with seriously distorted vascular architecture (Do *et al.*, 2007; Zhou *et al.*, 2010) but, surprisingly, without measurable formation of C-lignin. For whatever reason, lack of monolignol methylation appears to be better tolerated in the seed coat.

The Cactaceae represent a large but relatively unexplored plant family as regards their biochemistry. However, there has been considerable interest in the taxonomy of this family (Butterworth *et al.*, 2002; Nyffeler, 2002; Butterworth and Wallace, 2004; Edwards *et al.*, 2005), largely driven by the commercial availability of a large number of these attractive species for amateur growers and collectors, and this makes them an excellent subject for the study of seed coat lignin diversity and biosynthesis. Here, we determine the distribution of lignin composition types across 130 species of cactus, covering the two sub-families representative of the largest number of known species, and report the structures of the non-classical lignin polymers found. The results indicate that possession of C-lignin in the seed coat is a trait that is still evolving within the Cactaceae. We also show that some cactus species possess a novel natural lignin composed primarily of 5H units, which could be viewed as an intermediate stage in the evolution of seed coats with C-lignin.

RESULTS

Distribution of C-lignin within the family Cactaceae

We have recently described the presence of C-lignin in the seed coat of *V. planifolia* and some members of the genus the genus *Melocactus* (Chen *et al.*, 2012). To determine whether C-lignin is found throughout the family Cactaceae, we used thioacidolysis to determine the lignin composition from seed samples of 130 different species of cactus, representative of the two major sub-families (based on number of species) within the Cactaceae (Cactoideae and Opuntioideae), and including 47 different

genera. Table S1 presents the complete data set for the analysis of the seed lignins of the species analyzed, including overall lignin thioacidolysis yields and notes on seed color. Representative thioacidolysis traces are shown in Figure 2. Most species contained C-lignin as the only lignin type within the seed, whereas a smaller number contained G/S lignin (with no evidence for the presence of C-lignin). In a few species, the lignin levels appeared very low, and it was hard to assign a composition by thioacidolysis.

Table 1 places the lignin composition data for each genus studied in a phylogenetic context based on subfamilies, tribes and genera. From this analysis, it is clear that C-lignin is primarily found within the subfamily Cactoidae; we found no evidence for its presence within the subfamily Opuntioideae (in the genera A*ustrocylindropuntia*, *Grusonia*, *Maihueniopsis*, *Nopalea* or *Opuntia*). However, within the Cactoideae, members of the related genera *Coryphantha* and *Mammillaria* appear to contain predominantly G/S lignin in their seeds, and seeds of members of the genus *Escobaria* (also related to *Mammillaria* and *Coryphantha*) (Butterworth and Wallace, 2004) contain either G/S lignin or very low levels of lignin as assessed by thioacidolysis; the analyses suggested traces of both S and 5-hydroxy-guaiacyl (5H) units in the lignin (see below). These data are summarized in Figure 3, which places the compositional analysis on the broad taxonomy of the orders within the Cactaceae and on a detailed, previously published phylogeny of multiple genera within the Cactoideae based on plastid DNA sequence analysis (Crozier, 2005). Interestingly, although we found only G/S lignin (or very low lignin levels) in 15 *Mammillaria* species, the seeds of one species, *M. lasiacantha*, contained high levels of only C-lignin (Figure 2c). Thus, possession of C-lignin in the seed coat appears to be a trait that is still evolving in the genus *Mammillaria*.

C-lignin from cacti is a linear benzodioxane polymer

C-lignin in the seed coat of *V. planifolia* (Chen *et al.*, 2012) is a linear polymer derived almost totally from caffeyl alcohol monomers linked head to tail into benzodioxane chains via the ‘endwise’ radical coupling reactions that typify lignification (Figure 1b), and with a number average degree of polymerization in excess of 30 (Chen *et al.*, 2012). Analysis of lignin monomer composition by thioacidolysis indicated that the lignin in the seed coat of the Brazilian cactus *Melocactus salvadorensis* was likewise composed only of caffeyl alcohol-derived units (Figure 2a). Analysis of cell wall preparations from the stems, spines, spine base, stem/root interface, and roots of mature plants of Melocactus by thioacidolysis indicated, however, that the lignin in the vegetative tissues of the plant was comprised predominantly of G and S

units, with small amounts of H units and no evidence for the presence of C units (Figure S1).The main body of the plant contained very little lignin, most being found at the base of the stem and in the roots.

To determine whether *Melocactus* C-lignin is structurally similar to that found in *V. planifolia*, the entire cell walls from *M. salvadorensis* seed were analyzed by solution-state NMR via complete dissolution/acetylation using the DMSO/NMI solvent system (Lu and Ralph, 2003). The 2D HSQC spectra of the acetylated *M. salvadorensis* seeds (Figure 4a) displayed essentially identical signals to those previously observed for the *V. planifolia* C-lignin (Chen *et al.*, 2012). The aromatic HSQC spectrum clearly indicated that the seed lignin is almost exclusively composed of C units; typical G and S lignin aromatics are virtually nonexistent in this seed. In the high-field aliphatic regions of the HSQC spectrum, the signals from the benzodioxane units **VI** were clearly observed along with the signals from typical cell wall polysaccharides such as glucans, arabinans and xylans; the α-, β-, and γ-correlations from *trans*-benzodioxane rings **VI***t*, as well as lower-level contributions from *cis*-benzodioxane rings **VI***c*, were resolved. The benzodioxanes were the dominant lignin units accounting for over 98% of the identifiable dimeric units in the polymer. The normal acyclic β-aryl ether units **I**, which are the predominant linkage units in typical G/S lignins derived from coniferyl and sinapyl alcohols, were absent, whereas small amounts of resinol **III** units, which also can be generated by polymerization of caffeyl alcohol, are present. All these data suggest that the *Melocactus* seed lignin is essentially a homopolymer biosynthesized purely from caffeyl alcohol, as we previously determined for the Vanilla seed coat lignin.

Structures of other lignin types within the Cactaceae

The structure of the "classical" G/S lignin in seeds of *Mammillaria densispina* (Figure 2b) was interrogated by whole cell wall NMR analysis (Figure 4b); the lignin is typical G-rich G/S lignin and is seen as being rich in β-aryl ether units **I**, with more modest amounts of phenylcoumaran **II** and resinol **III** units, as is typical for most angiosperm lignins (Ralph and Landucci, 2010). The high level of C-lignin in the seeds of *M. lasiacantha* (Figure 2c) was confirmed by whole cell wall NMR analysis (Figure S2a).

The genus *Epithelantha* is phylogenetically distinct from *Mammillaria*, *Escobaria* and *Coryphantha* within the Cactoideae (Crozier, 2005) (Figure 3). Thioacidolysis of seeds of *Epithelantha micromeris* and *E. greggii* failed to release any lignin units (Figure 2d). Whole cell wall NMR confirmed the lack of lignin in these seeds, although an unassigned peak was detected in the aromatic region of the spectra, and the expected signals were seen for cell wall glucan, xylan and arabinan polymers (Figure 4c). Lignin levels were likewise extremely low to absent in *Mammillaria prolifera*, a species that, on the

basis of chloroplast DNA sequence, is very closely related to *M. lasiacantha* (Crozier, 2005), the only *Mammillaria* species we observed to possess C-lignin in seeds (Figure 3).

Two species of *Escobaria* (*albicolumnaria* and *vivipara* [subspecies *arizonica* and *neomexicana*]) examined contain high levels of G/S lignin their seeds (Table S1, Figure 2e). In contrast, the thioacidolysis yields from *E. dasyacantha*, *lloydii* and *zilziana* were very low and the GC traces were atypical of those from seed coats containing G/S or C-lignins. The GC traces for these *Escobaria* seeds displayed small amounts of G, S and unusual 5H (5-hydroxyguaiacyl) units (Figure 2f). Although poorly resolved, the S and 5H thioacidolyis products could be clearly identified by selected ion monitoring (Figure 2f, inset). The results suggest that the unusual monolignol, 5-hydroxyconiferyl alcohol, participates in lignification in these *Escobaria* seeds. Whole cell wall NMR strongly supports this finding. In the aromatic region of HSQC spectra of seed of *E. dasyacanthus* and *zilziana* (Figure 4d,e), clear signals from etherified and non-etherified 5H units are resolved and could be assigned tentatively using the NMR data from lignins incorporating 5-hydroxyconiferyl alcohol isolated from transgenic plants (Marita *et al.*, 1999, 2001, 2003; Ralph *et al.*, 2001). There are several aromatic correlations (gray) that are currently not assigned. The aliphatic regions of the spectra from those seeds clearly indicate the massive presence of benzodioxane units IV that could be generated from the polymerization of 5-hydroxyconiferyl alcohol in a similar way as those from caffeyl alcohol (Figure 1b). Whereas thioacidolysis released only small amounts of 5H monomers along with comparable amounts of G and/or S monomers (Figure 2f), whole cell wall NMR suggests that the lignin from seeds of *E. dasyacantha* is composed almost totally of 5H units (Figure 4d), and that from *E. zilziana* is also predominantly composed of 5H units, with low levels of S units and minimal levels of G units (and no C units) (Figure 4e).

Morphology of cactus seed coats with different lignin compositions

To determine whether possession of different lignin types was associated with any specific morphological feature in the cactus seed coat, we examined the coats from imbibed seeds of four species, *Melocactus salvadorensis* and *Mammillaria lasiacantha* (C-lignin only), *Escobaria lloydii* (5H and S lignin only), and *Escobaria vivipara arizonica* (G/S lignin only), by fluorescence and light microscopy (Figure 5). Lignin/phenolic autofluorescence was seen in confocal images of all seed coats, and was greatest in *E. vivpara arizonica* and *M. lasiacantha* (Figure 5c,d). The level of autofluoresence did not directly reflect the lignin thioacidolysis yield from the seed (Table S1). Staining with toluidine blue revealed deep blue staining around purple structures (corresponding to pits) in the seed coats of the related *Escobaria* and *Mammillaria* species, whereas the seed coat of the *Melocactus* species showed

much weaker staining and lack of the purple pit-like structures. These data indicate that the seed coat structure/staining is related to taxonomy and is probably little affected by lignin composition.

Discussion

Distribution of C-lignin within the Cactaceae

Our data indicate that C-lignin is widely distributed within the subfamily Cactoideae, but is apparently not found (or at least is not common) within the more ancestral subfamily Opuntioideae. We did not analyze lignin within the subfamilies Maihuenioideae (comprising only one species) or Pereskioideae; Pereskia is regarded as being basal to all other families within the Cactaceae (Edwards *et al.*, 2005) (Figure 3), and we assume that it will contain "classical" G/S lignin in seeds.

Within the Cactoideae, all members of fifty six genera analyzed contained only C-lignin within their seeds. These species cover an enormous geographical and climatic range, from Argentina (*Acanthocalycium*) through the tropics (*Melocactus*, *Discocactus*) to cold, mountainous regions of North America (*Pediocactus*). Thus, the possession of C-lignin does not appear to be a trait that has obviously been selected for on the basis of environment.

Our data indicate that the seed lignin in the ancestral clades within the Cactaceae is of the normal G/S type typical of dicotyledonous angiosperms, and that this has evolved to C-lignin in most of the genera within the subfamily Cactoideae, with the exception of *Coryphantha*, *Escobaria* and *Mammillaria*, where several species (possibly all in the case of C*oryphantha*) retain the G/S type. Of the sixteen species of *Mammillaria* examined, four had high levels of G/S lignin in the seeds, eleven had very low levels of G/S lignin, and one, *M. lasiacantha*, had high levels of C-lignin only. All seven Escobaria species analyzed had high levels of lignin; G/S in three species and an unexpected polymer showing neither C- nor G-signatures in the other four species. Thus, it is clear that seed lignin composition is still evolving in *Escobaria* and *Mammillaria*.

Identification of a novel, naturally occurring 5H lignin

NMR analysis clearly indicated that the lignin from *Escobaria dasyacantha* seeds consists almost entirely of 5H units. As in C-lignin, these are characterized by chains of benzodioxane units (from β–O–4-coupling followed by internal trapping of the quinone methide intermediate); this explains the apparently very

low lignin content of these samples when determined by thioacidolysis, as the linkage pattern of 5H residues makes them resistant to release by this approach (Marita *et al*., 2003). The *Escobaria* seed coat lignins represent, to the best of our knowledge, the first examples of naturally occurring all-5H lignins. Lignin polymers containing 5-hydroxyconiferyl alcohol-derived benzodioxane units have been created artificially though knock-out of 5-hydroxyconiferaldehyde 5-*O*-methyltransferase (COMT), either through mutation or RNA silencing, in vascular tissues of several plant species (Marita *et al.*, 1999, 2001, 2003; Ralph *et al.*, 2001). Arabidopsis transgenics with particularly high levels of benzodioxanes were created by up-regulation of F5H in COMT-deficient backgrounds (Vanholme *et al*., 2010; Weng *et al.*, 2010) but a lignin composed entirely of 5H units has never been reported. The lignin in seeds of *E. zilziana* was also composed primarily of 5H units, although in this case there were also S units and (minimal) G units present. Whether the lignin in *E. zilziana* seeds is of mixed type, or is composed of separate 5H and S/(G) polymers, remains to be determined. Isolation and more detailed characterization of *Escobaria* 5H lignins will be reported elsewhere.

Biochemical mechanisms underlying seed lignin types within the Cactaceae

Based on the currently accepted biochemical pathway to monolignols, we suggest the following hypothesis regarding the evolution of lignin types within the Cactaceae. The ancestral form is likely the common G/S type. 5H lignin differs from S lignin in the lack of a single methyl group on the hydroxyl at position 5 of the aromatic ring (Figure 1a), whereas C-lignin differs from G lignin in the lack of methyl substitution on the hydroxyl at position 3. The fact that the lignin in the vegetative tissues of plants that possess C-lignin in the seeds appears to be of the normal G/S type ((Chen *et al.*, 2012) and the present work) suggests that if a genetic mutation has occurred to block monolignol methylation at either position, it is likely seed-specific in nature. In this respect, it is interesting that, although knock out of COMT can result in relatively normal plants with very high levels of 5H lignins in vascular tissue (Marita *et al.*, 2001, 2003; Morreel *et al.*, 2004), knock out of both COMT and caffeyl Coenzyme-A 3-*O*-methyltransferase (CCoAOMT, the enzyme primarily responsible for the 3-*O*-methylation reaction in monolignol biosynthesis (Ye *et al.*, 1994)) does not result in plants producing C-lignin in vascular tissues. Rather, plant growth is seriously disrupted, and very low levels of normal G-lignins are produced in the vegetative tissues (Do *et al.*, 2007; Zhou *et al.*, 2010). Since the Arabidopsis COMT/CCoAOMT double knock-out mutant does not set seeds, it is not possible to investigate whether the seed coat of the mutant contains C-lignin (Do *et al.*, 2007). In only one report to date have C-units been reported in the

lignin of non-seed tissues, in liquid cultures of *Pinus radiata* down-regulated in CCoAOMT (Wagner *et al.*, 2011). Even so, this lignin is still predominantly of the G type (Wagner *et al.*, 2011). It is therefore possible that C-lignin formation in vascular tissue is detrimental to plant growth (possibly because of the chemical reactivity of the caffeyl alcohol monomer), but is somehow tolerated in the seed coat.

In the light of these considerations, we propose that seed-coat-specific loss of expression of COMT (either through mutation of a seed-specific form of the *COMT* gene or of a regulatory gene controlling COMT expression in the seed coat) will result in accumulation of the 5H lignins found in some species of *Escobaria*. Similarly, formation of C-lignins would require mutation of a seed-specific form of CCoAOMT or of a regulatory gene controlling CCoAOMT expression in the seed. Alternatively, it is possible to account for C-lignin formation through the loss of function or loss of expression of an enzyme of C1 metabolism that produces methyl groups specific for lignification. Although there have recently been major advances in our understanding of the transcriptional control of monolignol biosynthesis (Zhong and Ye, 2007; Zhong *et al.*, 2008; Demura and Ye, 2010; Zhao and Dixon, 2011), little is known about seed-specific regulation of the pathway. It is also possible that C-lignin formation might require loss of function of both COMT and CCoAOMT, because recent evidence has implicated COMT as also able to catalyze the 3-*O*-methylation of monolignol precursors (Parvathi *et al.*, 2001; Zhou *et al.*, 2010). In this latter case, the 5H lignin seen in some *Escobaria* species could be viewed as intermediate in the evolution of C lignin. It is important to note that we have, to date, seen no example of a cactus species that contains both C and G/S lignins in the seed coat, consistent with C-lignin formation's being the result of loss of function of a process that is also required for G/S lignin biosynthesis.

Although we favor the above models, which are supported by our data within the Cactaceae, our limited understanding of seed coat lignin composition across the plant kingdom makes it difficult to rule out the possibility that C and 5H seed coat lignins are ancestral, and that the methylation functions were more recently acquired but can also be easily lost.

The taxonomy of the Cactaceae is complex and constantly under revision (Butterworth *et al.*, 2002; Butterworth and Wallace, 2004; Crozier, 2005). Although we do not propose that lignin composition is a reliable taxonomic marker, it is interesting that this trait does in general correlate with the evolution of the family, and appears to be still evolving in at least two related genera (*Escobaria* and *Mammillaria*). Major questions for future study include whether seed lignin compositional traits provide any fitness advantages to cacti, and the molecular genetic mechanisms that result in tissue-specific silencing (or possibly avoidance) of monolignol *O*-methylation. Lignin deposition in plant seed coats is a

relatively unexplored area, but clearly one with potential for uncovering new regulatory mechanisms controlling lignin biosynthesis, with implications for advancing the field of bio-based products.

EXPERIMENTAL PROCEDURES

Plant Materials

Seeds of the following species were obtained from the corresponding author's (RAD) private collection (Sulphur, Oklahoma): *Astrophytum capricorne*, *Astrophytum myriostigma* var *columnare*, *Discocactus silicicola*, *Frailea heliosa*, *Mammillaria pilcayensis*, *Mammillaria spinossissima*, *Mammallaria glassii* var *ascencionis* L1186, *Mammillaria prolifera* var *texensis*, *Melocactus glaucescens*, *Melocactus guitartii*, *Melocactus obtusipetalus*, *Melocactus erythranthus*, *Melocactus ammotrophus*, *Melocactus neomontanus*, *Melocactus delessertianus*, *Melocactus curvispinus*, *Melocactus cesius*, *Melocactus levitestatus*, *Melocactus guaricensis*, *Melocactus salvadorensis*, *Melocactus loboguerrori*, *Melocactus depressus*, *Notocacus roseoluteus*, *Opuntia ficus- indica*, *Uebelmannia pectinifera* var *pseudopectinifera*, and *Wigginsia erinacea*. Seeds of the remaining species were purchased from Mesa Garden, Belen, NM; catalog numbers are listed in Table S1. Root, stem, stem base, spine and spine base tissues were dissected from a juvenile plant of *Melocactus salvadorensis*.

Plant vegetative tissue samples were ground to a fine powder using a freezer mill (SPEX SamplePrep, Metuchen, NJ) under liquid nitrogen. Seeds were broken down to powder by hand with a hammer. All samples were extracted with chloroform/methanol (2:1, v/v), 100% methanol, and water (3 times each), then freeze-dried before lignin analysis.

Light Microscopy

Seeds were planted on Arabidopsis growth medium (0.5% phyta-agar) consisting of 0.5 X Murashige-Skoog (MS) salts and 1% sucrose. After planting, petri dishes containing seeds were placed in the refrigerator for 3 days for vernalization before exposure to constant light at 24^{o}C. After seeds were germinated (up to 1 month), they were fixed in 2.5% (v/v) glutaraldehyde (Electron Microscopy Sciences, Hatfield, PA) and 4% (v/v) paraformaldehyde (Electron Microscopy Sciences) in phosphate buffered saline (PBS, pH 7.2). Samples were then washed with PBS, post fixed with 1% (v/v) osmium tetroxide, dehydrated in a graded ethanol series, and embedded in LR White resin (London Resin Co. Ltd., Reading, Berkshire, UK). The resin was polymerized at 55°C for 3 days. Serial 1 µ sections were cut with a diamond knife on a Leica EM UC7 ultramicrotome (Leica Mikrosysteme GmbH, Vienna, Austria),

and stained with 1% (w/v) toluidine blue O and observed with an Nikon Optophot-2 microscope (Nikon Instruments Inc., Tokyo, Japan). For visualization of total lignin autofluorescence, 1 μ sections were immersed in Citifluor (Electron Microscopy Science) and observed under a Leica TCS SP2 AOBS confocal laser scanning microscope (Leica Microsystems Heidelberg GmbH, Mannheim, Germany) illuminated with a 405 nm blue diode laser and emission detected at 460 nm. Laser intensity, pinhole and photomultiplier gain settings were kept constant between specimens.

Thioacidolysis

Analytical thioacidolysis was according to the method described previously (Lapierre *et al.*, 1985, 1986). Briefly, approximately 10 mg of freeze-dried samples were reacted with 3 ml of 0.2 M BF_3-etherate in an 8.75:1 dioxane/ethanethiol mixture at 100 ° C for 4 h. At the end of the reaction, 4 ml of water was added, and the mixture extracted three times with 4 ml of dichloromethane. The organic layers were combined and dried under a stream of nitrogen. After derivatization with BSTFA + 1% TMCS (Pierce Biotechnology Inc., Rockford, IL), the samples were subjected to gas chromatography/mass spectrometric (GC/MS) analysis. Lignin thioacidolysis monomers were quantified by GC/MS as their trimethylsilyl derivatives. GC/MS was performed on a Hewlett–Packard 5890 series II gas chromatograph with a 5971 series mass-selective detector (column: HP-1; 60 m × 0.25 mm; 0.25 μm film thickness), and mass spectra were recorded in electron impact mode (70 eV). Scanning range was from 50-650 m/z.

NMR Analysis

Selected seed samples were analyzed by solution-state NMR via complete dissolution/acetylation using the dimethylsulfoxide (DMSO)/*N*-methylimidazole (NMI) solvent system as described previously (Lu and Ralph, 2003; Mansfield *et al*., 2012). Briefly, pre-extracted seeds (~100 mg) were ball-milled (3 × 20 min, 10-min cooling cycle) using a Retsch PM100 ball-mill vibrating at 600 rpm with ZrO_2 vessels containing ZrO_2 ball bearings. The ball-milled seed tissue (~40 mg) was dissolved in DMSO/NMI (2:1, vol/vol, 3 ml) at room temperature, acetic anhydride (1 ml) was added, and then stirred for 2 h at room temperature. The mixture was poured into distilled water (1 l). The resultant precipitate was recovered by filtration, washed with ultrapure water (1 l) and then lyophilized to yield acetylated seed tissues (yield typically at 115-130%), which were totally dissolved in chloroform-*d* (600 μl) for NMR experiments. The NMR spectra were acquired on a Bruker Biospin AVANCE 500 MHz spectrometer fitted with a cryogenically cooled 5-mm TCI

gradient probe with inverse geometry (proton coils closest to the sample) and spectral processing used Bruker's Topspin 3.1 (Mac) software. The central chloroform peaks were used as internal reference (δ_C/δ_H: 77.0/7.26 ppm). Adiabatic 2D-HSQC ("hsqcetgpsisp2.2") experiments were carried out using the parameters described previously (Chen et al., 2012). Processing used typical matched Gaussian apodization in F2 (LB = −0.5, GB = 0.001), and squared cosine-bell and one level of linear prediction (32 coefficients) in F1. For quantification of aromatic distributions (Figure 2), the carbon-2 correlation from G, the carbon-2/6 correlation from S, the carbon-2 correlation from C, and the carbon-2 and carbon-6 correlations from etherified and non-etherified 5H units were volume-integrated, and the S integrals were logically halved; linear prediction was not used for spectra being integrated. For an estimation of the various interunit linkage types, the well-resolved side chain C_α–H_α contours (**I_α**, **II_α**, **III_α**) or, because of spectral overlap of the **IV_α** contours, the C_β–H_β (**IV_β**) (Fig. 2) were integrated; no correction factors were used (Mansfield *et al*., 2012).

Phylogenetic Analysis

The phylogeny of the Cactaceae has been under constant revision (Nyffeler and Eggli, 2010). The overall taxonomy used in this article is based on that of the International Cactaceae Systematics Group (ICSG) of the International Organization for Succulent Plant Study (http://en.wikipedia.org/wiki/Classification_of_the_Cactaceae), and a higher resolution phylogeny based on plastid DNA sequence analysis (Crozier, 2005) was used to infer evolutionary aspects of lignin composition within the genera *Mammillaria*, *Coryphantha* and *Escobaria*.

ACKNOWLEDGEMENTS

We thank Dr Jin Nakashima for assistance with light microscopy, and Drs Xiaoqiang Wang and Qiao Zhao for critical reading of the manuscript. This work was supported by the Samuel Roberts Noble Foundation, and the US Department of Energy's Bioenergy Sciences and Great Lakes Bioenergy Centers, supported by the Office of Biological and Environmental Research in the DOE Office of Science (BER DE-AC05-00OR22725 and DE-FC02-07ER64494, respectively). YT acknowledges Postdoctoral Fellowship support from the Japan Society for the Promotion of Science (JSPS).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Figure S1. Lignin content and composition of different organs of a Melocactus plant, as determined by thioacidolysis.

Figure S2. Short-range ^{13}C–^{1}H correlation (HSQC) spectra of acetylated whole seed cell walls from *Mammillaria lasiacantha* and *Escobaria vivipara*.

Table S1. Details of origin, seed color and lignin content/composition of all species analyzed in the present study.

REFERENCES

Abramson, M., Shoseyov, O. and Shani, Z. (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability. *Plant Sci.* **178**, 61-72.

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annu Rev Plant Biol* **54**, 519-546.

Butterworth, C.A., Cota-Sanchez, J.H. and Wallace, R.S. (2002) Molecular systematics of tribe Cacteae (Cactaceae: Cactoideae): A phylogeny based on rpl16 intron sequence variation. *Systemat. Bot.* **27**, 257-270.

Butterworth, C.A. and Wallace, R.S. (2004) Phylogenetic studies of *Mammillaria* (Cactaceae)-insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *Am. J. Bot.* **91**, 1086-1098.

Chen, F., Tobimatsu, Y., Havkin-Frenkel, D., Dixon, R.A. and Ralph, J. (2012) A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. USA.* **109**, 1772-1777

Crozier, B.S. (2005) Systematics of Cactaceae Juss.: Phylogeny, cpdna evolution, and classification, with emphasis on the genus Mammillaria Haw., PhD Thesis, University of Texas, Austin.

Demura, T and Ye, Z.H. (2010) Regulation of plant biomass production. *Curr Opin Plant Biol* **13**, 299-304.

Do, C.-T., Pollet, B., Thevenin, J., Sibout, R., Denoue, D., Barriere, Y., Lapierre, C. and Jouanin, L. (2007) Both caffeoyl Coenzyme A 3-*O*-methyltransferase 1 and caffeic acid 3-*O*-methyltransferase 1 are

involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta* **226**, 1117-1129.

Edwards, E.J., Nyffeler, R. and Donoghue, M.J. (2005) Basal cactus phylogeny: Implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *Am. J. Bot.* **92**, 1177-1188.

Humphreys, J.M. and Chapple, C. (2002) Rewriting the lignin roadmap. *Curr. Opin. Plant Biol.* **5**, 224-229.

Lapierre, C., Monties, B. and Rolando, C. (1985) Thioacidolysis of lignin: Comparison with acidolysis. *J. Wood Chem. Technol.* **5**, 277-292.

Lapierre, C., Monties, B. and Rolando, C. (1986) Thioacidolysis of poplar lignins: Identification of monomeric syringyl products and characterization of guaiacyl-syringyl lignin fractions. *Holzforschung* **40**, 113-118.

Lu, F. and Ralph, J. (2003) Non-degradative dissolution and acetylation of ball-milled plant cell walls: high-resolution solution-state NMR. *Plant J.* **35**, 535-544.

Mansfield, S.D., Kim, H., Lu, F. and Ralph, J. (2012) Whole plant cell wall characterization using 2D-NMR. *Nature Protocols* **7**, 1579-1589.

Marita, J.M., Ralph, J., Hatfield, R.D. and Chapple, C. (1999) NMR characterization of lignins in Arabidopsis altered in the activity of ferulate 5-hydroxylase. *Proc. Natl. Acad. Sci. USA* **96**, 12382-12332.

Marita, J.M., Ralph, J., Hatfield, R.D., Guo, D., Chen, F. and Dixon, R.A. (2003) Structural and compositional modifications in lignin of transgenic alfalfa down-regulated in caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase. *Phytochemistry* **62**, 53-65.

Marita, J.M., Ralph, J., Lapierre, C., Jouanin, L. and Boerjan, W. (2001) NMR characterization of lignins from transgenic poplars with suppressed caffeic acid *O*-methyltransferase activity. *J. Chem. Soc. Perkin Trans.* **1**(22), 2939-2945.

Morreel, K., Ralph, J., Lu, F., Goeminne, G., Busson, R., Herdewijn, P., Goeman, J.L., Van der Eycken, J., Boerjan, W. and Messens, E. (2004) Phenolic profiling of caffeic acid *O*-methyltransferase-deficient poplar reveals novel benzodioxane oligolignols. *Plant Physiol.* **136**, 4023-4036.

Nyffeler, R. (2002) Phylogenetic realtionships in the cactus family (Cactaceae) based on evidence from *trnk/matk* and *trnl-trnf* sequences. *Am. J. Bot.* **89**, 312-326.

Nyffeler, R and Eggli, U. (2010) A farewell to dated ideas and concepts: Molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. *Schumannia* **6**, 109-149.

Parvathi, K., Chen, F., Guo, D., Blount, J.W. and Dixon, R.A. (2001) Substrate preferences of *O*-methyltransferases in alfalfa suggest new pathways for 3-*O*-methylation of monolignols. *Plant J.* **25**, 193-202.

Ralph, J., Lapierre, C., Lu, F.C., Marita, .JM., Pilate, G., VanDoorsselaere, J., Boerjan, W. and Jouanin, L. (2001) NMR evidence for benzodioxane structures resulting from incorporation of 5-hydroxyconiferyl alcohol into lignins of *O*-methyltransferase-deficient poplars. *J. Agric. Food Chem.* **49**, 86-91.

Vanholme, R., Ralph, J., Akiyama, T., Lu, F., Rencoret Pazo, J., Christensen, J., Rohde, A., Morreel, K., DeRycke, R., Kim, H., Van Reusel, B and Boerjan, W. (2010) Engineering traditional monolignols out of lignins by concomitant up-regulation *F5H1* and down-regulation of *COMT* in Arabidopsis. *Plant J.*, 2010, **64**, 885-897.

Wagner A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., Pears, L,.. and Ralph, J. (2011) CCoAOMT suppression modifies lignin composition in *Pinus radiata*. *Plant J.* **67**, 119-129.

Weng, J.-K., Mo, H. and Chapple, C. (2010) Over-expression of F5H in COMT-deficient Arabidopsis leads to enrichment of an unusual lignin and disruption of pollen wall formation. *Plant J.* **64**, 898-911.

Ye, Z.H., Kneusel, R.E., Matern, U. and Varner, J.E. (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *Plant Cell* **6**, 1427-1439.

Zhao, Q. and Dixon, R.A. (2011) Transcriptional networks for lignin biosynthesis: More complex than we thought? *Trends Plant Sci.* **16,** 227-233

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L. and Ye, Z.-H. (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell* **20**, 2763-2782.

Zhong, R. and Ye, Z.H. (2007) Regulation of cell wall biosynthesis. *Curr. Opin. Plant Biol.* **10**, 564-572.

Zhou, R., Nakashima, J., Jackson, L., Shadle, G., Temple, S., Chen, F. and Dixon, R.A. (2010) Distinct cinnamoyl CoA reductases involved in parallel routes to lignin in *Medicago truncatula. Proc. Natl. Acad. Sci. USA* **107**, 17803-17808.

Legends to Figures and Table

Figure 1. Lignin biosynthesis. (a) Simplified scheme for the successive hydroxylation and *O*-methylation reactions in monolignol biosynthesis, and the structures and nomenclatures of the monolignol units in the lignin chain. (b) Assembly of C-lignin and 5-H lignin polymers from caffeyl alcohol and 5-hydroxyconiferyl alcohol, respectively.

Figure 2. Gas chromatogram traces of the thioacidolysis products of lignin from seeds of the cactus species indicated. IS, docosane internal standard. The inserts in 2f show the partial selective-ion-monitoring (SIM) chromatogram of m/z 299 (S monomer) and m/z 357 (5H monomer).

Figure 3. Phylogenetic relationships and seed lignin composition of cacti. The scheme shows the five orders currently recognized (from the basal Pereskioidaeae to the Cactoideae), with a more detailed phylogeny, based on chloroplast DNA sequence data, shown for certain closely related species within the Cactoideae (this only represents a small fraction of the order). The latter is taken from Crozier (2005), with additional species not analyzed here deleted for simplification. Lignin composition is indicated by color: red, C-lignin; blue, G/S lignin; green, 5H lignin; gray, no detectable lignin. The boxes with black outlines show the number of individual species analyzed with the lignin composition indicated. The two numbered boxes with no outline indicate the lignin composition and number of species in addition to those listed within a specific region of the detailed phylogenetic tree. See Table S1 for full details of species analyzed.

Figure 4. Short-range ^{13}C–^{1}H correlation (HSQC) spectra of acetylated whole seed cell walls from *Melocactus salvadorensis* (a), *Mammillaria densispina* (b), *Epithelantha micromeris* (c), *Escobaria dasyacantha* (d), and *Escobaria zilziana* (e).

Figure 5. Cactus seed coat morphology revealed by fluorescence microscopy and toluidine blue O staining. (a-d), autofluorescence of lignin and phenolic compounds. (e-h), staining with toluidine blue O. (a) and (e), *Melocactus salvadorensis* (C lignin only). (b) and (f), *Escobaria lloydii* (5H and S lignin). (c) and (g), *E. vivipara arizonica* (G and S lignin). (d) and (h), *Mammillaria lasiacantha* (C lignin only).

Table 1. Seed lignin compositions within 47 genera of cacti, organized according to the taxonomy suggested by the International Cactaceae Systematics Group (ICSG) of the International Organization for Succulent Plant Study. X = no lignin detectable.

Subfamily	Tribe	Genus	# of species	Seed lignin type
Cactoideae	**Browningiae**	Browningia	1	C
	Cacteae	Ariocarpus	3	C
		Astrophytum	3	C
		Coryphantha	11	G/S
		Echinocactus	1	C
		Echinomastus	1	C
		Epithelantha	2	X
		Escobaria	7	G/S, 5H
		Ferocactus	2	C
		Mammillaria	16	G/S, X, C
		Pediocactus	1	C
		Sclerocactus	5	C
		Stenocactus	2	C
		Strombocactus	1	C
		Thelocactus	3	C
		Turbinicarpus	4	C
	Calymmanthae (0)			
	Cereeae	Cereus	1	C
		Discocactus	1	C
		Melocactus	15	C
		Pilosocereus	1	C
		Uebelmannia	1	C
	Hylocereeae		0	
	Notocactaceae	Copiapoa	2	C
		Eriosyce	3	C
		Eulychnia	1	C
		Frailea	1	C
		Parodia	4	C
	Pachycereeae	Carnegia	1	C
		Echinocereus	3	C

		Escontria	1	C
		Pachycereus	1	C
		Polaskia	1	C
		Stenocereus	1	C
	Rhipsalideae		0	
	Trichocereeae	Acanthocalycium	1	C
		Cleistocactus	3	C
		Echinopsis	4	C
		Espostoa	2	C
		Haageocereus	1	C
		Harrisia	2	C
		Matucana	1	C
		Mila	1	C
		Oreocereus	3	C
		Rebutia	4	C
		Weberbauerocereus	1	C
Maihuenioideae			0	
Opuntioideae	**Austrocylindropuntieae**	Austrocylindropuntia	1	G/S
	Cylindropuntieae	Grusonia	1	G/S
	Opuntieae	Opuntia	2	G/S
	Pterocacteae		0	
	Tephrocacteae	Maihueniopsis	1	G/S
Pereskioideae			0	

Table 1. Seed lignin compositions within 47 genera of cacti, organized according to the taxonomy suggested by the International Cactaceae Systematics Group (ICSG) of the International Organization for Succulent Plant Study. X = no lignin detectable.