Calcium Sensors and Their Interacting Protein Kinases: Genomics of the Arabidopsis and Rice CBL-CIPK Signaling Networks

Üner Kolukisaoglu², Stefan Weinl², Dragica Blazevic, Oliver Batistic, and Jörg Kudla*

Abteilung Pflanzenphysiologie, Universität Rostock, Albert-Einstein-Strasse 3, 18051 Rostock, Germany (U.K.); Molekulare Botanik, Universität Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany (S.W., D.B., O.B., J.K.); and Institut für Botanik, Universität Münster, Schlossgarten 3, 48149 Münster, Germany (J.K.)

Calcium signals mediate a multitude of plant responses to external stimuli and regulate a wide range of physiological processes. Calcium-binding proteins, like calcineurin B-like (CBL) proteins, represent important relays in plant calcium signaling. These proteins form a complex network with their target kinases being the CBL-interacting protein kinases (CIPKs). Here, we present a comparative genomics analysis of the full complement of CBLs and CIPKs in Arabidopsis and rice (Oryza sativa). We confirm the expression and transcript composition of the 10 CBLs and 25 CIPKs encoded in the Arabidopsis genome. Our identification of 10 CBLs and 30 CIPKs from rice indicates a similar complexity of this signaling network in both species. An analysis of the genomic evolution suggests that the extant number of gene family members largely results from segmental duplications. A phylogenetic comparison of protein sequences and intron positions indicates an early diversification of separate branches within both gene families. These branches may represent proteins with different functions. Protein interaction analyses and expression studies of closely related family members suggest that even recently duplicated representatives may fulfill different functions. This work provides a basis for a defined further functional dissection of this important plant-specific signaling system.

All organisms use a network of signal transduction pathways to cope with their environment, to control their metabolism, and to realize their developmental programs. Calcium has emerged as an ubiquitous second messenger involved in many of these processes (Gilroy and Trewavas, 2001). In plants, intracellular calcium levels are modulated in response to various signals including abiotic stresses, light, pathogens, and hormones (Harper, 2001; Knight and Knight, 2001). In addition, physiological processes like guard cell regulation, root hair elongation, and pollen tube growth are accompanied by distinct spatiotemporal changes in calcium concentration (Evans et al., 2001; Sanders et al., 2002). The specific signatures of these calcium transients can encode information and contribute to the specificity required for efficient stimulus response coupling (McAinsh and Hetherington, 1998; Allen and Schroeder, 2001). Nevertheless, an additional level of regulation in calcium signaling is achieved via the function of calcium-binding proteins (Snedden and Fromm, 1998; Luan et al., 2002; Sanders et al., 2002). These sensor proteins likely recognize specific calcium signatures and relay these signals into downstream responses such as phosphorylation cascades and regulation of gene expression (Luan et al., 2002).

In the case of plant calcium-dependent protein kinases (CDPKs), which harbor a calcium-binding domain as well as a catalytic Ser/Thr kinase domain, calcium signals can be sensed and transmitted by a single protein (Harmon et al., 2000; Sanders et al., 2002). Functional analyses of different plant CDPKs have provided growing evidence for crucial functions of these proteins in divergent processes like hormone and stress signaling as well as pathogen response (Sheen, 1996; Romeis et al., 2001). Most of the other known calcium sensors, for example, calmodulin (CaM) and CaM-like proteins, do not have an enzymatic activity on their own. Binding of calcium ions to these proteins changes their affinity for and results in a subsequent activation or deactivation of their target proteins (Luan et al., 2002). Recently, the group of plant calcium sensor proteins has been extended by the identification of calcineurin B-like (CBL) proteins from Arabidopsis. These proteins are most similar to both the regulatory B subunit of calcineurin (CNB) and neuronal calcium sensors (NCS) of animals (Kudla et al., 1999). CBL proteins contain EF hand motifs as structural basis for calcium binding and interact specifically with a group of Ser/Thr protein kinases designated as CBL-interacting protein kinases (CIPKs; Shi et al., 1999; Kim et al., 2000). CIPKs most likely represent targets of calcium signals sensed and transduced by CBL proteins. The N-terminal kinase domain of these pro-
teins is most closely related to sucrose non-
fermenting-like and cAMP-dependent kinases from
various organisms (Shi et al., 1999). Because of this
structural feature, these proteins have also been as-
signed to the SnRK3 subgroup of plant SNF-like ki-
nases (Hrabak et al., 2003). Subsequent yeast two-
hybrid screens and interaction studies uncovered 15
CIPKs from Arabidopsis. Moreover, this analysis
identified a conserved 24-amino acid motif within the
C-terminal non-kinase region of the CIPKs as suffi-
cient and required to mediate CBL-CIPK interaction
(Albrecht et al., 2001). Because of the functionally
important and conserved amino acid motif Asn-Ala-
Phe, this protein region has been designated as NAF
domain (protein families database accession no.
PF03822). The high number of identified CBLs and
CIPKs suggested a rather complex web of potential
interactions. Remarkably, preferential complex for-
mation of individual CBLs with defined subsets of
CIPKs has been found to contribute to generating
specificity in this signaling network (Kim et al., 2000;
Albrecht et al., 2001).

To date, a physiological function has been estab-
lished only for a few CBL and CIPK proteins. The
CBL calcium sensor SOS3 (AtCBL4) and the CIPK-
type kinase SOS2 (AtCIPK24) have been identified in
genetic screens. They appear to be part of a calcium-
regulated signaling pathway that specifically medi-
ates salt stress signaling and adaptation (Liu and
Zhu, 1998; Halfter et al., 2000; Liu et al., 2000). Anal-
ysis of a CIPK3 loss-of-function allele established a
function of this kinase in regulating abscisic acid
(ABA) responses during seed germination and in
regulating stress-induced gene expression (Kim et al.,
2003). Two independent reverse genetic analyses of
CBL1 function revealed that disruption of the CBL1
gene renders plants hypersensitive to drought and
salt stress and impairs the regulation of early in-
duced transcription factors as well as other stress-
responsible genes (Albrecht et al., 2003; Cheong et al.,
2003). Moreover, these studies also identified this
calcium sensor as a regulator of plant cold stress
signaling and adaptation.

The biochemical analyses of CIPKs revealed that
both deletion of the CBL-interaction domain and sub-
stitution of a critical Thr in the putative activation
loop by Asp result in a constitutively hyperactive
kinase (Guo et al., 2001; Gong et al., 2002b, 2002c).
Phenotypical analyses of plants overexpressing such
a constitutively active CIPK (PKS11/CIPK8) sug-
gested a function of the analyzed protein in mediat-
ing plant resistance to high concentrations of Glc
(Gong et al., 2002a). However, the regulation, func-
tion, and interconnections of most of the known CBL
and CIPK proteins have not been determined (Luan
et al., 2002). Moreover, the identification and analysis
of the components of this signaling network has so
far been mainly restricted to the species Arabidopsis.

In this study, we present a systematic genome-
wide analysis of the Arabidopsis and rice (Oryza
sativa) CBL and CIPK genes and encoded proteins.
Our data provide a framework for future functional
dissection of this calcium-regulated plant signal
transduction system in both model plants.

RESULTS AND DISCUSSION

The Complement of CBLs and CIPKs Encoded in the
Arabidopsis Genome

In our previous studies, we identified and ana-
lyzed six CBLs and 15 CIPKs from Arabidopsis
(Kudla et al., 1999; Shi et al., 1999; Albrecht et al.,
2001). To uncover the full complement of these cal-
cium sensors and their interacting kinases encoded in
the Arabidopsis genome, we performed database
analyses using different BLAST algorithms (see “Ma-
terials and Methods”). Subsequent pair-wise DNA
sequence comparisons were applied to identify and
exclude double annotations. To distinguish CIPKs
from other related SNF-like protein kinases, can-
date sequences were additionally analyzed with the
“motif” algorithm for the presence of the NAF do-
main (Albrecht et al., 2001). These analyses led to the
identification of 10 loci potentially encoding CBLs
and 25 loci that could encode CIPKs (Tables I and II).

We performed a systematic reverse transcriptase
(RT)-PCR-based cDNA cloning of all potentially func-
tional CBL and CIPK genes to verify their active
expression and to determine the structure of the en-
coded mRNAs. Our cDNA cloning approach re-
vealed that all genes are transcribed and in several
cases led to a correction of the theoretical gene pre-
dictions. These data have been incorporated in public
databases, and the corresponding nucleotide and
protein identification numbers are depicted in Tables
I and II.

Our comparison of the genomic loci with their
encoded cDNAs revealed that almost all identified
Arabidopsis CBL genes harbor six or seven introns in
their coding region (Fig. 1A). In the case of AtCBL10,
an additional intron appears in its N-terminal coding
region. This intron lacks any homolog in other
AtCBLs. Four introns are absolutely conserved in
phase and position in all 10 CBL genes. Three other
introns are structurally conserved, but are absent in
either one or two members of the CBL gene family,
most likely due to intron loss events during evolu-
tion. Interestingly, the AtCBL7 gene, which is located
in tandem orientation with AtCBL3 in the Arabidop-
sis genome (see below), contains an intron in a posi-
tion not found in any other CBL locus. Remarkably,
all CBL genes for which the full-length cDNA se-
quences have been determined (AtCBL1-4 and
AtCBL9) harbor introns in their 5′-untranslated region
(UTR); one in AtCBLs1 and 9, and two in AtCBLs2-4.
Whether this unusual intron composition potentially

Kolukisaoglu et al.
reflects regulatory functions during gene expression remains to be investigated experimentally.

In contrast to the CBL genes, only eight of 25 AtCIPK genes harbor multiple intron sequences (Fig. 1B). Again, phase and position of these introns, if present, appear to be rather conserved. The position of seven introns is absolutely invariant. The other six introns are only absent in one to three intron-containing members of the CIPK gene family. One exception is represented by an additional intron in the AtCIPK23 gene, which is not shared by any other CIPK gene. Moreover, the AtCIPK16 gene harbors only a single intron not found in other CIPK genes. This might indicate a recent intron acquisition in this

### Table I. The Arabidopsis CBL gene family

<table>
<thead>
<tr>
<th>Name</th>
<th>Arabidopsis Genome Initiative Identification No.</th>
<th>Nucleotide Identification No.</th>
<th>cDNA</th>
<th>Protein Identification No.</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCBL01</td>
<td>At4g17510</td>
<td>AF076251</td>
<td>642</td>
<td>AAC26008</td>
<td>213</td>
</tr>
<tr>
<td>AtCBL02</td>
<td>At5g55990</td>
<td>AF076252</td>
<td>681</td>
<td>AAC26009</td>
<td>226</td>
</tr>
<tr>
<td>AtCBL03</td>
<td>At4g26570</td>
<td>AF076253</td>
<td>681</td>
<td>AAC26010</td>
<td>226</td>
</tr>
<tr>
<td>AtCBL04</td>
<td>At5g24270</td>
<td>AF192886</td>
<td>669</td>
<td>AAG28402</td>
<td>222</td>
</tr>
<tr>
<td>AtCBL05</td>
<td>At4g01420</td>
<td>AF192885</td>
<td>612</td>
<td>AAG28401</td>
<td>203</td>
</tr>
<tr>
<td>AtCBL06</td>
<td>At4g16350</td>
<td>AF192884</td>
<td>681</td>
<td>AAG28400</td>
<td>226</td>
</tr>
<tr>
<td>AtCBL07</td>
<td>At4g26560</td>
<td>AF290434</td>
<td>745</td>
<td>AAG10059</td>
<td>214</td>
</tr>
<tr>
<td>AtCBL08</td>
<td>At1g64480</td>
<td>AF411957</td>
<td>645</td>
<td>AAL10300</td>
<td>214</td>
</tr>
<tr>
<td>AtCBL09</td>
<td>At5g47100</td>
<td>AF411958</td>
<td>642</td>
<td>AAL10301</td>
<td>213</td>
</tr>
<tr>
<td>AtCBL10</td>
<td>At4g33000</td>
<td>AF490607</td>
<td>741</td>
<td>AAT72364</td>
<td>246</td>
</tr>
</tbody>
</table>

* Systematic designation given to the genes by the Arabidopsis Genome Initiative (http://mips.gsf.de/proj/thal/db/index.html).  
* Length of the cDNA (coding region) in basepairs.  
* No. of amino acids of the predicted protein; for alternatively spliced mRNAs, the longest predicted amino acid sequence is given.  
* Alternative name, SOS2 (Liu et al., 1998).

### Table II. The Arabidopsis CIPK gene family

<table>
<thead>
<tr>
<th>Name</th>
<th>Arabidopsis Genome Initiative Identification No.</th>
<th>Nucleotide Identification No.</th>
<th>cDNA</th>
<th>Protein Identification No.</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCIPK01</td>
<td>At3g17510</td>
<td>AF302112</td>
<td>1,335</td>
<td>AAG28776</td>
<td>444</td>
</tr>
<tr>
<td>AtCIPK02</td>
<td>At5g07070</td>
<td>AF286050</td>
<td>1,371</td>
<td>AAF86506</td>
<td>456</td>
</tr>
<tr>
<td>AtCIPK03</td>
<td>At2g26980</td>
<td>AF286051</td>
<td>1,326</td>
<td>AAF86507</td>
<td>441</td>
</tr>
<tr>
<td>AtCIPK04</td>
<td>At4g14580</td>
<td>AY007221</td>
<td>1,281</td>
<td>AAG01367</td>
<td>426</td>
</tr>
<tr>
<td>AtCIPK05</td>
<td>At5g10930</td>
<td>AF285105</td>
<td>1,296</td>
<td>AAF86504</td>
<td>431</td>
</tr>
<tr>
<td>AtCIPK06</td>
<td>At4g30960</td>
<td>AF285106</td>
<td>1,326</td>
<td>AAF86505</td>
<td>441</td>
</tr>
<tr>
<td>AtCIPK07</td>
<td>At3g23000</td>
<td>AF290192</td>
<td>1,290</td>
<td>AAK16682</td>
<td>429</td>
</tr>
<tr>
<td>AtCIPK08</td>
<td>At4g24400</td>
<td>AF290193</td>
<td>1,338</td>
<td>AAK16683</td>
<td>445</td>
</tr>
<tr>
<td>AtCIPK09</td>
<td>At1g01140</td>
<td>AF295664</td>
<td>1,350</td>
<td>AAK16684</td>
<td>449</td>
</tr>
<tr>
<td>AtCIPK10</td>
<td>At5g58380</td>
<td>AF295665</td>
<td>1,440</td>
<td>AAK16685</td>
<td>479</td>
</tr>
<tr>
<td>AtCIPK11</td>
<td>At2g30360</td>
<td>AF295666</td>
<td>1,308</td>
<td>AAK16686</td>
<td>435</td>
</tr>
<tr>
<td>AtCIPK12</td>
<td>At4g18700</td>
<td>AF295667</td>
<td>1,470</td>
<td>AAK16687</td>
<td>489</td>
</tr>
<tr>
<td>AtCIPK13</td>
<td>At2g34180</td>
<td>AF295668</td>
<td>1,509</td>
<td>AAK16688</td>
<td>502</td>
</tr>
<tr>
<td>AtCIPK14</td>
<td>At5g01820</td>
<td>AF295669</td>
<td>1,329</td>
<td>AAK16689</td>
<td>442</td>
</tr>
<tr>
<td>AtCIPK15</td>
<td>At5g01810</td>
<td>AF302111</td>
<td>1,266</td>
<td>AAK16692</td>
<td>421</td>
</tr>
<tr>
<td>AtCIPK16</td>
<td>At2g25090</td>
<td>AY030304</td>
<td>1,410</td>
<td>AAF19215</td>
<td>469</td>
</tr>
<tr>
<td>AtCIPK17</td>
<td>At1g48260</td>
<td>AY036958</td>
<td>1,299</td>
<td>AAK64513</td>
<td>432</td>
</tr>
<tr>
<td>AtCIPK18</td>
<td>At1g29230</td>
<td>AY034099</td>
<td>1,563</td>
<td>AAK59695</td>
<td>520</td>
</tr>
<tr>
<td>AtCIPK19</td>
<td>At5g45810</td>
<td>AY030303</td>
<td>1,452</td>
<td>AAK50347</td>
<td>483</td>
</tr>
<tr>
<td>AtCIPK20</td>
<td>At5g45820</td>
<td>AY035225</td>
<td>1,320</td>
<td>AAK61493</td>
<td>439</td>
</tr>
<tr>
<td>AtCIPK21</td>
<td>At5g57630</td>
<td>AY034100</td>
<td>1,251</td>
<td>AAK59696</td>
<td>416</td>
</tr>
<tr>
<td>AtCIPK22</td>
<td>At2g38490</td>
<td>AF450478</td>
<td>1,338</td>
<td>AAL47845</td>
<td>445</td>
</tr>
<tr>
<td>AtCIPK23</td>
<td>At1g30270</td>
<td>AY035226</td>
<td>1,449</td>
<td>AAK61494</td>
<td>482</td>
</tr>
<tr>
<td>AtCIPK24</td>
<td>At5g35410</td>
<td>AF395081</td>
<td>1,341</td>
<td>AAK72257</td>
<td>446</td>
</tr>
<tr>
<td>AtCIPK25</td>
<td>At5g25110</td>
<td>AF448226</td>
<td>1,467</td>
<td>AAL41008</td>
<td>488</td>
</tr>
</tbody>
</table>

* Systematic designation given to the genes by the Arabidopsis Genome Initiative (http://mips.gsf.de/proj/thal/db/index.html).  
* Length of the cDNA (coding region) in basepairs.  
* No. of amino acids of the predicted protein; for alternatively spliced mRNAs, the longest predicted amino acid sequence is given.  
* Alternative name, ATPK10 (Mizoguchi et al., 1994).  
* Alternative name, ATPK10 (Mizoguchi et al., 1994).  
* Alternative name, ATPK10 (Mizoguchi et al., 1994).  
* Alternative name, ATPK10 (Mizoguchi et al., 1994).  
* Alternative name, SOS2 (Falke et al., 2000).
In addition, we identified a single intron in the 5′-UTR of AtCIPK15, which otherwise does not harbor intron sequences in its coding region. Taken together, these findings suggest the preservation of a relatively constant exon-intron composition of the CBL and CIPK genes during the evolution of the Arabidopsis genome.

In addition, we assayed the Arabidopsis CBLs and CIPKs for the occurrence of alternative mRNA splicing. To this end, the cDNA sequences obtained during this study were compared with all available full-length cDNA and expressed sequence tag (EST) sequence data. These investigations identified alternatively spliced mRNAs for AtCBL2, AtCBL4, and AtCBL10 as well as for AtCIPK3 and AtCIPK9. All alternatively spliced cDNA sequences have been deposited in the respective GenBank entries. The nucleotide identification numbers of the analyzed ESTs and cDNA sequences are listed in Supplemental Table SI. The differential splicing events of the AtCBL10 and AtCIPK3 transcripts are exemplarily depicted in Figure 2. In the case of AtCBL10, the processing of an alternative exon leads to two potential protein isoforms with different N-terminal sequences (AtCBL10-1, MEQVSSRSSSLT...; AtCBL10-2, MTTGRPNILALKISTRRSSSLT...). The differential splicing of CBL2 and CBL4 transcripts affects the composition of their mature 5′-UTR. Interestingly, one spliced intron in the 5′-UTR of AtCBL4/SOS3 encompasses 3.7 kb of genomic sequence, a fact that needs to be considered when attempting to study expression of this gene by the analyses of promoter-reporter gene constructs in transgenic plants. For AtCIPK3, we detected three alternatively spliced forms, which differ in the C-terminal part of the protein. An alignment of these three protein sequences is depicted in Supplemental Fig. S1. These isoforms arose by differential splicing of intron 11 within the otherwise completely processed full-length mRNAs. In isoform AtCIPK3-2, the recognition of an alternative splice site leads to an RNA lacking parts of exon 12 and the complete exons 13 and 14. Therefore, the potentially specified AtCIPK3-2 protein would suffer a truncation of 69 amino acids. In the AtCIPK3-3 isoform, retaining of intron 11 results in an in-frame stop codon within this intron sequence. We identified three differentially spliced transcript forms of AtCIPK9, which are generated by alternative splicing of introns 7 and 13. Compared with AtCIPK9-1, the resulting variants would encode proteins either lacking two amino acids (AtCIPK9-2) or harboring four additional amino acids (AtCIPK9-3). Only a relative low number of full-length ESTs or cDNAs have been available for our analyses of potential alternative splicing events. This limits the statistical reliability of any prediction about the ratios of certain transcripts in planta. Nevertheless, the only four cDNAs identified for AtCBL10 represent two alternatively spliced forms (AtCBL10-1, two cDNAs; AtCBL10-2, two cDNAs), and the six identified AtCIPK3 cDNAs represent three isoforms (AtCIPK3-1, three cDNAs; AtCIPK3-2, one cDNA; AtCIPK3-3, two cDNAs). The rather even distribution of differentially spliced mRNAs argues against a for-
tuous detection of “mis-spliced” RNAs and makes it more likely that these alternatively spliced mRNAs are functional. Currently, it remains to be established how widespread the occurrence of alternative splicing is within the CBL and CIPK families and what the functional significance of the resulting protein isoforms could be. However, our identification of multiple differentially spliced isoforms is likely to add an additional level of complexity to the regulation of this signaling system.

Chromosomal Distribution and Evolution of CBLs and CIPKs

Analyses of the almost completely sequenced Arabidopsis genome have revealed that segmental duplications of chromosomal regions during evolution have significantly shaped the current structure of this genome (Grant et al., 2000; Bowers et al., 2003). These processes have been discussed as important mechanisms that significantly contributed to the evolution of Arabidopsis gene families (Blanc et al., 2000). We therefore analyzed the genomic distribution and evolutionary history of the CBL and CIPK genes. Most of the CBL genes are concentrated on chromosomes IV and V (Fig. 3A). While chromosomes II and III do not harbor CBL loci, AtCBL8 represents the only CBL gene located on chromosome I. In contrast, the 25 CIPK genes are distributed among all five Arabidopsis chromosomes (Fig. 3B).

Two independent duplications affecting different parts of chromosomes IV and V led to the formation of the AtCBL1/AtCBL9 and AtCBL2/AtCBL3 gene pairs, respectively (Supplemental Fig. S2). In contrast, a duplication affecting regions of chromosomes IV and I was followed by loss of one CBL gene copy (Supplemental Fig. S2). Therefore, the AtCBL8 gene, whose ancestor has been involved in this rearrangement, at present represents a singular CBL lacking a closely related counterpart. We also identified a single local tandem duplication leading to the formation of the AtCBL3/AtCBL7 gene pair. Interestingly, the two gene pairs AtCBL9 and AtCBL3 that arose by segmental duplications represent the most closely related CBL pairs. They exhibit a significantly higher degree of conservation than the AtCBL7 pair encoded by tandemly oriented genes. While both the AtCBL1/AtCBL9 and the AtCBL2/AtCBL3 protein pairs share 89% and 92% identical amino acid residues, AtCBL3 and AtCBL7 exhibit only 60% sequence identity (Supplemental Table SII). This might point to a relatively ancient origin of this tandemly oriented gene pair as compared with the identified segmental duplications.

Our whole-genome sequence comparison for the 25 CIPK loci uncovered altogether eight segmental duplication events (Fig. 3B; Supplemental Fig. S3). Six of these chromosomal rearrangements led to the presence of pairs of closely related CIPKs in the genome, whereas two loci duplications were followed by subsequent deletions of the affected kinases during the evolution of the Arabidopsis genome. In addition, we identified two CIPK gene pairs (AtCIPK14/AtCIPK15 and AtCIPK19/AtCIPK20), which arose by tandem duplications. As observed for the CBLs, all segmentally duplicated CIPK gene pairs (AtCIPK1/AtCIPK17, AtCIPK2/AtCIPK10, AtCIPK4/AtCIPK7, AtCIPK5/AtCIPK25, AtCIPK12/AtCIPK19, and AtCIPK13/AtCIPK18) encode the most closely conserved protein pairs, whereas proteins encoded by tandemly oriented genes (AtCIPK14/AtCIPK15 and AtCIPK19/AtCIPK20) exhibit a lower degree of conservation (Supplemental Table SIV). Taken together, our analysis of the chromosomal evolution of the CBL and CIPK gene families indicates that segmental duplications have predominantly contributed to the current
complexity of both Arabidopsis gene families. Only two of the 10 Arabidopsis CBL genes and four of the 25 AtCIPKs are encoded in tandem orientation. In contrast, four AtCBLs and 12 AtCIPKs were identified as segmentally duplicated loci, indicating the importance of this evolutionary mechanism for the growth and diversification of these gene families. A recent genome-wide analysis of NBS-Leu-rich repeat (LRR) genes in Arabidopsis (Meyers et al., 2003) reported frequent microscale rearrangements and extensive duplications as driving forces for the evolution of this large gene family. Moreover, these amplification processes led to the clustering of NBS-LRR sequences in specific chromosomal regions. Our analysis indicates that the evolution of the CBL and CIPK genes in Arabidopsis did not follow such a pattern. One potential explanation for this difference could be a strong selection pressure for gene dosages effects of CBLs that would select against extensive duplications of functional identical genes. Moreover, each newly formed CIPK would have to co-evolve with the corresponding CBL(s) regulating this kinase. These functional constraints might have influenced the evolutionary processes forming the extant CBL-CIPK signaling network.

To extend our studies on CBLs and CIPKs beyond Arabidopsis, we searched the publicly available EST (and genomic) databases for homologous sequences in other organisms. For CBLs, the applied criteria included a significantly higher similarity with the Arabidopsis CBLs than with CaM as well as the CBL-specific spacing of the EF hand motifs. Homologous CIPK-type kinases from other species were defined by both the similarity with the AtCIPKs and the presence of the NAF domain. Although these analyses systematically excluded partial ESTs covering only N-terminal parts of the encoded proteins, we nevertheless identified ESTs encoding CBLs and CIPKs in plant species like Medicago truncatula (nine different CBLs and 11 distinct CIPKs), wheat (Triticum aestivum; 11 CBLs and 29 CIPKs), barley (Hordeum vulgare; nine CBLs and 14 CIPKs), soybean (Glycine max; seven CBLs and 13 CIPKs), the gymnosperm Pinus sp. (two CBLs and seven CIPKs), and the moss Physcomitrella patens (four CBLs and three CIPKs). The accession numbers of the identified ESTs are displayed in the Supplemental Table S1b. The occurrence of multiple isoforms of CBLs and CIPKs in all analyzed plant species indicates that these proteins generally form a complex signaling network. We did not identify EST or genomic sequences encoding CBLs or CIPKs outside the plant kingdom, suggesting that the function of these signaling components is confined to plants but arose early in plant evolution.

We next investigated the occurrence of CBLs and CIPKs in the extensively sequenced genome of the monocotyledonous plant rice. To this end, we separately analyzed the available sequence data for the two rice subspecies O. sativa subsp. japonica and O. sativa subsp. indica (for details see “Materials and Methods”). In both genomes, we identified the same set of 10 calcium sensor proteins and 30 interacting kinases. The general features of the rice CBLs and CIPKs are summarized in Tables III and IV. The available sequences of both rice species do not cover their complete genomes and are only partially assembled. Nevertheless, the consistency of identified CBL and CIPK genes suggests that we uncovered the full complement of both gene families in rice. The rice genome is not only significantly larger than the Arabidopsis genome but also predicted to encode a greater number of genes (32,000–55,000 as compared with 27,300 in Arabidopsis; Goff et al., 2002; Yu et al., 2002; Wortman et al., 2003). As reported for several other gene families, for example, P-type ATPases and ABC transporters (Baxter et al., 2003; Jasinski et al., 2003), this general increase in the number of genes in rice obviously did not significantly affect the complexity of the CBL and CIPK gene families.

Comparison of the exon-intron composition of Arabidopsis and rice genes revealed a rather high degree

<table>
<thead>
<tr>
<th>Name</th>
<th>Clone Identification No.</th>
<th>Position</th>
<th>Intron</th>
<th>cDNA</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsCBL1</td>
<td>AC084763 (AAA01004016)</td>
<td>94244-103391</td>
<td>7</td>
<td>639</td>
<td>213</td>
</tr>
<tr>
<td>OsCBL2</td>
<td>AL713904 (AAA01001368)</td>
<td>5789-9843</td>
<td>7</td>
<td>678</td>
<td>226</td>
</tr>
<tr>
<td>OsCBL3</td>
<td>AC092390 (AAA01005520)</td>
<td>28690-34606</td>
<td>7</td>
<td>675</td>
<td>225</td>
</tr>
<tr>
<td>OsCBL4</td>
<td>AC097111 (AAA01001069)</td>
<td>26564-28044</td>
<td>7</td>
<td>633</td>
<td>210</td>
</tr>
<tr>
<td>OsCBL5</td>
<td>AP003504 (AAA01005392)</td>
<td>59100-61578</td>
<td>7</td>
<td>642</td>
<td>214</td>
</tr>
<tr>
<td>OsCBL6</td>
<td>AL713906 (AAA01000906)</td>
<td>69628-73277</td>
<td>7</td>
<td>678</td>
<td>226</td>
</tr>
<tr>
<td>OsCBL7</td>
<td>AP004212 (AAA01011541)</td>
<td>16316-18296</td>
<td>7</td>
<td>642</td>
<td>214</td>
</tr>
<tr>
<td>OsCBL8</td>
<td>CLC10421 (AAA01023033)</td>
<td>30197-32277</td>
<td>7</td>
<td>642</td>
<td>214</td>
</tr>
<tr>
<td>OsCBL9</td>
<td>AP002908 (AAA01061971)</td>
<td>12168-123750</td>
<td>8</td>
<td>870</td>
<td>290</td>
</tr>
<tr>
<td>OsCBL10</td>
<td>AP002901 (AAA0102074)</td>
<td>128202-130821</td>
<td>8</td>
<td>747</td>
<td>249</td>
</tr>
</tbody>
</table>

* Alternative rice bacterial artificial chromosome clone from the second subspecies (accessions beginning with AAAA.. refer to O. sativa subsp. indica, all other accessions indicate O. sativa subsp. japonica).  
* Gene position on the corresponding rice bacterial artificial chromosome clone listed in the “Clone Identification No.” column.  
* No. of introns within the coding region.  
* Length of the cDNA (coding region) in basepairs.  
* No. of amino acids of the predicted protein.
of structural conservation (data not shown). The rice genes OsCBL1 to 7 all harbor seven introns, which are absolutely invariant in phase and position when compared with the respective conserved seven introns present in six of the 10 AICBLs. This situation suggests that the single intron losses observed in AICBL3, 5, 6, and 7 occurred after the divergence of both evolutionary lineages. These seven introns are conserved in OsCBL9 and OsCBL10. Interestingly, these two genes share the additional first intron within their N-terminal extension with their closely related counterpart AICBL10.

As observed for the Arabidopsis CIPKs, the members of the rice CIPK gene family can also be divided into intron-less (eight OsCIPKs) and intron-harboring (22 OsCIPKs) representatives (Table IV). The latter exhibit a high degree of intron conservation with their relative Arabidopsis gene members and harbor 11 to 13 introns, which are all conserved in phase and position. Gene structures with 11 or 12 introns remaining can clearly be attributed to single or double intron loss events. For example, OsCIPK1 and OsCIPK21 have lost just one of these conserved introns (homologous to intron 8 in AICPK8).

In our analysis, we identified the two genes OsCIPK14 and OsCIPK15 as located on different chromosomes in the both rice subspecies. Both predicted proteins are almost identical and differ only by a single amino acid substitution and in their C-terminal nine amino acids. Therefore, this CIPK pair might represent an example for a very recent duplication event in this species. Our genomic analyses of rice CBL and CIPK genes did not detect any tandemly duplicated OsCBL genes. Of the 30 OsCIPKs, three gene pairs appeared to be organized in tandem orientation (OsCIPK5/13, OsCIPK12/30, and OsCIPK2/29). Moreover, these gene pairs are only relatively distantly related, suggesting a rather ancient origin of the duplications (see below). Therefore, despite the larger genome size of rice as compared with Arabidopsis, microscale recombination events do not seem to have significantly affected the evolution of both rice gene families.

<table>
<thead>
<tr>
<th>Name</th>
<th>Clone Identification No.</th>
<th>Clone Identification No.</th>
<th>Position</th>
<th>Introns</th>
<th>cDNA</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsCIPK01</td>
<td>AP003428</td>
<td>(AAAA01031723)</td>
<td>11550-120039</td>
<td>12</td>
<td>1,383</td>
<td>461</td>
</tr>
<tr>
<td>OsCIPK02</td>
<td>AP003367</td>
<td>(AAAA0103105)</td>
<td>48660-49991</td>
<td>13</td>
<td>1,329</td>
<td>443</td>
</tr>
<tr>
<td>OsCIPK03</td>
<td>AP002818</td>
<td>(AAAA0100189)</td>
<td>85611-85857</td>
<td>12</td>
<td>1,335</td>
<td>445</td>
</tr>
<tr>
<td>OsCIPK04</td>
<td>AL837528</td>
<td>(AAAA0102226)</td>
<td>145269-146633</td>
<td>12</td>
<td>1,362</td>
<td>454</td>
</tr>
<tr>
<td>OsCIPK05</td>
<td>AP003052</td>
<td>(AAAA01008215)</td>
<td>13923-12538</td>
<td>12</td>
<td>1,383</td>
<td>461</td>
</tr>
<tr>
<td>OsCIPK06</td>
<td>AAAA01000701</td>
<td>(CL83888)</td>
<td>17837-19192</td>
<td>12</td>
<td>1,353</td>
<td>451</td>
</tr>
<tr>
<td>OsCIPK07</td>
<td>AAAA01012457</td>
<td>(CL017116)</td>
<td>3882-4219</td>
<td>12</td>
<td>1,353</td>
<td>455</td>
</tr>
<tr>
<td>OsCIPK08</td>
<td>AP003449</td>
<td>(AAAA01001679)</td>
<td>74597-82361</td>
<td>13</td>
<td>1,326</td>
<td>441</td>
</tr>
<tr>
<td>OsCIPK09</td>
<td>AC104427</td>
<td>(AAAA01000865)</td>
<td>32041-35803</td>
<td>13</td>
<td>1,368</td>
<td>456</td>
</tr>
<tr>
<td>OsCIPK10</td>
<td>AC131175</td>
<td>(AAAA01023197)</td>
<td>68596-69691</td>
<td>12</td>
<td>1,506</td>
<td>502</td>
</tr>
<tr>
<td>OsCIPK11</td>
<td>AP003230</td>
<td>(AAAA01008532)</td>
<td>102333-103861</td>
<td>12</td>
<td>1,506</td>
<td>502</td>
</tr>
<tr>
<td>OsCIPK12</td>
<td>AP003256</td>
<td>(AAAA01031578)</td>
<td>38475-36913</td>
<td>12</td>
<td>1,560</td>
<td>520</td>
</tr>
<tr>
<td>OsCIPK13</td>
<td>AP003052</td>
<td>(AAAA01023672)</td>
<td>2348-3883</td>
<td>12</td>
<td>1,533</td>
<td>511</td>
</tr>
<tr>
<td>OsCIPK14</td>
<td>BX000511</td>
<td></td>
<td>11199-113310</td>
<td>12</td>
<td>1,317</td>
<td>439</td>
</tr>
<tr>
<td>OsCIPK15</td>
<td>BX000512</td>
<td></td>
<td>111783-112587</td>
<td>12</td>
<td>1,302</td>
<td>434</td>
</tr>
<tr>
<td>OsCIPK16</td>
<td>AAAA01003250</td>
<td>(CL01893)</td>
<td>7861-9231</td>
<td>12</td>
<td>1,368</td>
<td>456</td>
</tr>
<tr>
<td>OsCIPK17</td>
<td>AAAA01004506</td>
<td>(CL03998)</td>
<td>10945-14154</td>
<td>11</td>
<td>1,362</td>
<td>454</td>
</tr>
<tr>
<td>OsCIPK18</td>
<td>AAAA01010065</td>
<td>(CL24701)</td>
<td>22322-23695</td>
<td>11</td>
<td>1,371</td>
<td>457</td>
</tr>
<tr>
<td>OsCIPK19</td>
<td>AB011967</td>
<td>(AAAA01011190)</td>
<td>3641-5267</td>
<td>12</td>
<td>1,527</td>
<td>509</td>
</tr>
<tr>
<td>OsCIPK20</td>
<td>AAAA01036134</td>
<td>(HTC152200-B01)</td>
<td>1-1361/1360-1401</td>
<td>12</td>
<td>1,398</td>
<td>466</td>
</tr>
<tr>
<td>OsCIPK21</td>
<td>AP003749</td>
<td>(AAAA01019996)</td>
<td>64996-69312</td>
<td>12</td>
<td>1,491</td>
<td>497</td>
</tr>
<tr>
<td>OsCIPK22</td>
<td>AAAA01005133</td>
<td>(CL008902)</td>
<td>3820-5175</td>
<td>12</td>
<td>1,353</td>
<td>451</td>
</tr>
<tr>
<td>OsCIPK23</td>
<td>AP003703</td>
<td>(AAAA01001330)</td>
<td>16522-20217</td>
<td>13</td>
<td>1,350</td>
<td>450</td>
</tr>
<tr>
<td>OsCIPK24</td>
<td>AAAA01002034</td>
<td>(CL024530)</td>
<td>1909-7271</td>
<td>13</td>
<td>1,359</td>
<td>453</td>
</tr>
<tr>
<td>OsCIPK25</td>
<td>AAAA01000413</td>
<td>(CL018106)</td>
<td>16907-18457</td>
<td>12</td>
<td>1,542</td>
<td>514</td>
</tr>
<tr>
<td>OsCIPK26</td>
<td>AAAA01003195</td>
<td>(HTC028888-A01)</td>
<td>25212-26693</td>
<td>12</td>
<td>1,479</td>
<td>493</td>
</tr>
<tr>
<td>OsCIPK27</td>
<td>CL000365</td>
<td>(AAAA0101351)</td>
<td>11848-13185</td>
<td>12</td>
<td>1,335</td>
<td>445</td>
</tr>
<tr>
<td>OsCIPK28</td>
<td>AAAA01001836</td>
<td>(CL024530)</td>
<td>16944-13001</td>
<td>12</td>
<td>1,305</td>
<td>435</td>
</tr>
<tr>
<td>OsCIPK29</td>
<td>AP003757</td>
<td>(AAAA01003105)</td>
<td>32507-33838</td>
<td>12</td>
<td>1,329</td>
<td>443</td>
</tr>
<tr>
<td>OsCIPK30</td>
<td>AP003256</td>
<td>(AAAA01019275)</td>
<td>30049-28619</td>
<td>12</td>
<td>1,428</td>
<td>476</td>
</tr>
</tbody>
</table>

* Alternative name, OsPK07 (Ohba et al., 2000).  
* Alternative name, OsPK04 (Ohba et al., 2000).  
* Gene position on the corresponding rice bacterial artificial chromosome clone listed in column “Clone Identification No.”  
* Length of the cDNA (coding region) in basepairs.  
* No. of amino acids of the predicted protein.  
* Copyright © 2004 American Society of Plant Biologists. All rights reserved.
To further elucidate the relationships of CBLs and CIPKs, we performed phylogenetic analyses to infer clustering patterns reflecting sequence similarity and evolutionary ancestry. Plant CaM as well as animal CNB and NCS were included as outgroups in the analysis of the calcium sensor proteins. The resulting unrooted phylogenetic tree is depicted in Figure 4. (To maintain clarity, this tree does not present bootstrap values. A cladogram displaying all bootstrap values is presented in Supplemental Fig. S4.) Plant CBLs clearly form a monophyletic group separated from CaMs and calcium sensor proteins from animals. Moreover, Arabidopsis CBL pairs (AtCBL1/AtCBL9 and AtCBL2/AtCBL3) resulting from segmental duplications of chromosomal regions appear to be most closely related, again pointing to a relatively recent formation of these gene pairs. In contrast, the CBL pair AtCBL3/AtCBL7, which evolved by tandem duplication, is far less related.

The evolutionary comparison of rice and Arabidopsis CBL genes clearly points to the occurrence of multiple species-specific duplication events. Of course, these duplications could also be diagnostic for larger evolutionarily related groups, like monocotyledonous and dicotyledonous species. However, the lack of additional plant genome sequences currently prevents a further investigation of this issue. This situation will most likely impede a precise prediction of common CBL gene functions based solely on sequence information from rice and Arabidopsis. For example, the two closely related calcium sensor proteins pairs AtCBL2/3 and OsCBL2/3 result from duplication events that occurred independently in both species analyzed. OsCBL9 and 10 are clearly related to AtCBL10 by both sequence similarity and intron composition. However, it is currently impossible to address whether the rice-specific duplication leading to OsCBL9 and 10 could result in functional overlaps or, alternatively, reflects a functional specialization. Likewise, the single OsCBL1 gene is most closely related to the duplicated AtCBL1 and AtCBL9 genes, which probably fulfill different functions in Arabidopsis (see below). Nevertheless, in combination with additional information, such as CBL-CIPK interaction specificity and gene expression data, knowledge of these evolutionary relations will help to design reverse genetic strategies to unravel CBL gene functions.

Interestingly, the distribution of myristoylated Arabidopsis and rice CBL proteins coincides with two neighboring branches in the dendrogram. The only predicted non-myristoylatable calcium sensor in these branches (AtCBL8) still harbors a recognizable degenerated myristoylation consensus sequence. This might suggest a single early event during the evolution of the CBL genes leading to the loss of this lipid modification in the remaining CBL lineages. Such a diversification would have enabled the separate evolution of membrane-associated and membrane-independent calcium-signaling pathways.

Plant SnRK1 and SnRK2 kinases were included as outgroups in the phylogenetic analysis of the Arabidopsis and rice CIPKs. The resulting unrooted phylogenetic tree is depicted in Figure 5. (To maintain clarity this tree does not present bootstrap values. A cladogram displaying all bootstrap values is presented in Supplemental Fig. S5.) This dendrogram clearly indicates a monophyletic origin of the com-

---

**Figure 4.** Phylogenetic relationships of Arabidopsis and rice CBLs with related calcium-binding proteins. Alignment of full-length protein sequences and phylogenetic analyses were performed as described in “Materials and Methods.” The accession numbers of the Arabidopsis and rice CBLs are presented in Tables I and III. Arabidopsis CaM 1 (AtCaM1, accession no. NP_198594), human CaM 1 (HsCaM1, accession no. AAA51918), yeast CNB (ScCNB, accession no. P25296), human NCS 1 (HsNCS1, accession no. NP_055101), and human recoverin (HsRCV, accession no. BAA19460) were included in the analyses as the most closely related non-CBL-type calcium-binding proteins. Stars indicate myristoylated AtCBLs and potentially myristoylated OsCBLs.
plex family of plant CIPKs. Interestingly, members of both the rice and the Arabidopsis CIPKs form two distinct subgroups based on their sequence similarity. This clustering of CIPKs strictly coincides with the presence (or absence) of introns in the genes of these subgroups, indicating that the ancestors of both groups have been present in the genome before the evolutionary separation of the mono- and dicotyledonous lineages. Moreover, several groups of closely related Arabidopsis and rice CIPKs form well-supported separate branches. This might suggest a relatively early diversification of the CBL and CIPK gene families during plant evolution.

The resolution of evolutionary relationships between individual CIPK proteins within these branches varies significantly. For example, the rice and Arabidopsis representatives of CIPK8, 9, 23, and 24 are clearly recognizable as distinct orthologous gene pairs. This might allow for prediction of similar functions of their gene products in both species. In contrast, in both the CIPK4/7 and the CIPK1/17/21 branches, our analysis indicates parallel species-specific duplication events. This prevents an interspecific gene-to-gene assignment of distinct CIPKs. However, it appears noteworthy that in rice and Arabidopsis, the evolution of these branches followed the same pattern resulting in the closely related gene groups CIPK4/7 and CIPK1/17/21 in both species. The slightly increased number of CIPKs in rice compared with Arabidopsis can be largely attributed to rice-specific duplications in the evolutionary branch encompassing AtCIPK2, 6, 15, and 20 and the respective rice homologs. Although this branch just represents four CIPKs from Arabidopsis, it encompasses 12 members of the rice kinase family. It needs to be analyzed if the amplification of this rice CIPK subgroup just results in functional overlaps or instead points to an evolutionary specialization. Finally, our data indicate that the complexity of this signaling network is not restricted to Arabidopsis and therefore is likely to reflect functional diversification.

Comparative Analyses of the Proteins

The CBL protein family appears to be rather conserved in size and structure. Most of the genes encode predicted polypeptides ranging from 23.5 to 26 kD in Arabidopsis and 23.9 to 25.9 kD in rice (Tables I and III). The only noted exceptions are represented by AtCBL10 and the recently duplicated OsCBL9 and OsCBL10, which, due to an N-terminal extension, have predicted molecular masses of 31.7, 32.9 and 29.9 kD, respectively. Although the N termini of AtCBL10 and OsCBL9/10 do not harbor any discernible sequence or signal motif, it appears conceivable that this domain is of functional relevance. The amino acid sequence identity of different CBLs ranges from 29% to 92% (33%–93% similarity) in Arabidopsis and from 40% to 92% (44%–94% similarity) in rice (Supplemental Tables SII and SIII). In Arabidopsis, the most closely related proteins AtCBL1 and AtCBL9 as well as AtCBL2 and AtCBL3
Calcium-binding proteins contain EF hand motifs as structural basis for calcium binding. Each EF hand consists of a loop of 12 amino acids flanked by two α-helices. A single Ca\(^{2+}\) ion is bound to each EF hand via the loop domain. In CaM, which harbors four EF motifs, these sites act pair wise resulting in a cooperativity of calcium binding (Resh, 1999). Myristoylation of the mammalian calcium sensor recoverin is a prerequisite for the occurrence of the calcium-myristoyl switch leading to a calcium-dependent membrane association of recoverin (Ames et al., 1997). Our analysis revealed that four of the 10 Arabidopsis CBL proteins harbor a conserved myristoylation motif (Fig. 6A). In rice, five CBL proteins (OsCBL1, 4, 5, 7, and 8) represent potential substrates for this lipid modification. Myristoylation has been shown to be required for the function of AtCBL4/SOS3, although the exact role of myristoylation of this protein remains a mystery (Ishitani et al., 2000). Moreover, AtCBL1, 5, and 9 appear to become myristoylated when translated in vitro in a rabbit reticulocyte lysate (O. Batistic and J. Kudla, unpublished data). On the other hand, interaction of AtCBL1 with its target kinases occurs independently of myristoylation (Albrecht et al., 2001). Therefore, a possible function of CBL myristoylation could be the membrane targeting of CBL-CIPK complexes. The hydrophobicity of a myristoylated peptide alone is usually not sufficient to anchor the protein to the plasma membrane. Stable membrane association of myristoylated proteins is often accompanied by palmitoylation of an adjacent Cys residue (Resh, 1999). This palmitoylatable Cys residue is present in all four myristoylated CBLs from Arabidopsis as well as in all potentially myristoylated OsCBLs, suggesting that these proteins could be membrane targeted by virtue of dual lipid modification (Fig. 6A).

Calcium-binding proteins contain EF hand motifs as structural basis for calcium binding. Each EF hand consists of a loop of 12 amino acids flanked by two α-helices. A single Ca\(^{2+}\) ion is bound to each EF hand via the loop domain. In CaM, which harbors four EF motifs, these sites act pair wise resulting in a cooperativity of calcium binding (Haeseleer et al., 2002). Calcium sensors of the NCS/recoverin/guanylate cyclase activating protein-type exhibit a different mode of calcium binding, although they harbor four EF hands like CaM (Burgoyne and Weiss, 2001). In recoverin, attachment of calcium is commonly initiated only on EF hands 2 and 3 and leads to significant conformational changes of the affected protein. Because of a Cys-Pro substitution in the 12 amino acid loop within the first EF hand, this motif is nonfunctional in terms of calcium binding in all NCSs. Instead, this EF hand can mediate the interaction with their target proteins, like for example guanylate cyclase in the case of GCAP-2 (Ermilov et al., 2001). Several members of the NCS/recoverin/GCAP calcium sensor group exhibit variations from the canonical EF hand sequence in single EF hands (Haeseleer et al., 2002). These mutations modulate but do not abolish the calcium-binding affinity of each specific calcium sensor, thereby providing a mechanism to differentially decipher various calcium signals.

Our comparative analyses of the EF hand composition in the Arabidopsis CBL protein family revealed extensive similarities to the NCS-type calcium sensors. All 10 AtCBL proteins harbor four EF hand motifs (Fig. 6A). Interestingly, the size of the linker regions between the EF hand loops is absolutely conserved in all proteins and appears to be unique to this family of calcium sensor proteins. EF1 and EF2 are 22 amino acids apart, whereas 25 amino acids separate EF2 from EF3, and 32 amino acids are inserted between EF3 and EF4. The number of EF hands as well as their spacing is also absolutely conserved in all predicted OsCBL proteins (data not shown). Size variation of CBL proteins is therefore exclusively caused by extension or reduction of the N- and C-terminal regions (Fig. 6A). This might suggest a conserved three-dimensional structure of all plant CBLs. Also, like in NCS proteins, the first EF hand, although conserved in all CBLs, appears to be unable to bind Ca\(^{2+}\) ions. This is caused by either single-amino acid substitutions at critical calcium-binding positions (either positions X, Y, and Z or −Y, −X, and −Z; for the Arabidopsis representatives, see Fig. 6B) or larger structural changes (Fig. 6B). In most cases, the oxygen-containing calcium-binding residues at the Y and/or X positions in this EF loop are altered to Ala, Val, Ile, or Cys (as observed in EF1 of GCAP proteins). In addition, in AtCBL5, the essential amino acid at the −Z position is replaced by the positively charged residue Lys. Both AtCBL6 and AtCBL7 display structural changes in the first EF hand. In EF1, four amino acids of the canonical loop in AtCBL6 appear to be deleted. In AtCBL7, EF1 harbors five additional amino acids between the fourth and fifth amino acids (Z position) of the loop, thereby dividing the EF hand motif. Furthermore, several amino acids have been deleted from the entering helix, which could lead to a protein with a different structural conformation in the region of EF1.

AtCBL6, AtCBL7, and AtCBL8 also contain additional mutations in other EF hands (Fig. 6B). In the EF3 loop of AtCBL8, the normally conserved Y residue is changed to the non-oxygen-containing amino acid glutamine. Moreover, 13 amino acid deletions appear to be present in the EF4 loop of AtCBL8. These structural changes might explain the site-specific phosphotyrosine phosphorylation of CBL proteins by CIPK kinases (Kudla, unpublished data). On the other hand, interaction of AtCBL4 with its target kinases is independent of myristoylation (Albrecht et al., 2001). Therefore, a possible function of CBL myristoylation could be the membrane targeting of CBL-CIPK complexes. The hydrophobicity of a myristoylated peptide alone is usually not sufficient to anchor the protein to the plasma membrane. Stable membrane association of myristoylated proteins is often accompanied by palmitoylation of an adjacent Cys residue (Resh, 1999). This palmitoylatable Cys residue is present in all four myristoylated CBLs from Arabidopsis as well as in all potentially myristoylated OsCBLs, suggesting that these proteins could be membrane targeted by virtue of dual lipid modification (Fig. 6A).
acid His. In AtCBL6, the conserved Glu at the \(-Z\) position of EF2 is altered to Ala, and at the same position in the EF3 loop of AtCBL7, the calcium-binding residue is changed to Gly. Both mutations could lead to a change in the loop conformation and therefore to a markedly reduced affinity for calcium (Lewit-Bentley and Réty, 2000). In addition, we detected a remarkable variation of the EF hand sequences in all AtCBL proteins (Fig. 6B). Most of the observed deviations from the canonical EF hand sequence are predicted to result in a reduced affinity toward calcium ions. For example, in the remaining EF hands of AtCBL2, 3, 4, and 5, the oxygen-containing amino acid at the Y position is replaced by...
the positively charged amino acid Lys or Arg. The same mutation is also observed in EF2 and EF4 of AtCBL6, EF3 and EF4 of AtCBL7, EF2 of AtCBL8, and EF2 and EF3 of AtCBL10. Moreover, in the binding loop of EF2 of AtCBL1 and AtCBL9, the oxygen amino acid at the Z position is changed to Lys (Fig. 6B). Taken together, this analysis revealed that in AtCBL1 and AtCBL9, EF binding loops 3 and 4 represent a normal canonical sequence. The canonical EF hand sequence can also be identified in the EF-binding loop 3 of AtCBL6, EF2 of AtCBL7, and EF4 of AtCBL8 and AtCBL10. Therefore, the CBL protein family can be divided into proteins with one canonical EF hand (AtCBL6, 7, 8, and 10), with two canonical EF hands (AtCBL1 and 9) and without any canonical EF hand (AtCBL2, 3, 4, and 5; Fig. 6A). A similar variation of EF hand sequences appears to occur in the predicted rice CBL proteins (data not shown). These differences in the EF hand composition in individual CBLs could lead to different affinities toward calcium ions. Whether such differences in calcium-binding affinity contribute to deciphering the different calcium signals in response to various environmental stimuli awaits further experimental analyses. So far, no quantitative Ca\textsuperscript{2+}-affinity studies have been reported for any CBL protein, and side-by-side comparisons of Ca\textsuperscript{2+} binding have not been performed. AtCBL1 has been shown to bind calcium in vitro in gel-overlay assays and displays a shift during electrophoresis after calcium binding (Kudla et al., 1999). Moreover, in vitro studies suggest that interaction of AtCBL1 with AtCIPK1 requires micromolar levels of Ca\textsuperscript{2+} (Shi et al., 1999). Up to now, the function of CBL1-CIPK1 interaction in regulating CIPK1 kinase activity has not been investigated, because the identity of CIPK1 substrates has remained elusive. In contrast, in vitro studies of SOS3/AtCBL4 revealed that this protein binds calcium in overlay assays, but that this binding is too weak to induce a mobility shift during electrophoretic separation (Ishitani et al., 2000). Moreover, at least in vitro, interaction of SOS3/AtCBL4 with SOS2/AtCIPK24 appears to occur calcium independently (Halfter et al., 2000).

Nevertheless, in vitro, the autophosphorylation activity of SOS2/AtCIPK24 and its kinase activity toward synthetic substrates are further stimulated by interaction with both SOS3/AtCBL4 and calcium (Guo et al., 2001, Gong et al., 2002c). Therefore, it appears conceivable that significant variations in Ca\textsuperscript{2+} affinity of individual CBLs and variation in Ca\textsuperscript{2+} dependence of CBL-CIPK interaction and activation might exist. This potential aspect of Ca\textsuperscript{2+} regulation in this signaling system clearly deserves more attention.

Interestingly, CBL proteins resulting from gene duplications (AtCBL1 and AtCBL9) harbor an absolutely identical amino acid composition in the EF hand loops 1 and 2. Because EF hands in animal NCSes and CNBs form functional pairs, it appears possible that the N-terminal parts of AtCBL1 and 9 fulfill identical functions (e.g. protein-protein interaction via EF1). Contrarily, the two EF hands at the C-terminal ends of theses proteins have diverged during evolution, resulting in EF motifs with potentially distinct calcium-binding affinities. Therefore, the differences in the last two EF hand loops could lead to a different affinity toward calcium ions and consequently could convey different functions of the highly related calcium sensors AtCBL1 and AtCBL9.

Like the CBLs, the 25 AtCIPK and the predicted 30 OsCIPK proteins display a high degree of conservation. In Arabidopsis, the proteins exhibit an overall amino acid sequence identity ranging from 33% to 78%. The sizes of the predicted proteins vary from 46 to 59 kD (Table 1; Supplemental Table SIV). Pair-wise amino acid alignments of the highly conserved N-terminal kinase domain revealed 51% to 90% sequence identity (63%–96% similarity). In contrast, the C-terminal non-catalytic region appears to be much less conserved, exhibiting an amino acid identity of 24% to 58% (36%–69% similarity). The only exception is represented by the 24-amino acid NAF domain (58%–86% identity; 66%–94% similarity). A similar degree of conservation in size and structure is also typical for the rice CIPKs (Table IV; Supplemental Table). All CIPK proteins harbor the 11 conserved subdomains typical for Ser/Thr kinases. Biochemical analyses of AtCIPKs expressed in bacteria revealed a rather low enzymatic activity of the recombinant protein (Shi et al., 1999; Guo et al., 2001). For several AtCIPKs, it has been established experimentally that a substitution of a conserved Thr (within the activation loop) to Asp results in a hyperactive enzyme (Gong et al., 2002b, 2000c). Thus phosphorylation-dependent activation of CIPKs thus might provide a mechanism to convey the cellular cross-talk with other signaling systems like CDPKs and mitogen-activated protein kinases.

To investigate the potential localization and modification of the CIPKs, we performed bioinformatics analyses using the PSORT and the MOTIF algorithms. None of the CIPKs appears to harbor any conserved lipid modification (e.g. myristoylation, palmitoylation, and farnesylation) motif or displays any detectable unambiguous cellular sorting sequence. Therefore, the determination of the sub-cellular targeting of the CIPKs will require experimental confirmation for each of these kinases. Nevertheless, a combination of CBL localization data together with information about the specificity of CBL-CIPK complex formation should allow experimentally testable predictions about the sub-cellular targeting of specific CIPKs.

Sequence Similarity of CBLs and CIPKs Does Not Necessarily Result in Functional Redundancy

Our previous yeast two hybrid analyses on a subset of Arabidopsis CBLs and CIPKs have established that specific CBL-CIPK complex formation contributes to
generating specificity in this signaling network and that this differential calcium sensor/kinase affinity is also reflected in in vitro experiments with purified proteins (Albrecht et al., 2001). To exemplarily unravel the interaction preferences of two closely related CBL proteins with the complete set of Arabidopsis CIPKs, we studied AtCBL1 and AtCBL9 interactions with AtCIPK1 to AtCIPK25 in yeast two-hybrid interaction analyses. To this end, full-length cDNAs of AtCBL1, AtCBL9, and all AtCIPKs were cloned in the respective DNA-binding domain and GAL4 activation domain plasmids. After transformation into the yeast strain SMY3, interaction was monitored by growth on interaction-selecting media lacking His. As shown in Figure 7, in these analyses, AtCBL1 exhibited a significant interaction only with a subset of six CIPKs (AtCIPK1, 7, 8, 17, 18, and 24). For four additional CIPKs, we observed a weak tendency of complex formation, whereas 12 CIPKs did not show any affinity toward AtCBL1. Our data did not reproduce the AtCBL1-AtCIPK15 interaction reported by Guo et al. (2002; in this study designated as SCaBP5 and PKS3). This difference could, for example, result from the different vector systems and procedures used in both studies.

Interestingly, the observed preferential complex formation does not appear to coincide with the phylogenetic relationships of the AtCIPKs, because, for example, members of the evolutionary separated branches of intron-harboring (AtCIPK1, AtCIPK8, AtCIPK17, and AtCIPK24) as well as intron-free CIPKs (AtCIPK7 and AtCIPK18) interact with AtCBL1. In addition, of the closely related duplicated CIPK pairs, only AtCIPK1/AtCIPK17 exhibited a

![Figure 7. Comparative yeast two-hybrid interaction analysis of AtCBL1 and AtCBL9 with all Arabidopsis CIPKs. The yeast strain SMY3 containing the indicated plasmid combinations were grown on SC medium –L and –W to an A_{600} of 2. Ten microliters of 10-fold dilution series was spotted onto selective (–H, –L, and –W) and nonselective (–L and –W) media. Decreasing cell densities in the dilution series are illustrated by narrowing triangles. LW depicts a representative dilution series on nonselective plates. The observed interaction intensities are illustrated at the right. XX indicates strong interaction (growth detectable to at least 10^{-4} dilution). X indicates weak interaction (growth detectable at 10^{-3} dilution; –, no interaction).]
similar affinity toward AtCBL1, whereas, for example, the pairs AtCIPK4/AtCIPK7 and AtCIPK13/AtCIPK18 displayed a rather different interaction profile. These data indicate that sequence similarity and evolutionary history are not sufficient to predict CBL-CIPK interactions. Therefore, the exact structural features determining the specificity of complex formation remain to be uncovered. In addition, these findings might suggest that even a high conservation of certain CBL or CIPK family members does not necessarily indicate functional redundancy.

In our analyses, AtCBL9 displayed strong interaction with six different CIPKs and some degree of interaction with additional six CIPKs (Fig. 7). Of the strongly interacting kinases, only a subset of four CIPKs (AtCIPK1, 8, 18, and 24) exhibited a similar affinity toward both CBL proteins. In contrast, AtCIPK7 and AtCIPK17 preferentially interact with AtCBL1, whereas AtCIPK21 and AtCIPK23 appear to interact more efficiently with AtCBL9.

Although the interaction observed in yeast does not necessarily mean that these proteins also interact in planta, these data nevertheless indicate that even if several kinases are expressed simultaneously in a single cell, differential interaction affinity can facilitate an efficient signal channeling. Moreover, despite the high similarity of AtCBL1 and AtCBL9, our data suggest that both calcium sensor proteins could target specific kinases with different efficiency. Kinases that are expressed in the same cells and interact equally strongly with both CBL proteins may potentially represent signaling nodes for different stimuli received by the two CBL proteins.

To further address this aspect, we comparatively analyzed the expression patterns of the closely related AtCBL1 and AtCBL9 genes in response to different environmental cues. Two-week-old seedlings were challenged with cold, drought, or salt stress, and RNA was extracted from samples harvested at the indicated time points (Fig. 8). Semi quantitative RT-PCR analyses were performed with gene-specific primers to determine the expression levels of AtCBL1, AtCBL9, and, as a constitutively expressed control, actin2. The expression of the stress-induced RD29A gene was analyzed as a control for the effectiveness of the performed treatments. These analyses revealed that whereas AtCBL1 gene expression was strongly induced under these conditions, AtCBL9 expression did not respond to these stimuli, providing additional evidence for different functions of the two proteins. This assumption is further corroborated by the recent finding that loss of AtCBL1 gene function impairs the ability of the affected mutant plants to respond to drought, cold, and salt stress (Albrecht et al., 2003; Cheong et al., 2003). These findings suggest that CBL1 represents an upstream regulator of stress gene expression and serves as a rate-limiting factor in multiple abiotic stress response pathways. Interestingly, despite the observed effects of both cbl1 knockout mutants on salt, drought, and cold stress responses, these mutant were indistinguishable from wild-type plants with regard to their responsiveness to the phytohormone ABA. This observation points to a function of CBL1 upstream or independent of any modulation of stress signaling by ABA. In contrast, in another investigation of CBL1 function (there designated as SCaBP5) the authors postulated a function of this calcium sensor protein as a global regulator of ABA responses (Guo et al., 2002). These contradictory findings might be explained by the RNA interference-based experimental approach pursued in the study by Guo et al. (in contrast to the T-DNA induced loss-of-function allele studies in the two other reports; discussed in detail by Albrecht et al. [2003]). Accordingly, our preliminary characterization of a T-DNA-induced cbl9 knockout mutant revealed a specific impairment of ABA responses in these plants (C. D’Angelo and J. Kudla, unpublished data).

Taken together, our data suggest that functional redundancy might only occur to a limited extent within the CBL-CIPK signaling network. Therefore, its components should be amenable to functional analyses by reverse genetic approaches using T-DNA or transposon-induced knockout lines. Our genomic and bioinformatics analyses of CBL and CIPK genes and proteins presented in this work can provide an important foundation for the further functional dissection of these important plant-specific signaling pathways.

**MATERIALS AND METHODS**

**Database Searches and Computational Analyses of CBL and CIPK Sequences and Genomic Loci**

The complement CBL and CIPK sequences encoded in the Arabidopsis genome were identified by sequence comparisons of previously identified AtCBLs and AtCIPKs with entries in the GenBank database (http://www.ncbi.nlm.nih.gov) and the Arabidopsis MIPS database (http://mips.gsf.de/proj/thal/db/index.html) using BLASTP and TBLASTN and the "mutil" algorithms (Altschul et al., 1990, 1997; Falquet et al., 2002). CBL and CIPK encoding sequences from rice (Oryza sativa) were identified using the same algorithms to search the rice genome sequences deposited in three different databases: (a) the nonredundant rice sequence subset of the Gen-
Bank database, containing the results of the International Rice Genome Sequencing Project (http://www.nbci.nlm.nih.gov/blast/), (b) the draft rice genome sequence of the Torrey Mesa Research Institute (http://portal.tmri.org/rice/RicePublicAccess.html; Goff et al., 2002), and (c) the draft rice genome sequence of the Chinese Academy of Sciences (http://www.ncbi.nlm.nih.gov/blast/Genome/PlantBlast.shtml; Yu et al., 2002). The first two sequence sets were derived from *O. sativa* subsp. *japonica* and the third from *O. sativa* subsp. *indica*. If possible, predictions of OsCBL and OsCIPK coding sequences were verified with available EST and full-length cDNA sequences. For verification and identification of special features in the UTRs or of splicing variants of the analyzed genes, BLASTN searches for corresponding cDNA full-length and EST sequences were performed in the GenBank database.

Deduced protein sequences were aligned with ClustalX (Thompson et al., 1997). Members of the CBL gene family were identified based on sequence comparison with other calcium-binding proteins as well as strictly conserved distances between the four EF hand loops. In addition to sequence conservation, the presence of the NAF domain (Albrecht et al., 2001) was the decisive criterion to include a kinase sequence into the CIPK gene family. Resulting multiple alignments were processed with BIOEDIT to determine the homology scores. For construction of phylogenetic trees, the CBL or CIPK amino acid sequences were aligned together with chosen representatives from related gene families. Tree reconstruction using the neighbor-joining method and confirmation of tree topology by bootstrap replications (1,000 replicates) were performed with ClustalX (default settings).

Analysis of chromosomal segment duplications affecting gene family. Resulting multiple alignments were processed with BIOEDIT 2001) was the decisive criterion to include a kinase sequence into the CIPK or CBL amino acid sequences. For construction of phylogenetic trees, the CBL or CIPK amino acid sequences were aligned together with chosen representatives from related gene families. Tree reconstruction using the neighbor-joining method and confirmation of tree topology by bootstrap replications (1,000 replicates) were performed with ClustalX (default settings). Analysis of chromosomal segment duplications affecting CBL or CIPK loci was performed as described (Kolukisaoglu et al., 2002).

cDNA Cloning, Sequence Analyses, Plasmid Constructions, and Yeast Two-Hybrid Assays

To identify conserved protein modification motifs and targeting sequences, the CBL and CIPK proteins were analyzed with the PROSITE (Falcon et al., 2002) and PSORT (http://psort.nibb.ac.jp/) programs, respectively. Potential N-terminal palmitoylation sites were identified by comparison of the N-terminal sequences with the N-terminal sequences of the known myristoylated and palmitoylated proteins Ara6 (Ueda et al., 2001; accession no. NP_567008) and lck (Paige et al., 1993; Yurchak and Setton, 1995; accession no. 148845).

The identification and cloning of *AtCBL1* to 6 and *AtCIPK* to 15 has been described previously (Albrecht et al., 2001). A complete list of primers used for RT-PCR, cDNA cloning, and expression studies during the course of the work is provided in Supplemental Table S1. RNA isolation, RT-PCR, plasmid constructions for *AtCBL7* to 10 and *AtCIPK* to 24, and yeast two-hybrid studies were performed as previously described (Albrecht et al., 2001). All yeast two-hybrid studies were performed in duplicate.

Structural and Motif Analysis of Proteins

The position of the EF hands within the CBL proteins was determined by comparing the complete sequence of all AtCBL proteins with the EF consensus sequence, available at PROSITE (PDOC00018). Comparisons of AtCBL EF hands 1, 2, 3, and 4 were performed based on the calculated consensus sequence. These consensus sequence was calculated with the ClustalW algorithm (Thompson et al., 1994) based on an alignment of all EF hands from HsCNB, AtCaM, and the three calcium-binding EF hands from HsNCS-1, HsGCAP-2, and the canonical EF hands identified in the different CIPKs.

Stress Treatments and Quantitative RT-PCR Analyses

In all stress experiments described, tissue samples of Arabidopsis cv Wassilewskija were taken after 0, 1, 3, 6, 12, and 24 h, immediately frozen in liquid nitrogen, and stored at −80°C until use. Cold treatments were performed with 14-d-old soil-grown seedlings, which were transferred from 21°C to 2°C and kept under these conditions for 24 h in the dark. For drought assays, plants were grown hydroponically in Magenta boxes for 14 d on one-half-strength Murashige and Skoog medium in 35% Hoagland’s Suc. Magenta boxes were transferred into a flow chamber, and plants were exposed to a stream of dry air in a flow bench for 20 min (resulting in approximately 10% fresh weight loss) and then returned to Magenta boxes. Afterward the Magenta boxes were closed and samples were taken at the described time points. For salt stress response analyses, the hydroponic cultivation medium was supplemented with 100 mM NaCl, and samples were taken at the respective time points.

RNA for quantitative RT-PCR analyses was isolated from 150-mg tissue samples using TriFast solution (Peqlab, Erlangen, Germany). cDNA syntheses were performed using 3 μg of random hexamer primer and 200 units of Moloney murine leukemia virus RT (Promega, Madison, WI) for 1 h at 42°C. Two hundred nanograms of cDNA was used as template in all PCR reactions with gene-specific primers. In preliminary experiments, we first determined for each gene to be analyzed the cycle number that was within the linear range of PCR product amplification. Subsequently, PCR amplifications (1 min, 94°C; 1 min 30 s, 58°C; 1 min, 72°C; and volume, 50 μL) were performed for 23 cycles. The six transcripts (corresponding to each time point) were assayed simultaneously in separate tubes using identical cDNA aliquots and the respective gene-specific primers. Aliquots (10 μL) of each amplification reaction were separated by agarose gel electrophoresis and transferred onto nylon membranes. All RT-PCR reactions were performed in triplicate. Hybridizations were carried out with an equimolar mixture of DNA fragments specific for the *AtCBL1*, *AtCBL5*, *AtCIPK2*, and *RD25A* genes and the actin gene (Schepers and Reaay, Primers from AM Moon). For signal detection and quantification, membranes were either exposed to x-ray films or evaluated by phosphorimaging.

**ACKNOWLEDGEMENTS**

We thank Dr. Veronica Albrecht for her support during the initial two-hybrid analyses, Dr. Ralph Rock for stimulating discussions and critical reading of the manuscript, and the Arabidopsis stock centers for providing various EST clones and cDNA libraries. We also thank Drs. Russel Jones and Yong-Sic Hwang for sharing unpublished results and for helpful discussions of the rice CBL nomenclature.

Received September 8, 2003; returned for revision October 8, 2003; accepted October 20, 2003.

**LITERATURE CITED**


Chikano H, Ogawa M, Ikeda Y, Koizumi N, Kusano T, Sano H (2001) Two novel genes encoding SNI1 related protein kinases from *Arabidopsis thaliana*: differential accumulation of AtSRI1 and AtSRI2 transcripts in...
Kolukisaoglu et al.

response to cytokinins and sugars, and phosphorylation of sucrose synthase by AtS2R. Mol Gen Genet 264: 674–681


