Evolutionary and functional approach to poplar circadian clock system

2009. 3

Division of Thermo-Biosystem Relations

Specialty of Science of Cryobiosystems

United Graduate School of Agricultural Sciences

(Iwate University)

Naoki TAKATA

Table of contents

Page

Chapter 1 Introduction 1

Trees in the plant kingdom 1

Seasonal dormancy induction and cold acclimation ..... 2 in boreal and temperate trees

Circadian clock system in model plant, *Arabidopsis thaliana* 2

Circadian clock related-genes in angiosperms 4

Circadian clock system and dormancy induction in trees 6

Outline of the thesis 7

Figure 9

Chapter 2 Molecular phylogeny and expression of poplar circadian clock genes, *LHY1* and *LHY2* 10

Summary 10

Introduction 11

Materials and Methods 13

Results 16

Discussion 20

Tables and Figures 25

Chapter 3 Phylogenetic footprint of the plant clock system in angiosperms: Evolutionary processes of *Pseudo-Response Regulators* 37

Summary 37

Introduction 38

Materials and Methods 40

Results 42

Discussion 47

Table and Figures 52

Chapter 4 ***PRR1/TOC1*, a key component of the *Arabidopsis* clock system, does not show a diurnal expression pattern in poplar** **66**

Summary 66

Introduction 67

Materials and Methods 69

Results 74

Discussion 77

Table and Figures 80

Chapter 5 General discussion **86**

Evolutionary process of the plant circadian clock system 86

Circadian clock system in *Populus* 88

Chapter 6 Conclusion remarks and perspectives **92**

Acknowledgements **93**

References **95**

Chapter 1

Introduction

Trees in the plant kingdom

When we look around our garden, we can classify plants on the basis of their appearance into groups such as "trees" or "herbs". This classification is intuitive and practical but neither systematic nor standardized. When we consider the relationships between plants' traits and classification, we can see that plants termed "trees" or "herbs" are dispersed among diverse taxa. Indeed, both trees and herbs are polyphyletic and mixed in plant phylogeny (Angiosperm Phylogeny Group, 2003; Soltis *et al.*, 2005). The gap between morphological characteristics and plant classification is a longstanding mystery in the plant kingdom (Groover, 2005).

In the plant kingdom, trees are described as vascular plants that have a long life span and radial growth at the vascular cambium (Fukushima *et al.*, 2003). They are phylogenetically categorized into a wide variety of taxa within seed plants. The seed plants are comprised of 700 living species of gymnosperms and 250,000 living species of angiosperms (Heywood, 1978). In gymnosperms, most species in the extant orders (Cycadales, Ginkgoales, Gnetales, and Coniferales) are classified into trees, which are described as softwood species (Coulter & Chamberlain, 1901). Within the angiosperms, hardwood trees show a polyphyletic relationship. The current angiosperm phylogeny demonstrates that among the eudicot lineage, the conversion of trees into herbs or herbs into trees occurred within independent orders and/or families (Angiosperm Phylogeny Group, 2003; Soltis *et al.*, 2005). For example, families within two predominant groups of eudicots (rosids and asterids) include both tree and herb species; Fabales, Rosales, and Brassicales in the rosids, and Cornales, Asterales, and Solanales in the asterids. These phylogenetic footprints suggest that the histologic and genetic features of living trees have been expressed in diverse taxa of land plants during the plant speciation process, rather than simply being retained throughout the diversification and evolution of trees. The aim of the present study was to

determine the common regulatory mechanisms conserved in boreal and temperate trees throughout the process of plant speciation using evolutionary and functional analyses.

Seasonal dormancy induction and cold acclimation in boreal and temperate trees

Each year, trees in boreal and temperate zones undergo changes that enable them to cope with the extreme temperature and light conditions in winter. Trees respond to the sequential environmental signals such as shortening daylength, low non-freezing and freezing temperatures. These responses result in a change in trees' developmental status from a growing to a dormant stage, and acquisition of maximum freezing resistance (Sakai & Larcher, 1987; Weiser, 1970). Growth of boreal and temperate trees ceases during the first stage of this process, and terminal buds form in response to shortening daylength (Perry, 1971). Trees enter endodormancy after a few weeks of short-day conditions. Endodormancy is controlled by endogenous factors, and freezing tolerance slightly increases as trees enter this stage (Fuchigami *et al.*, 1971; Junttila & Kaurin, 1990; Li *et al.*, 2002; Welling *et al.* 2002). The initial freezing tolerance is rapidly enhanced by subsequent exposure to low non-freezing and freezing temperatures (Harrison *et al.*, 1978). As the freezing tolerance increases under low temperature conditions, the dormant stage simultaneously shifts from endodormancy to ecodormancy, an imposed rest under unfavorable circumstances such as low and freezing temperatures (Rinne *et al.*, 1997). A few months later, the increase in mean temperatures in spring releases ecodormancy and leads to a gradual decrease in the freezing tolerance of dormant trees (Beck *et al.*, 2004; Li *et al.*, 2003).

Circadian clock system in model plant, *Arabidopsis thaliana*

Natural daylength is a key environmental cue for plants to gauge the correct time and season in nature. To accurately measure daylength, plants have an endogenous mechanism known as the "circadian clock". Circadian clock systems regulate various biological processes, such as photomorphogenic processes, floral transition, leaf movements, stomatal conductance, photosynthetic capacity, and volatile emissions (reviewed in Yakir *et al.*, 2007). The seasonal transition from

growing to dormant stages in boreal and temperate trees is also categorized as a plant photoperiodic response (Garner & Allard, 1923). Since some of these phenomena are widely conserved among diverse groups of plants, it is assumed that plants share a basic circadian clock mechanism that functions as a master controller of photoperiodic responses.

In the last few decades, molecular genetic analyses using a model plant, *Arabidopsis thaliana*, have identified a wide variety of circadian clock-related genes (reviewed in McClung, 2006; Más, 2005). The clock-related genes are classified into several types including transcriptional factors, photoreceptors, posttranslational regulators (kinases and F-box proteins), and others. On the basis of the clock models (transcriptional feedback systems) in cyanobacteria, *Drosophila,* and mammals (reviewed in Dunlap, 1999), it is assumed that the transcriptional factors identified in *A. thaliana* make up negative-positive feedback loops that are underpinned by photoreceptors and posttranslational regulators. Recently, a computational model of the plant clock system has been developed, which contains the main transcriptional feedback loop and additional loops associated with the main loop (Fig. 1; Locke *et al.*, 2006; Zeilinger *et al.*, 2006).

The main transcriptional feedback loop ("Loop I") is composed of two single Myb genes [*Late Elongated Hypocotyl* (*LHY*) and *Circadian Clock Associated 1* (*CCA1*)], a Pseudo-Response Regulator [*Pseudo-Response Regulator 1/Timing of CAB2 Expression 1* (*PRR1/TOC1*)] and an unknown factor "X" (Fig. 1; Alabadí *et al.*, 2001; Locke *et al.*, 2006; Zeilinger *et al.*, 2006). In this loop, the morning-acting *LHY* and *CCA1* genes, which have partially redundant functions, directly repress the evening-acting *PRR1/TOC1* gene and are in turn induced by *PRR1/TOC1* through the unknown factor "X" (Alabadí *et al.*, 2001). It has been proposed that the unknown factor "X" is *Early Flowering 3* (*ELF3*), *Early Flowering 4* (*ELF4*), *Lux Arrhythmo/Phytoclock 1* (*LUX/PCL1*) and/or *Time For Coffee* (*TIC*) (McClung, 2008). *LHY* and *CCA1* entrain the light signal via the photoreceptor (phytochrome B, PhyB) and Phytochrome-Interacting Factor 3 (PIF3) signaling pathway, which results in their morning expressions (Martínez-García *et al.*, 2000). It was reported that the LHY and CCA1 proteins are posttranslationally phosphorylated by Casein Kinase 2 (CK2) to stabilize their

binding activities to the *PRR1/TOC1* promoter (Daniel *et al.*, 2004), and that the proteasome-dependent degradation of LHY and CCA1 proteins is inhibited by De-Etiolated 1 (DET1) (Song & Carré, 2005). On the other hand, the PRR1/TOC1 protein accumulates early in the night and interacts with the phosphorylated PRR3 protein, a paralogous gene of *PRR1/TOC1*, to stabilize its function (Fujiwara *et al.*, 2008). PRR1/TOC1 is gradually dissociated from PRR3 during the night, and is then posttranslationally degraded through the proteasome pathway mediated by Zeitlupe (ZTL) (Fujiwara *et al.*, 2008; Más *et al.*, 2003).

Loop III (the "morning loop") associates with the main feedback loop (*LHY/CCA1-PRR1/TOC1*-X) by sharing morning-expressed *LHY* and *CCA1* (Fig. 1; Locke *et al.*, 2006; Zeilinger *et al.*, 2006). In this loop, *LHY* and *CCA1* positively regulate *PRR7* and *PRR9* in the *PRR* gene family and are repressed by these *PRR* genes in a feedback manner (Farré *et al.*, 2005). The expression of *PRR9* is up-regulated by the light signal, resulting in its morning expression (Makino *et al.*, 2001). After *PRR9* has been expressed at peak level, *PRR7* transcripts begin to increase. PRR7 protein is degraded through the proteasome pathway during the early night (Farré & Kay, 2007).

Loop II (the "evening loop") comprises *PRR1/TOC1* and an unknown factor "Y", and is connected to the other loops via *PRR1/TOC1* (Fig. 1; Locke *et al.*, 2006; Zeilinger *et al.*, 2006). It has been proposed that the unknown factor "Y" is *Gigantea* (*GI*) and/or *PRR5*, however, the results of a recent study suggest that these two genes do not completely fulfill the role of factor "Y" (Kawamura *et al.*, 2008).

Using numerous genetic and biological analyses, the complete mechanism of the plant clock system has been deduced for *A. thaliana*. However, it is still unclear whether this complicated clock network in *A. thaliana* is conserved among clock systems in other plant species.

Circadian clock-related genes in angiosperms

Substantial genomic resources such as expression sequence tags (ESTs) and protein and genomic sequences have been accumulated for many plants. In addition, entire genomic sequences are available for some angiosperms, including *Oryza sativa* (rice; commelinids in monocots), *Sorghum bicolor* (sweet sorghum;

commelinids in monocots), *Vitis vinifera* (grape; rosids in eudicots), *Populus trichocarpa* (poplar; eurosids I in eudicots), *Carica papaya* (papaya; eurosids II in eudicots) and *A. thaliana* (thale cress; eurosids II in eudicots) (Arabidopsis Genome Initiative, 2000; Jaillon *et al.*, 2007; Ming *et al.*, 2008; Tuskan *et al.*, 2006; Yu *et al.*, 2005). Analyses of such genomic resources allow us to understand the evolutionary conservation and phylogenetic relationships of genes of interest (Arnaud *et al.*, 2007; Yang *et al.*, 2005; Yang *et al.*, 2008). Among circadian clock-related genes, comparative genomic analyses show that there are widely conserved homologs of the *Arabidopsis* clock related-genes in both eudicots and monocots. For example, the genomes of the monocot *O. sativa* and the eudicot *C. papaya* contain almost all of the known clock-related genes, although the copy number varies among different species (Ming *et al.*, 2008; Murakami *et al.*, 2007). In addition, the clock-related genes are also found in the EST database of model legumes (pea, soybean, and *Medicago*; eurosids I in eudicots) (Hecht *et al.*, 2005). These findings imply that a set of the clock-related genes that play key roles in the *Arabidopsis* clock system was established before the divergence of monocotyledonous and eudicotyledonous plants.

Most circadian clock-related genes show typical diurnal/circadian expression patterns. In *A. thaliana*, *LHY/CCA1* genes in the main loop show typical diurnal rhythmicity with peak expression around dawn, and expression of *GI*, *ELF3*, *ELF4,* and *LUX/PCL1* peaks around dusk (Doyle *et al.*, 2002; Fowler *et al.*, 1999; Hicks *et al.*, 2001; Onai & Ishiura, 2005; Schaffer *et al.*, 1998; Wang & Tobin, 1998). The five *PRR* genes in *A. thaliana* show diurnal and sequential expression patterns from dawn to dusk in the following order: *PRR9→PRR7→PRR5→ PRR3→PRR1* (Matsushika *et al.*, 2000). Interestingly, these expression patterns were also found in orthologous genes of other plant species. In the monocot *O. sativa*, *OsCCA1*, which is the ortholog of *LHY/CCA1* in *A. thaliana,* exhibits a typical morning expression, and the five *OsPRR* genes show sequential expressions in the following order: *OsPRR73* (*OsPRR37*)*→OsPRR95* (*OsPRR59*)*→OsPRR1,* similar to the *PRR* genes in *A. thaliana* (Murakami *et al.*, 2003). Furthermore, the *LHY/CCA1* genes in *Phaseolus vulgaris*, *Castanea sativa*, *Lemna gibba,* and *Lemna paucicostata* also show diurnal expressions with peak expression around dawn (Kaldis *et al.*, 2003; Miwa *et al.*, 2006; Ramos *et al.*,

2005). These expression patterns of clock-related genes are important molecular events within the plant circadian clock system. The diurnal/circadian expression of these genes in diverse plant species suggests functional conservation and common clock systems among angiosperms. However, it is still unknown how the ancestral clock system was structured in ancient angiosperms, and how it was modified and refined in each plant species through the course of diversification.

Circadian clock system and dormancy induction in trees

In boreal and temperate trees, the first step towards establishing endodormancy is triggered by the perception of short-day signals (Weiser, 1970). The light signals are perceived by photoreceptors, such as cryptochromes and phytochromes, and are subsequently entrained into the plant clock system (Martínez-García *et al.*, 2000). In the last decade, studies on transgenic *Populus* have been carried out to clarify the involvement of phytochromes in photoperiodic regulation in trees (Olsen *et al.*, 1997). In the first report, the *PhyA* gene was overexpressed in hybrid aspen, and transgenic lines did not stop growing even under short-day conditions (6-h daylength) (Olsen *et al.*, 1997). A subsequent study showed that the *PhyA* overexpressors did not form a terminal bud in short-day conditions (Mølmann *et al.*, 2005; Olsen *et al.*, 1997; Welling *et al.*, 2002). Eriksson (2000) produced *PhyA*-repressed plants by introducing a *PhyA*-antisense construct, and found that transformants established bud formation earlier than wild-type plants in response to short-day signals. These results were similar to those reported for *PhyA*-overexpressing plants. These transgenic analyses suggest that the *PhyA* gene is required for the photoperiodic control of seasonal growth cessation and bud set in *Populus*.

Recent studies further clarified that signaling components downstream from the plant clock system have roles in regulating photoperiodic responses (Böhlenius *et al.*, 2006; Ruonala *et al.*, 2008). The *CONSTANS* (*CO*) and *Flowering Locus T* (*FT*) genes are necessary for floral transition in angiosperms (reviewed in Searle & Coupland, 2004). Interestingly, the *Populus CO* and *FT* genes play a key role in regulating seasonal endodormancy transition (Böhlenius *et al.*, 2006). Overexpression of the *Populus CO* and *FT1* genes in hybrid aspen inhibited short day-induced growth cessation and bud set. Similarly, transgenic

plants in which the levels of *CO* and *FT* were down-regulated were more sensitive to shortening daylength. In an extreme case, some *FT1*-repressed lines set a terminal bud even under 16-h long-day conditions, whereas wild-type plants continued to grow. In nature, poplar ecotypes originating from various latitudes differ in their critical daylength for growth cessation. The correlation between the active phase of *CO* expression and the light period is critical in determining expression of its target *FT1* and for growth to continue. These results indicate that the *CO/FT* regulatory module is the endogenous timekeeper responsible for seasonal dormancy transition.

Together, the upstream and downstream signaling pathways of the plant circadian cock system are necessary for seasonal photoperiodisms in trees. However, there is no direct evidence showing that the endogenous clock system regulates seasonal dormancy cycles in boreal and temperate trees.

Outline of the thesis

The aim of this thesis was to uncover the molecular mechanisms underlying induction of seasonal dormancy in boreal and temperate trees. First, I focused on the circadian clock system in boreal trees that is required for seasonal photoperiodic responses and induction of endodormancy. This study examines the plant clock system in the model tree *Populus,* and reconstructs it using evolutionary and functional approaches. Evolutionary analyses revealed the evolutionary processes of the angiosperm clock system, including the biological clock in the *Populus* genus (Chapters 2 and 3). These analyses showed that key components in the *Arabidopsis* clock system are altered in *Populus*. In addition, functional approaches were used to determine whether these differences would result in an altered regulation mechanism in the *Populus* clock system (Chapter 4). These analyses showed that the regulation of expression and protein function of the *Populus PRR1/TOC1* are modified in *Populus*. However, this modification does not affect the diurnal expression patterns of most clock-related genes in *Populus*. Collectively, evolutionary and functional approaches have been used to reconstruct the fundamental mechanism of the *Populus* clock system. Although the *Populus* clock system slightly differs from the *Arabidopsis* clock system, the system might be necessary to predict daily and seasonal environmental changes

and to regulate plant photoperiodisms in *Populus*. The next step will be to determine how the *Populus* clock system reconstructed in this study regulates molecular mechanisms of seasonal dormancy induction in boreal and temperate trees.

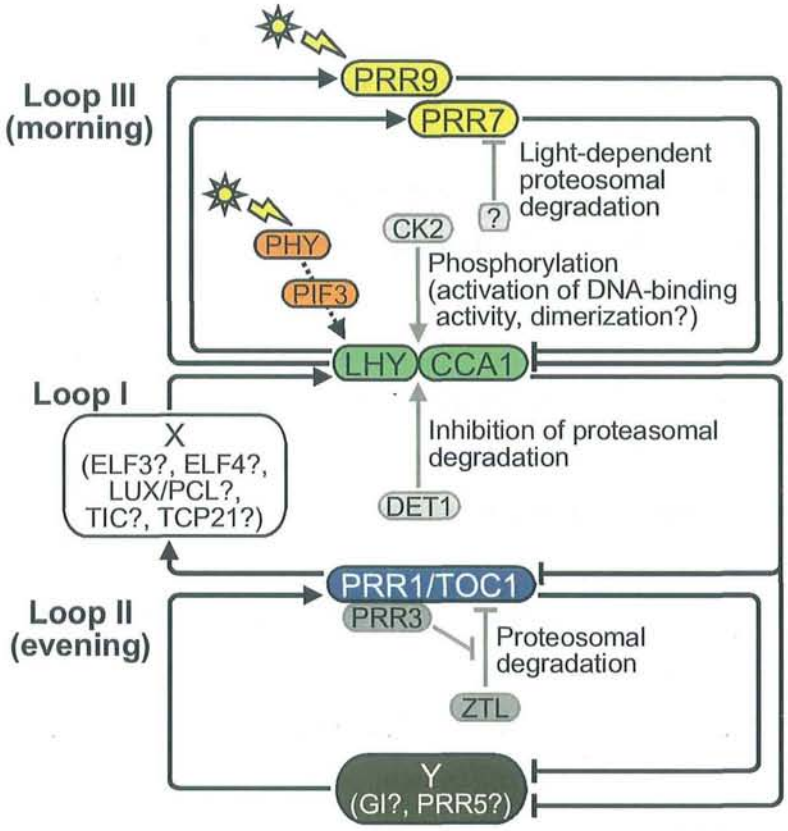


Figure 1. Schematic representation of the proposed model of the plant circadian clock. Yellow lightning indicate light input.

Chapter 2

Molecular phylogeny and expression of poplar circadian clock genes, *LHY1* and *LHY2*

Summary

• *LHY/CCA1* genes play a key role in the plant circadian clock system and are highly conserved among plant species. However, the evolutionary process of the *LHY/CCA1* gene family remains unclear in angiosperms. To obtain details of the phylogeny of these genes, this study characterized *LHY/CCA1* genes in a model woody plant, *Populus* tree.

• The evolutionary process of angiosperm *LHY/CCA1* genes was elucidated using three approaches: comparison of exon-intron structures, reconstruction of phylogenetic trees and examination of syntenic relationships. In addition, the molecular evolutionary rates and the expression patterns of *Populus LHY*s were analyzed.

• Gene duplication events of *Populus LHY*s and *Arabidopsis LHY/CCA1* had occurred independently by different chromosomal duplication events arising in each evolutionary lineage. *Populus LHY*s were under purifying selection by estimating substitution rates of these genes. Further, *Populus LHY*s conserved diurnal expressions in leaves and stems but the transcripts of *LHY2* were more abundant than those of *LHY1* in *Populus* plants.

• This study uncovered phylogenetic relationships of the *LHY/CCA1* gene family in angiosperms. In addition, the transcript abundance and the evolutionary differences between *Populus LHY1* and *LHY2* imply that *Populus LHY2* rather than *LHY1* may have a major role in the *Populus* clock system.

Introduction

Circadian rhythms, one of the most widespread phenomena in living organisms, are generated by endogenous circadian clock systems. The circadian clock systems in plants control the timing of endogenous responses under complex circumstances such as 24-h light-dark cycles and daily temperature fluctuations. Those responses include various diurnal biological processes, such as leaf movements, stomatal conductance, photosynthetic capacity and volatile emission (reviewed in Yakir *et al.*, 2007). Furthermore, plant circadian clock systems regulate long-term developmental processes, such as transition from vegetative to reproductive development and from growing to dormant stages, in response to long-period circannual changes in environmental factors (Böhlenius *et al.*, 2006; Hecht *et al.*, 2007).

In the past decade, aided by *Arabidopsis* genetics and systems biology, a wealth of information about plant clock systems has been accumulated (reviewed in McClung, 2006; Más, 2005). The plant clock system is proposed to be three transcriptional-feedback loop model (loop I, II and III) in *Arabidopsis thaliana* (Locke *et al.*, 2006; Ueda, 2006; Zeilinger *et al.*, 2006). In this model system, loop I couples together the evening oscillator (loop II) and the morning oscillator (loop III). Loop I, as the center of the three loops, consists of two morning-expressed genes, *Late Elongated Hypocotyl* (*AtLHY*) and *Circadian Clock Associated 1* (*AtCCA1*), and an evening-expressed gene, *Pseudo-response regulator 1*/*Timing of CAB2 Expression 1* (*AtPRR1*/*TOC1*). *AtLHY* and *AtCCA1* are paralogous genes and have a partial redundant function to generate robust circadian rhythms in various environments (Gould *et al.*, 2006; Schaffer *et al.*, 1998). The morning expressions of *AtLHY* and *AtCCA1* result from a direct activation by light and an indirect activation by the partner, *AtPRR1*/*TOC1* (Alabadí *et al.*, 2001; Kim *et al.*, 2003; Wang & Tobin, 1998). In addition, AtLHY/CCA1 proteins directly bind to the evening element on the promoter region of *AtPRR1*/*TOC1*, resulting in repression of its transcription during the daytime (Alabadí *et al.*, 2001). Thus, it is clear that *AtLHY* and *AtCCA1* play a key role in the entrainment of environmental cues and the regulation of the clock system itself in the main loop of the clock system.

Homologs of *Arabidopsis LHY*/*CCA1* genes are conserved not only in

eudicotyledonous but also in monocotyledonous plants. In eudicotyledonous plants, *LHY/CCA1* genes have been isolated from *Phaseolus vulgaris* (Kaldis *et al.*, 2003) and *Castanea sativa* (Ramos *et al.*, 2005) in addition to *Arabidopsis*. In monocotyledonous plants, *LHY/CCA1* genes have been isolated from *Oryza sativa* (Izawa *et al.*, 2002), *Lemna gibba* and *Lemna paucicostata* (Miwa *et al.*, 2006). The proteins encoded by these genes have a conserved Myb DNA-binding domain at their N-terminus. Furthermore, these genes exhibit rhythmicity with peak expression around dawn, which is consistent with the expression patterns of *Arabidopsis LHY* and *CCA1* (Schaffer *et al.*, 1998; Wang & Tobin, 1998). These observations imply that functions of *LHY/CCA1* genes in the plant clock system are highly conserved among angiosperm species.

Unlike the conservation of their function, the number of *LHY/CCA1* genes per genome varies in plant species. In eudicots, one copy of the *LHY/CCA1* gene is found in *P. vulgaris* (eurosids I) and two copies exist in *A. thaliana* (eurosids II) (Kaldis *et al.*, 2003; Schaffer *et al.*, 1998). In addition, two copies of *LHY/CCA1* gene are annotated as predicted genes in the available genomic sequence database of *Populus trichocarpa* (eurosids I; Tuskan *et al.*, 2006; see http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). In monocots, one copy of the *LHY/CCA1* gene has been isolated from *O. sativa* and two copies have been isolated from *L. gibba* and *L. paucicostata* (Miwa *et al.*, 2006; Murakami *et al.*, 2007). Thus, the evolutionary process of the *LHY/CCA1* gene family appears to be complicated in both eudicots and monocots.

Gene duplication can result from unequal crossing-over, retroposition, and chromosomal or whole genome duplications (reviewed in Zhang, 2003). In the genome of *A. thaliana*, three polyploidy events (so-called α, β and γ) are assumed to have occurred in angiosperm evolutionary lineages (Blanc & Wolfe, 2004; Bowers *et al.*, 2003; De Bodt *et al.*, 2005). Although the correct timing of α and β polyploidy events had been under dispute, recent completion of the draft genome sequence of *Carica papaya* revealed that these polyploidy events had arisen after divergence of *Arabidopsis* and *Carica* in eurosids II (Tang *et al.*, 2008). On the other hand, the γ polyploidy event is believed to have occurred in eudicot lineages after divergence of monocots and eudicots, although the correct timing is still under debate (Jaillon *et al.*, 2007). In the *Populus* lineage of eurosids I, the

Salicoid polyploidy event occurred within Salicaceae after divergence between Fabales and Malpighiales, (Tuskan *et al.*, 2006). Since the conservation of gene orders on the duplicated chromosomes results from the chromosomal duplication events (Adams & Wendel, 2005), comparisons of the gene orders around duplicated genes provide molecular evolutionary information for understanding the phylogenetic relationships (Bocock *et al.*, 2008; Sampedro *et al.*, 2005).

To clarify the evolutionary relationships of angiosperm *LHY/CCA1* genes, here I first isolated two full-length *LHY/CCA1* genes from the genus *Populus* and characterized their genomic structures. I next compared the exon-intron structures of *LHY/CCA1* genes, reconstructed a phylogenetic tree using sequence data of angiosperm *LHY/CCA1* genes and examined syntenic relationships in the neighboring regions of *LHY/CCA1* genes across plant species. Furthermore, I analyzed the molecular evolutionary rates, the diurnal expression patterns and the expression levels of these genes to verify the biological function of two *LHY* genes in *Populus* plants. This study uncovered not only the evolutionary processes of the *LHY/CCA1* gene family in angiosperms but also the differential expression patterns of two *LHY* genes in *Populus* plants.

Materials and Methods

Plant material

Poplar (*Populus nigra* var. *italica*) plants were grown aseptically in agar medium containing Murashige and Skoog basal salt (Murashige & Skoog, 1962), Murashige and Skoog vitamin, 20 mM MES-KOH (pH 5.8), 0.5 $mg \cdot l^{-1}$ indole-3-butyric acid, 3% (w/v) sucrose and 0.8% (w/v) agar at 22°C under 16 h-light /8 h-dark conditions (100 $\mu mol \cdot m^{-2} \cdot s^{-1}$).

Isolation of full-length cDNA

The *Populus LHY/CCA1* genes were isolated from a full-length enriched cDNA library constructed from mRNA of *P. nigra* (Nanjo *et al.*, 2007). The *LHY/CCA1* genes were subjected to dideoxy-nucleotide sequencing using a primer walking method and nucleotide sequences were assembled by ATSQ software (Genetyx, Tokyo, Japan).

Isolation of genomic DNA encoding LHY*s*

To determine the exon-intron boundaries of *Populus LHY*s, the full-length genomic regions were isolated from *P. nigra* genomic DNA using PCR. Genomic DNA was extracted by the CTAB (hexadecyltrimethyl-ammonium bromide) method from the leaves of 1-month-old poplar plants maintained on agar medium (Murray & Thompson, 1980). PCR was performed by Takara LA Taq® polymerase (Takara Bio, Shiga, Japan) according to the manufacturer's instructions using the primer sets, 5'-TTGGCTTTCTCTTCTCACTGCC-3' and 5'-CCATGCAAGGCCAATTCAATAC-3' for *PnLHY1* and 5'-GATGGAGTGTG-TCTAACTGGT-3' and 5'-CCGTGGAAGGCCAATTCAATACT-3' for *PnLHY2*. The PCR condition was: 94°C for 1 min; 30 cycles of 98°C for 15 sec, 68°C for 10 min; and 72°C for 10 min. The amplified PCR fragments were subcloned into pGEM®-T Easy vector (Promega, Madison, WI, USA) and subjected to dideoxy-nucleotide sequencing using a primer walking method. The nucleotide sequences were assembled by ATSQ software (Genetyx, Tokyo, Japan).

Phylogenetic analysis

Amino acid sequences were deduced from cDNA sequences of *LHY/CCA1* genes and aligned using the ClustalW program. The numbers of amino acid substitutions between each pair of LHY/CCA1 proteins were estimated by the Jones-Taylor-Thornton (JTT) model (Jones *et al.*, 1992) with the complete-deletion option. From estimated numbers of amino acid substitutions, a phylogenetic tree was reconstructed by the neighbor-joining (NJ) method (Saitou & Nei, 1987). The bootstrap values were calculated with 1,000 replications (Felsenstein, 1985). These procedures were performed using MEGA4 software (http:// www.megasoftware.net/index.html) (Tamura *et al.*, 2007). A phylogenetic tree was also reconstructed by the Maximum Likelihood (ML) method using PhyML (http://atgc.lirmm.fr/phyml/) (Guindon *et al.*, 2005) applying the JTT model for amino acid substitutions. The bootstrap values for this phylogenetic tree were calculated with 100 replications. Rates of nonsynonymous (d_N) and synonymous (d_S) substitutions were calculated by the modified Nei-Gojobori method with the transition/transversition ratio equal to 1.2 and Jukes-Cantor correction. This analysis was performed using MEGA4 software (Tamura *et al.*, 2007).

Boxshade program (http://www.ch.embnet.org/software/BOX_form.html) was used to highlight the conserved amino acid residues in the aligned amino acid sequences encoded by *LHY/CCA1* genes.

Identification of chromosomal synteny

I searched the syntenic relationship in *Arabidopsis* or *Oryza* genomes using the genomic regions containing *AtLHY* and *AtCCA1* or *OsCCA1* on the basis of previous studies (Blanc & Wolfe, 2004; Bowers *et al.*, 2003; Salse *et al.*, 2008; Yu *et al.*, 2005).

To identify the syntenic relationship of the genomic regions containing *PtLHY1* and *PtLHY2* in the *Populus* genome (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), I first performed a bi-directional TBLASTN search against the genomic sequence of *Populus* using predicted genes located in neighboring regions of the *PtLHY1* and *PtLHY2* genes as queries. If a gene(s) showing high sequence similarity to a query sequence (E-value lower than 10^{-5}) was found near the *LHY/CCA1* gene, the TBLASTN search was further performed in the reverse direction using the best-hit gene as a query. When the best-hit gene in the second TBLASTN search showed high similarity (E-value lower than 10^{-5}) to the gene used as the query in the first TBLASTN search, I considered this gene pair as orthologs.

The syntenic relationships across *Populus*, *Arabidopsis* and *Oryza* genomes were examined as described above using the databases of *Arabidopsis* (MIPS *Arabidopsis thaliana* genome database, http://mips.gsf.de/proj/plant/jsf/athal/index.jsp) and *Oryza* (Rice Annotation Project Database, http://rapdb.dna.affrc.go.jp/).

RNA extraction and real-time PCR

To investigate the expression pattern of *LHY/CCA1* genes in poplar, leaves and stems of *P. nigra* growing on the campus of Iwate University (Morioka, Japan) were collected at three-hour intervals from 9:00 AM on July 7, 2006 to 9:00 AM on July 9, 2006. The samples were immediately frozen in liquid nitrogen and stored at -80°C until use. During sampling, the natural mean day length and temperature were 14 h 52 min and 20.3°C, respectively.

Total RNA was isolated from samples using a NucleoSpin® RNA Plant kit (MACHEREY-NAGEL, Düren, Germany) with in-column DNase I digestion. First-strand cDNA was synthesized using ReverTra Ace-α® (TOYOBO, Osaka, Japan) according to the manufacturer's instructions. Real-time PCRs were performed using a Thermal Cycler Dice® Real Time System (Takara Bio, Shiga Japan) according to the manufacturer's instructions. The gene-specific primers for real-time PCR were: *PnLHY1* (forward, 5'-GTGAGTTTTCATGTGAGTTTCCG-G-3'; reverse, 5'-CTACCAATAAGCCGTCGTCTTG-3'), *PnLHY2* (forward, 5'-CTCCATTGAGCTGCCTGAAACA-3'; reverse, 5'-CGACCGCATAGACTCCA-ATTC-3') and *ubiquitin 11* (*UBQ*) (forward, 5'-GGTTGATTTTTGCTGGGAAGC -3'; reverse, 5'-GATCTTGGCCTTCACGTTGT-3'). The *UBQ* gene was used as a normalization control. Each RNA sample was assayed in triplicate. RNAs were assayed from two biological replicates.

Real-time PCRs were also performed to examine expression levels of *PnLHY1* and *PnLHY2*. cDNA fragments of *PnLHY1*, *PnLHY2* and *UBQ* were amplified using Takara Ex Taq® polymerase (Takara Bio, Shiga, Japan) by the gene-specific primer pairs shown above. The amplified fragment of *PnLHY1* was subcloned into pGEM®-T Easy vector (Promega, Madison, WI, USA) and those of *PnLHY2* and *UBQ* were subcloned into pTAC-1 vector (Biodynamics Laboratory, Tokyo, Japan). These fragments were subjected to dideoxy-nucleotide sequencing. The vectors containing *PnLHY2* and *UBQ* were digested with *Sal*I and *Xho*I to cut out DNA fragments containing *PnLHY2* and *UBQ*, which were subsequently introduced into the *Sal*I site of pGEM®-T Easy vector harboring *PnLHY1*. The pGEM-T easy vector containing the fragments of *PnLHY1*, *PnLHY2* and *UBQ* was used to generate a standard curve of real-time PCR amplification. Transcript levels of *PnLHY1* and *PnLHY2* were normalized with that of *UBQ*. Each RNA sample was assayed in triplicate. RNAs were assayed from two biological replicates.

Results

Characterization of Populus LHY/CCA1 *genes*

In the genomic sequence database of *P. trichocarpa*, two *LHY/CCA1* homologs, *PtLHY1* and *PtLHY2*, were predicted (Table 2-1). With this information, cDNA

clones of two *LHY/CCA1* genes (*PnLHY1* and *PnLHY2*) were isolated from a full-length enriched cDNA library of *P. nigra* (Nanjo *et al.*, 2007). *PnLHY1* and *PnLHY2* exhibited 97.0% and 96.1% homology at nucleotide sequence level to predicted *PtLHY1* and *PtLHY2*, respectively. Although *PtLHY1* has been annotated as two partial *LHY/CCA1* genes in its genomic region, I confirmed that a single, complete *LHY/CCA1* gene was encoded in this region.

Exon-intron structures of LHY/CCA1 *genes*

Next, I determined the exon-intron boundaries of *PnLHY1* and *PnLHY2* on the genome of *P. nigra*. Both *PnLHY1* and *PnLHY2* contained seven introns and eight exons, with the coding sequences (CDSs) lying from the third to eighth exons (designed as coding regions I to IV; Fig. 2-1a). Comparison of the exon-intron structures of *LHY/CCA1* genes among angiosperm species revealed that the CDS of *OsCCA1* consist of six exons similar to *PnLHY*s (Fig. 2-1a). In contrast, both *AtLHY* and *AtCCA1* have an additional intron within the predicted coding region V of *PnLHY*s and *OsCCA1* (Fig. 2-1a).

Further comparisons of *PnLHYs*, *AtLHY/AtCCA1*, and *OsCCA1* at the amino acid sequence level revealed that these genes have a conserved Myb DNA-binding domain at the N-terminal region (Fig. 2-1b). In addition, their exon-intron boundaries were conserved among *PnLHY*s, *AtLHY/AtCCA1* and *OsCCA1*, except for the sites of the additional intron of *AtLHY/AtCCA1*. The nucleotide lengths of coding regions I, II and III, which encode the Myb DNA-binding domain, were the same in all three species (39, 112 and 62 base pairs, respectively; Fig. 2-1a).

Phylogenetic analysis of LHY/CCA1 *gene family in angiosperms*

To infer evolutionary relationships of angiosperm *LHY/CCA1* genes, a phylogenetic trees were constructed with the NJ and ML methods using 12 genes from monocots, *O. sativa* and *Sorghum bicolor*; core eudicots, *Mesembryanthemum crystallinum*; rosids, *Vitis vinifera*; eurosids I, *Phaseolus vulgaris*, *Castanea sativa*, *P. nigra* and *P. trichocarpa*; and eurosids II, *A. thaliana* (Table 2-2). Among these genes, *OsCCA1* and *SbCCA1* were used as an outgroup of the phylogenetic tree to place a root because the divergence between monocots

and eudicots has been established from various studies (Angiosperm Phylogeny Group, 2003; Soltis *et al.*, 2005).

The reconstructed phylogenetic trees confirmed that the topologies of these trees obtained by the two different tree building methods were essentially the same (Fig. 2-2). The phylogenetic tree revealed that *Arabidopsis CCA1* and *LHY* were distantly related to each other and diverged earlier from remaining eudicotyledonous *LHY/CCA1* genes. In addition, eudicotyledonous *LHY/CCA1* genes were more closely related to *Arabidopsis LHY* than to *Arabidopsis CCA1*. On the other hand, *Populus LHY1* and *LHY2* were more closely related than other *LHY/CCA1* genes, indicating that the gene duplication event that produced *Populus LHY1* and *LHY2* occurred after the divergence of *Populus* and the other eurosids I (*P. vulgaris* and *C. sativa*). Consequently, the topology of the phylogenetic tree implies that the duplication event of *Arabidopsis LHY/CCA1* does not coincide with that of *Populus LHY*s.

Chromosome syntenies among the genomes of Populus, Arabidopsis *and* Oryza

To obtain further information on evolutionary relationships of angiosperm *LHY/CCA1* genes, I next investigated chromosomal syntenies of *P. trichocarpa*, *A. thaliana* and *O. sativa* by examining the physical positions of the orthologous genes surrounding the *LHY/CCA1* genes. In the *Arabidopsis* genome, *LHY* and *CCA1* are located on chromosomes 1 and 2, respectively (Murakami *et al.*, 2007). The neighboring genes of *LHY* shared a syntenic relationship with those of *CCA1* (Fig. 2-3; Blanc & Wolfe, 2004; Bowers *et al.*, 2003). Furthermore, the flanking region of *LHY* and *CCA1* retained a chromosomal synteny with a partial region of chromosomes 4 and 3, respectively. However, the regions of these two chromosomes did not contain *LHY/CCA1* genes. In the *Populus* genome, *LHY1* and *LHY2* of *P. trichocarpa* are located on chromosomes 2 and 14, respectively (Table 2-1). Comparison of the gene organizations around *PtLHY1* and *PtLHY2* revealed that the flanking region of *PtLHY1* showed chromosomal synteny with that of *PtLHY2* (Fig. 2-3). Furthermore, the physical positions of the orthologous genes surrounding *Populus LHY1/LHY2* and *Arabidopsis LHY/CCA1* were relatively well conserved across plant species due to integration of the chromosomal syntenies within *Populus* or *Arabidopsis* genomes. These results

suggest that gene duplications of *Arabidopsis LHY/CCA1* and *Populus LHY*s have derived from ancient chromosomal duplication events.

The gene organizations surrounding *Populus LHY1/LHY2* and *Arabidopsis LHY/CCA1* were next compared against the *Oryza* genomic sequence. *OsCCA1* is located on chromosome 8 (Murakami *et al.*, 2007). I found that four genes located adjacent to *OsCCA1* showed homologous relationships to the genes that resided near the *LHY/CCA1* genes in the *Populus* and *Arabidopsis* chromosomes (Fig. 2-3). However, I did not find an extensive syntenic relationship around the *LHY/CCA1* genes between the *Oryza* and *Populus/Arabidopsis* genomes. In addition, the neighboring genes of *OsCCA1* had no syntenic regions with other chromosomes in the *Oryza* genome (Salse *et al.*, 2008; Yu *et al.*, 2005).

Molecular evolutionary rates of Populus LHY*s*

To examine the selection forces of *Populus LHY1/LHY2* and the other eurosids *LHY/CCA1* genes, I estimated ratios of nucleotide substitution rate in non-synonymous (d_N) versus synonymous (d_S) mutations between *LHY/CCA1* of rosids (*V. vinifera*) and eurosids (eurosids I, *P. nigra*, *C. sativa* and *P. vulgaris*; eurosids II, *A. thaliana*; Table 2-2). The d_N/d_S ratios of these genes were smaller than 0.4 and similar among *LHY/CCA1*s of eurosids (Table 2-3). This result implies that not only *Populus LHY*s but also other *LHY/CCA1* genes in eurosids are under purifying selection.

Expression patterns of Populus LHY*s*

To reveal the functional conservation of *Populus LHY*s in a clock system, I next determined expression patterns of *PnLHY1* and *PnLHY2* under field conditions in summer using real-time PCR. *PnLHY1* and *PnLHY2* showed typical diurnal expressions both in leaves and stems of *P. nigra* (Fig. 2-4). The transcripts of these genes began to increase gradually at midnight and reached peaks of diurnal rhythms around dawn.

Subsequently, I carried out quantitative analysis of *PnLHY1* and *PnLHY2* expressions around their peak expression time (at 9:00 AM on July 8, 2006) to determine whether there are differences in the expression levels of these genes. Interestingly, the transcripts of *PnLHY2* were at least 5 times more abundant than

those of *PnLHY1* in both leaf and stem tissues (Fig. 2-5). Furthermore, the expression levels of *PnLHY*s were 5 to 7-fold higher in leaves than in stems. These findings are consistent with expression data in the poplar eFP Browser that *Populus LHY2* is expressed in more abundance than *LHY1* in all tissues (mature leaf, young leaf, root, dark-grown seedling, continuous light-grown seedling, female catkins, male catkins and xylem) of *Populus* plants (Fig. 2-6). Thus, these results suggest that *Populus LHY2* but not *LHY1* is the predominant gene expressed in *Populus* plants.

Discussion

It is well known that the *LHY/CCA1* gene family plays a key role in the angiosperm circadian clock system (reviewed in Yakir *et al.*, 2007). However, the phylogenetic relationship of the *LHY/CCA1* gene family among eudicots and monocots remains to be determined. In the present study, I isolated two *LHY*s from the poplar tree and then elucidated the evolutionary process of the *LHY/CCA1* genes in angiosperms using three approaches: comparison of exon-intron structures, conventional phylogenetic reconstruction, and examination of syntenic relationships.

In eudicots (*Populus* and *Arabidopsis*) and monocots (*Oryza*), the exon-intron structures of their *LHY/CCA1* genes were well conserved within their CDSs (Fig. 2-1a). Five exon-intron boundaries were shared in all of the *LHY/CCA1* genes examined (Fig. 2-1b). The similarity in the exon-intron organization of *Populus LHY*s, *Arabidopsis LHY/CCA1* and *Oryza CCA1* implies that there is a common ancestral gene of *LHY/CCA1* in eudicots and monocots.

Phylogenetic trees of angiosperm *LHY/CCA1* genes reconstructed by the NJ and ML methods exhibited a distant relationship between *Arabidopsis LHY/CCA1* and other eudicotyledonous genes (Fig. 2-2). This evolutionary relationship in regard to the divergence of *Arabidopsis LHY* and *CCA1* has been shown in previous studies (Boxall *et al.*, 2005; Miwa *et al.*, 2006). In this study, by reconstructing the phylogenetic tree, I revealed a close relationship of *Populus LHY1* and *LHY2*. These results indicate that gene duplication of *Arabidopsis LHY* and *CCA1* would not coincide with that of *Populus LHY1* and *LHY2*.

To obtain further details of the phylogenetic relationships of *LHY/CCA1* genes

in angiosperms, I analyzed the chromosomal syntenies among three model plants, *P. trichocarpa*, *A. thaliana* and *O. sativa*. Chromosomal syntenies that are conserved across plant species are a powerful tool for studying the evolutionary process of a gene family (Bocock *et al.*, 2008; Sampedro *et al.*, 2005). In the *Arabidopsis* genome, the flanking regions of *AtLHY* and *AtCCA1* retained the synteny that was derived from the β polyploidy event (Figs. 2-3 and 2-7; Bowers *et al.*, 2003; described as "old" duplication in Blanc *et al.*, 2003; Blanc & Wolfe, 2004; De Bodt *et al.*, 2005). In addition, completion of the draft genome sequence of *C. papaya* suggests that the β polyploidy event would have taken place after divergence of *Arabidopsis* and *Carica* within Brassicales of eurosids II (Tang *et al.*, 2008). Therefore, I speculate that the ancestral *LHY/CCA1* gene was duplicated into *LHY* and *CCA1* in the lineage leading to *Arabidopsis* but not in that leading to *Carica*. This evolutionary footprint in Brassicales is consistent with the results of a recent study that *Carica* retains only one copy of *LHY/CCA1* gene in its genome (Ming *et al.*, 2008). The results in the present study also demonstrate that other chromosomal syntenies were found in *Arabidopsis* between chromosomes 1 and 4 and between chromosomes 2 and 3 (Fig. 2-3). These syntenies are assigned to the α polyploidy event that have arisen after the β polyploidy event (Bowers *et al.*, 2003; described as "recent" duplication in Blanc *et al.*, 2003; De Bodt *et al.*, 2005). However, since there are no *LHY/CCA1* genes within the syntenic regions of chromosomes 3 and 4, duplicated *LHY* and *CCA1* produced by the α polyploidy event may have been lost from the ancient *Arabidopsis* genome during the evolutionary process (Fig. 2-7).

The conserved syntenic relationships within the *Populus* genome lead us to hypothesize that *Populus LHY1* and *LHY2* were duplicated in the Salicoid duplication event that is believed to have occurred after the divergence of Fabaceae and Salicaceae within eurosids I (Figs. 2-3 and 2-7; Sterck *et al.*, 2005; Tuskan *et al.*, 2006). This hypothesis is consistent with the topology of the phylogenetic trees; the gene duplication of *Populus LHY*s occurred in a lineage of *Populus* (Fig. 2-2). Furthermore, the syntenic relationships of the *Populus* chromosomes are shared with four *Arabidopsis* chromosomes (Fig. 2-3). The present study showed that the chromosome duplication events that produced the syntenic relationships in the *Populus* and *Arabidopsis* genomes had occurred

independently in each lineage (Fig. 2-7). Thus, I propose that a common ancestral *LHY/CCA1* gene of eurosids I and II had been located on the common ancestral chromosome that was subsequently duplicated into two *Populus* chromosomes and four *Arabidopsis* chromosomes.

The evolutionary process of angiosperm *LHY/CCA1* genes that was deduced by the syntenic relationships differs from the topology of the phylogenetic tree with regard to the timing of the gene duplication of *Arabidopsis LHY/CCA1* (Figs. 2-2 and 2-7). Although the syntenic relationships indicated that the ancestral *LHY/CCA1* gene was duplicated into *LHY* and *CCA1* after divergence of *Arabidopsis* and *Carica* in eurosids II, the phylogenetic trees implied that the gene duplication had occurred before the divergence of eurosids I and II. Several studies have recently shown that the substitution rate among paralogous genes was accelerated in the *Arabidopsis* genome compared with the *Populus* genome, which could affect reconstruction of the phylogenetic tree (Tuskan *et al.*, 2006; Van de Peer *et al.*, 1996). Estimation of the synonymous substitution rates in the *LHY/CCA1* genes in rosids (*V. vinifera*) and eurosids (eurosids I, *P. nigra*, *C. sativa* and *P. vulgaris*; eurosids II, *A. thaliana*) indicated that *Arabidopsis LHY* and *CCA1* had a higher synonymous substitution rate than that of *LHY/CCA1* genes in eurosids I (d_S values in Table 2-3). Thus, the difference in synonymous substitution rates would affect the topology of the phylogenetic tree, resulting in inconsistency in the timing of duplication events of *Arabidopsis LHY* and *CCA1* genes estimated by phylogenetic tree and syntenic relationships.

Unfortunately, the evolutionary process of the *LHY/CCA1* gene family in monocots still remains unclear. The chromosome synteny analyses revealed a reduced level of conserved synteny in the flanking regions of the *LHY/CCA1* genes between the *Oryza* and eurosids genomes (Fig. 2-3). A reduced level of conserved synteny between the *Arabidopsis* and the *Oryza* genomes has been shown previously in genome-wide surveys (Salse *et al.*, 2002; Vandepoele *et al.*, 2002). Two *Lemna* plants have two *LHY/CCA1* genes (*LHYH1* and *LHYH2*) that show typical morning expressions (Miwa *et al.*, 2006). A preliminary analysis indicated that the gene duplication event of *Lemna LHYH*s did not coincide with that of *Arabidopsis LHY/CCA1* and *Populus LHY*s (data not shown). This is because *Arabidopsis LHY/CCA1* and *Populus LHY*s were derived from the β

polyploidy and Salicoid polyploidy events, respectively, and both polyploidy events occurred only in eudicot lineages (Bowers *et al.*, 2003; Blanc & Wolfe, 2004; Tuskan *et al.*, 2006). Furthermore, phylogenetic analysis suggests that the gene duplication of *Lemna LHYH*s occurred in a common ancestor of *Oryza* and *Lemna* (Miwa *et al.*, 2006; unpublished data).

Morning expressions of *LHY/CCA1* genes have been observed in many plant species (Kaldis *et al.*, 2003; Izawa *et al.*, 2002; Miwa *et al.*, 2006; Ramos *et al.*, 2005; Schaffer *et al.*, 1998; Wang & Tobin, 1998). Both *Populus LHY1* and *LHY2* also showed typical diurnal rhythmicity with peak expression around dawn with a slight difference (Fig. 2-4). Intriguingly, however, the results of the present study and the available microarray data showed that the amount of *Populus LHY2* transcripts was higher than that of *LHY1* in *Populus* plants (Figs. 2-5 and 2-6). It has been reported that some duplicated genes in the Salicoid polyploidy event showed asymmetric expressions as with *Populus LHY1/LHY2* (Arnaud *et al.*, 2007; Oakley *et al.*, 2007; Rajinikanth *et al.*, 2007; Tuskan *et al.*, 2006). Furthermore, asymmetric expressions of duplicated genes are observed not only in plants but also in mammals (Chung *et al.*, 2006; Ganko *et al.*, 2007). Because these expression differences indicate evolutionary divergences of duplicated genes, the difference in transcript level between *Populus LHY1* and *LHY2* may imply a functional diversity of these genes in *Populus* plants.

Alterations of a promoter region can contribute to differential expression of duplicated genes (reviewed in Zhang, 2003). In the promoter regions of *PnLHY1* and *PnLHY2*, a difference in the composition of cis-regulatory elements was found. This is presumably due to nucleotide substitutions, insertions and deletions (Fig. 2-8). It has been reported that the duplicated genes derived from the Salicoid polyploidy event showed differential expression patterns in *Populus* plants depending on the evolutionary changes of their promoter regions (Ohmiya *et al.*, 2003; Ohmiya, personal communication). Thus, it is possible that differences in the promoter regions of *Populus LHY*s may affect the transcript levels between *Populus LHY1* and *LHY2*.

In addition, diversity of duplicated genes can be resulted from alterations of a protein coding region (reviewed in Zhang, 2003). Both of *Populus LHY1* and *LHY2* genes are assumed to be functional in *Populus* plants because these genes

are under purifying selection (Table 2-3). However, the analysis of specific phosphorylation sites that would be required for the physical function of LHY/CCA1 protein in angiosperms (Daniel *et al.*, 2004) revealed that the serine residue located at the upstream of the Myb DNA-binding domain is highly conserved among angiosperm LHY/CCA1 proteins except for the *Populus* LHY1 proteins (Fig. 2-9). Wang *et al.* (2005) have proposed that, in some genes, the loss of a phosphorylation site in a protein would contribute to acquisition of a modified function of the protein during gene and species evolution. Thus, the mutation of the phosphorylation site in the *Populus* LHY1 protein may result in the divergence of these duplicated genes in poplar tree.

Collectively, alterations of protein coding regions and promoter regions of *Populus LHY1* and *Populus LHY2* indicate that these genes have been subjected to a different evolutionary fate after gene duplication. The conservation of phosphorylation sites and higher expression of *Populus LHY2* may suggest a major role of this gene in the *Populus* clock system. However, it is not clear at the moment whether *Populus LHY1* plays a role that is redundant with or different from *Populus LHY2* in *Populus* tree.

In summary, I have demonstrated that *Populus* has two *LHY*s produced by the Salicoid polyploidy event and that the two genes share a common ancestor with *Arabidopsis LHY* and *CCA1*. The Salicoid polyploidy event affected nearly 92% of the *Populus* genome and nearly 8,000 pairs of the Salicoid duplicated genes were identified out of 45,555 genes predicted in the present *Populus* genome (Tuskan *et al.*, 2006). They also revealed that the Salicoid duplicated genes were under purifying selection, which is similar to results for *LHY1* and *LHY2* in the present study (Table 2-3). On the other hand, they found using whole-genome microarray analyses that 5% of duplicated genes from the Salicoid polyploidy event (nearly 400 pairs of genes) showed differential expression patterns in *Populus* plants (Tuskan *et al.*, 2006). The present study also elucidated that the transcripts of *LHY2* were more abundant than those of *LHY1* in *Populus* plants but, interestingly, both *LHY*s conserved typical diurnal expressions in leaf and stem tissues (Figs. 2-4 and 2-5). Further studies are clearly needed to understand functional differences or redundancies between *LHY1* and *LHY2* in *Populus* plants.

Table 2-1 Genomic location of *LHY/CCA1*-like genes in the genus *Populus*

Populus nigra			*Populus trichocarpa*			
Gene name	Accession number	Nucleotide homology with *P. trichocarpa* ortholog (%)	Gene name	Predicted gene ID	Chromosome number	Location
PnLHY1	AB429410	97.0	*PtLHY1*	eugene3.00021682*, eugene3.00021683*	2	14176675-14185271
PnLHY2	AB429411	96.1	*PtLHY2*	estExt_Genewise1_v1.C_LG_XIV1950*	14	4344378-4351655

*Predicted gene IDs were obtained from JGI (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html).

Table 2-2 *LHY/CCA1*-like genes used in the phylogenetic analysis

Gene	Species	Classification[a]	Accession number/Gene ID
OsCCA1	*Oryza sativa*	Monocots	Os08g0157600[b]
*SbCCA1**	*Sorghum bicolor*	Monocots	Sbi_0.39254[c], TA26762_4558[d], TA31430_4558[d]
McCCA1	*Mesembryanthemum crystallinum*	eudicots/core eudicots	AY371287
VvLHY	*Vitis vinifera*	eudicots/rosids	GSVIVT00026185001[e]
CsLHY	*Castanea sativa*	eudicots/eurosids I	AY611029
PvLHY	*Phaseolus vulgaris*	eudicots/eurosids I	AJ420902
PnLHY1	*Populus nigra*	eudicots/eurosids I	AB429410
PnLHY2			AB429411
*PtLHY1**	*Populus trichocarpa*	eudicots/eurosids I	eugene3.00021682[f], eugene3.00021683[f]
PtLHY2			estExt_Genewise1_v1.C_LG_XIV1950[f]
LHY	*Arabidopsis thaliana*	eudicots/eurosids II	At1g01060[g]
CCA1			At2g46830[g]

[a]Plant classification refers to APGII (http://www.mobot.org/MOBOT/Research/APweb/welcome.html).
[b]Gene ID obtained from RAP-DB (http://rapdb.dna.affrc.go.jp/).
[c]Gene ID obtained from phytozome (http://www.phytozome.net/index.php).
[d]Gene IDs obtained from TIGR Plant Transcript Assemblies (http://plantta.tigr.org/).
[e]Gene ID obtained from Grape Genome Browser (http://www.genoscope.cns.fr/externe/English/Projets/Projet_ML/index.html).
[f] Gene IDs obtained from JGI (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html).
[g]Gene IDs obtained from TIGR (http://www.tigr.org/tdb/e2k1/ath1/).
*Genes that seemed to be inaccurately annotated were manually modified in this study.

Table 2-3 Rates of nonsynonymous and synonymous substitutions among eudicot *LHY/CCA1* genes

	Eurosids I				Eurosids II	
	P. nigra LHY1	*P. nigra LHY2*	*C. sativa LHY*	*P. vulgaris LHY*	*A. thaliana LHY*	*A. thaliana CCA1*
	d_N/d_S	d_N/d_S	d_N/d_S	d_N/d_S	d_N/d_S	d_N/d_S
Rosids	0.19/0.639	0.2/0.628	0.172/0.524	0.216/0.732	0.38/1.138	0.473/1.227
V. vinifera LHY	(0.297)	(0.319)	(0.328)	(0.295)	(0.334)	(0.385)

Rates of nonsynonymous (d_N) and synonymous (d_S) substitutions were calculated by the modified Nei-Gojobori method with the transition/ transversition ratio equal to 1.2 and Jukes-Cantor correction. d_N/d_S ratios are indicated in parentheses.

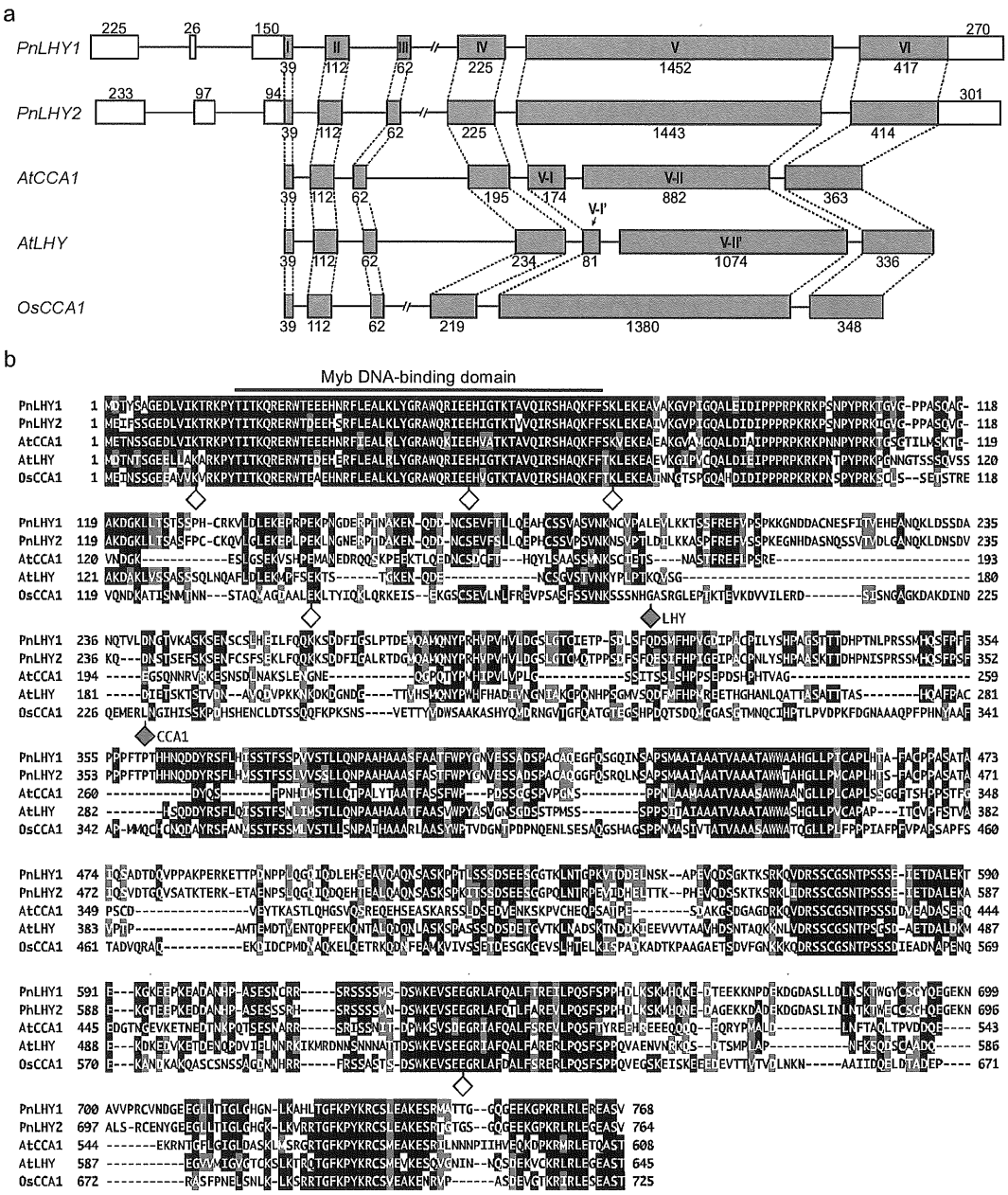


Figure 2-1. Comparison of the exon-intron structures of *LHY/CCA1* genes. (a) Schematic diagram of the exon-intron structures of *PnLHY1*, *PnLHY2*, *AtLHY*, *AtCCA1* and *OsCCA1*. White and gray boxes indicate exons for UTRs and CDSs, respectively. Coding regions I-VI are shown in exons 3-8. Lengths of nucleotide sequences are indicated above and below (above: UTR, below: exon). (b) Alignments of the amino acid sequences encoded by *PnLHY1*, *PnLHY2*, *AtLHY*, *AtCCA1* and *OsCCA1*. Amino acid sequences were aligned using the ClustalW program. Identical and similar amino acid residues are highlighted with black and gray backgrounds, respectively. White and gray diamonds indicate exon-intron boundaries shared in all genes and exon-intron boundaries found specifically in *AtLHY* and *AtCCA1*, respectively.

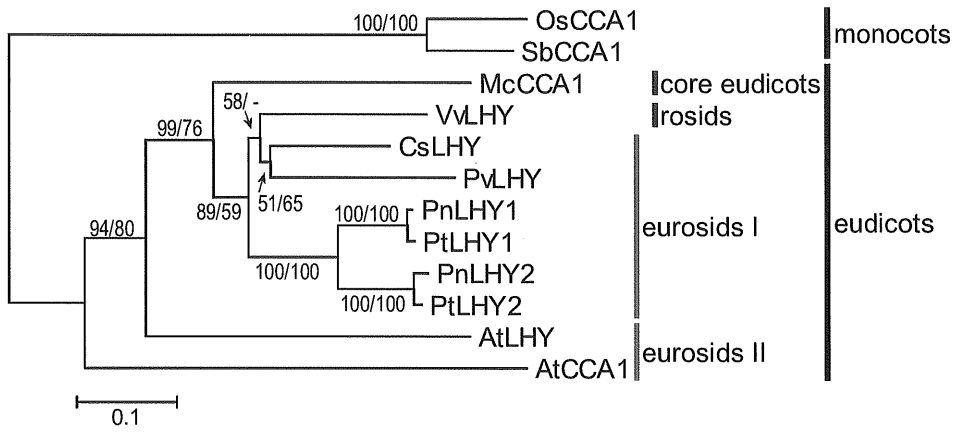


Figure 2-2. Phylogenetic tree of *LHY/CCA1* genes in angiosperms. Full-length amino acid sequences were aligned using the ClustalW program. A phylogenetic tree was reconstructed by the NJ method from numbers of amino acid substitutions estimated by applying the JTT model. A phylogenetic tree was also reconstructed by the ML method, and the tree showed the same topology as that of the tree obtained by the NJ method. The numerals at the branch indicate bootstrap values calculated by the NJ method with 1,000 replications (left) and by the ML method with 100 replications (right). Bootstrap values >50% are shown. To place a root on the phylogenetic tree, LHY/CCA1 proteins in monocots (OsCCA1 and SbCCA1) were utilized as an outgroup. Accession numbers of *LHY/CCA1* genes used in the phylogenetic analysis are shown in Table 2-2.

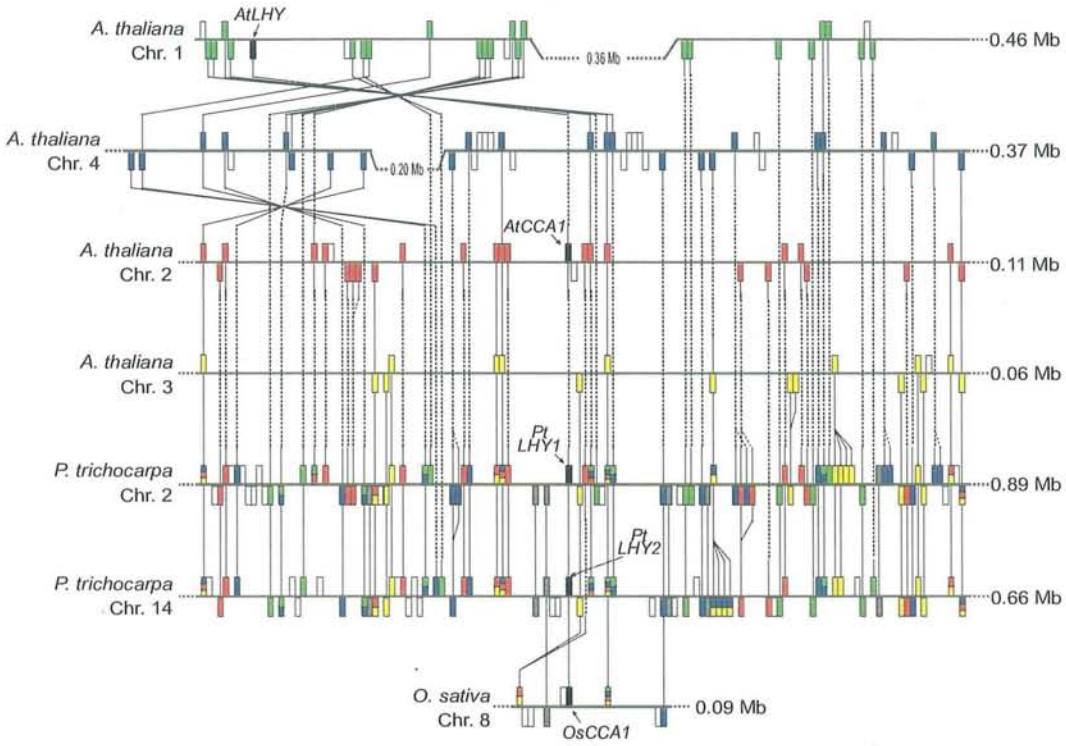


Figure 2-3. Gene organizations surrounding *LHY/CCA1* genes of *A. thaliana, P. trichocarpa* and *O. sativa*. Boxes indicated above and below the thick lines designate genes encoded on forward and reverse DNA strands, respectively. Orthologous genes that were identified by a bidirectional TBLASTN search in the present study and were reproduced from previous studies are colored as follows: green, *Arabidopsis* chromosome (Chr.) 1; blue, *Arabidopsis* Chr. 4; red, *Arabidopsis* Chr. 2; yellow, *Arabidopsis* Chr. 3; gray, *Populus* Chr. 2. Black boxes marked with arrows indicate *LHY/CCA1* genes. Orthologous genes are connected by lines to show the relative positions of each gene pair among different chromosomes. The lengths of the genomic regions are shown on the right.

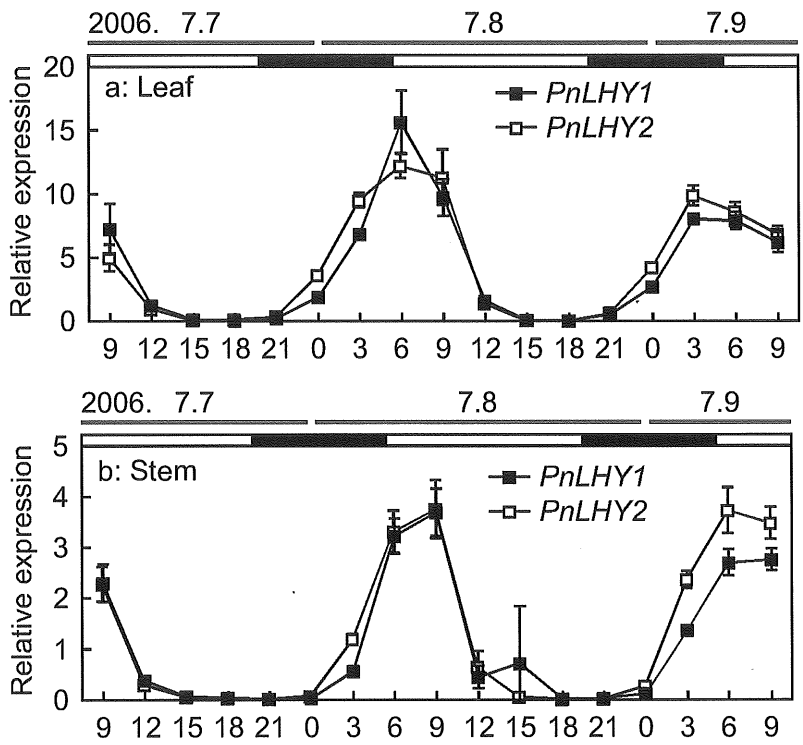


Figure 2-4. Diurnal expressions of *PnLHY1* and *PnLHY2* in leaf (a) and stem (b) tissues. With total RNA isolated from leaves and stems, transcript accumulation of each gene was investigated by real-time PCR using gene-specific primers and normalized to that of *ubiquitin 11* (*UBQ*). The date and time of collection of samples are shown above and below, respectively. White and black bars indicate day and night, respectively. Each RNA sample was assayed in triplicate. RNAs were assayed from two independent biological replicates. Values are means ± SD.

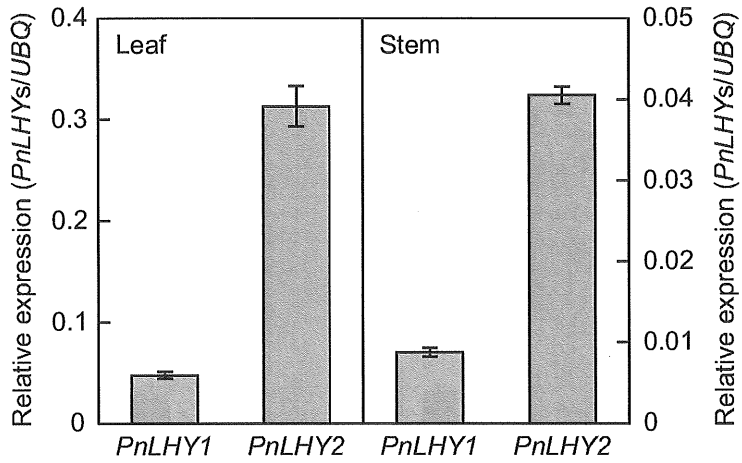


Figure 2-5. Relative expression levels of *PnLHY1* and *PnLHY2* in leaf and stem tissues. With total RNA isolated from the sample (collected at 9:00 AM on July 8, 2006), transcript accumulation of each gene was investigated by real-time PCR using specific primers and normalized to that of *ubiquitin 11* (*UBQ*). A vector harboring fragments of *PnLHY1*, *PnLHY2* and *UBQ* was used to generate a standard curve of real-time PCR amplification. Each RNA sample was assayed in triplicate. RNAs were assayed from two independent biological replicates. Values are means ± SD.

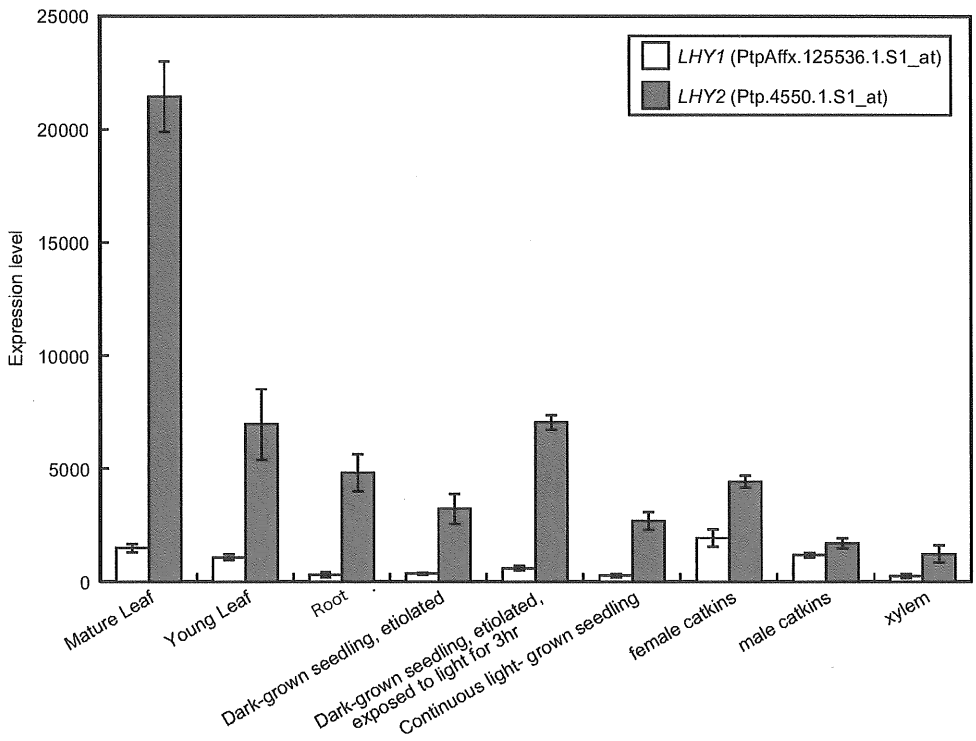


Figure 2-6. Expression levels of *Populus LHY1* and *LHY2* in *Populus* plants. Microarray data are available from the poplar eFP Browser (http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi). Probe Set IDs of *Populus LHY1* and *LHY2* were identified from the database in Affymetrix (http://www.affymetrix.com/). Expression data of *Populus LHY*s were retrieved using Probe Set ID (*Populus LHY1*; PtpAffx.125536.1.S1_at: *Populus LHY2*; Ptp.4550.1.S1_at). Values are means ± SD.

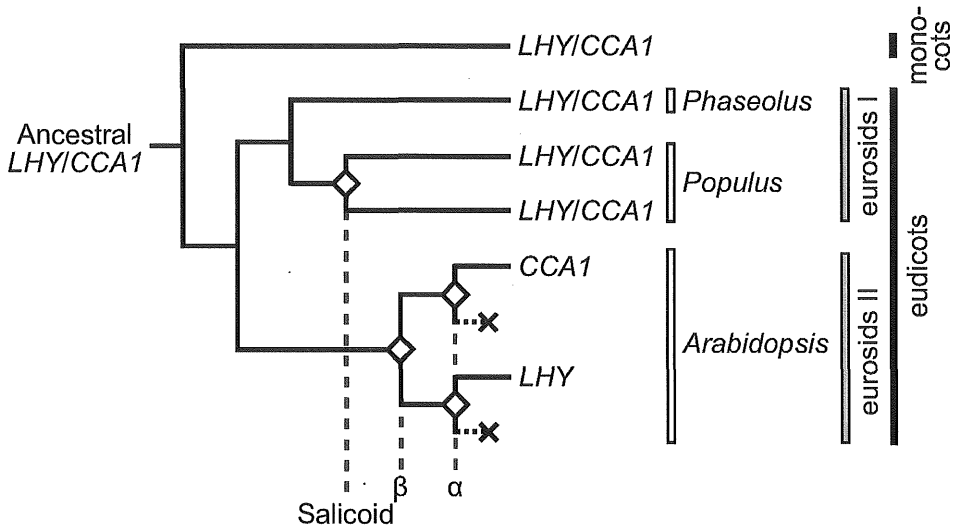


Figure 2-7. Evolutionary scenario of *LHY/CCA1* genes in angiosperms. Diamonds and crosses indicate gene duplication and gene loss events, respectively. Timings of the polyploidy events (Salicoid, β and α) are indicated by broken lines.

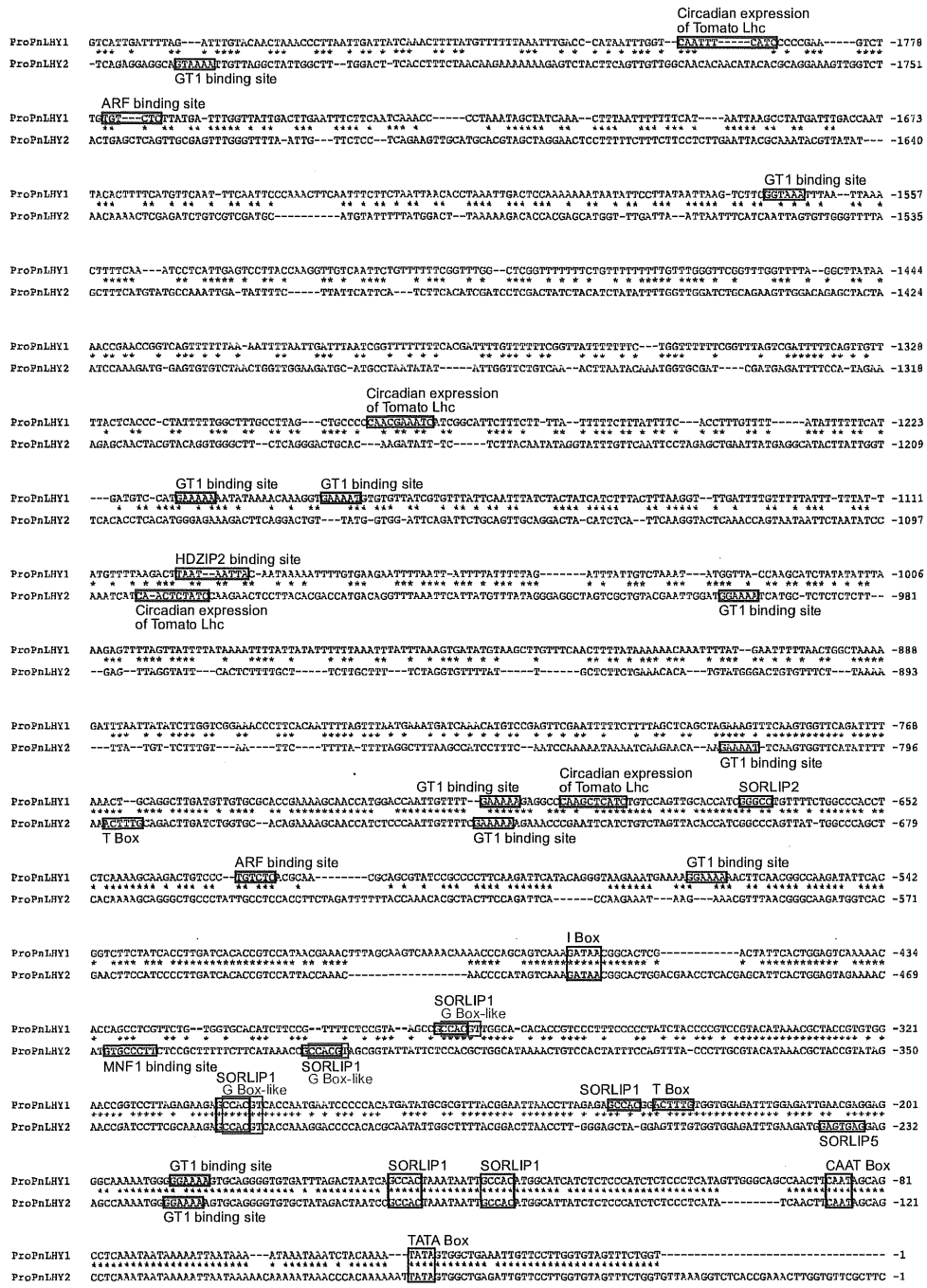


Figure 2-8. Alignment of the promoter sequences of *PnLHY1* and *PnLHY2*. A prediction program (PLACE signal scan; http://www.dna.affrc.go.jp/PLACE/index.html; Higo *et al.*, 1999) was used to identify putative regulatory cis-sequence elements in the *PnLHY* promoters.

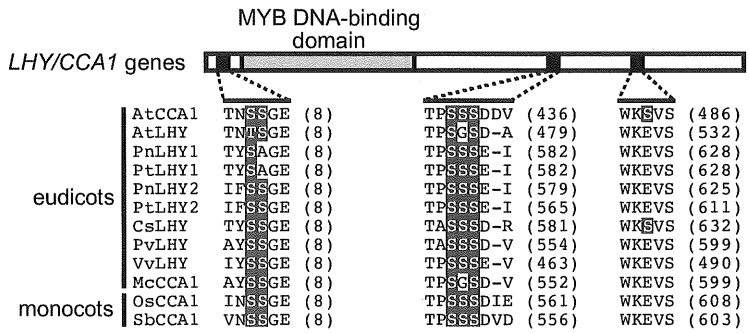


Figure 2-9. Conservation of phosphorylation sites that were identified in the LHY/CCA1 proteins. Shadings in gray on the alignment indicate putative phosphorylation sites. The numbers in parentheses indicate the count of amino acids.

Chapter 3

Phylogenetic footprint of the plant clock system in angiosperms: Evolutionary processes of *Pseudo-Response Regulator*s

Summary

• *PRR* genes have roles in regulating the plant circadian clock system. However, the evolutionary process of *PRR* genes in angiosperm evolutionary lineages is still unclear. The aim of the present study was to investigate the molecular phylogeny of these genes and to deduce the evolutionary process of the plant clock system.

• The molecular phylogeny of angiosperm *PRR* genes was examined by comparison of exon-intron structures, reconstruction of phylogenetic trees, and examination of syntenic relationships.

• Phylogenetic analyses revealed that *PRR* genes had diverged into three clades before the speciation of monocots and eudicots. Furthermore, copy numbers of some *PRR* genes independently increased in eudicots and monocots as a result of ancient chromosomal duplication events.

• The present study uncovered phylogenetic relationships among angiosperm *PRR* genes. Reconsideration of the molecular phylogenies of the plant circadian clock related-genes *LHY/CCA1* and *PRR* leads to a hypothesis on the evolutionary process of the angiosperm plant clock system.

Introduction

Organisms have an endogenous time-keeping mechanism, a circadian clock, to gauge daily and seasonal environmental changes. Circadian clock systems in plants have roles in regulating various photoperiodic responses, such as photomorphogenic processes, floral transition, leaf movements, stomatal conductance, photosynthetic capacity, and volatile emissions (reviewed in Yakir *et al.*, 2007). Because some of these photoperiodisms are conserved across plant species, it is widely thought that plants' circadian clocks share a basic mechanism that controls photoperiodic responses.

In the past decade, numerous molecular genetic analyses of the model plant *Arabidopsis thaliana* have uncovered the basic molecular network of the plant circadian clock (reviewed in McClung, 2006; Más, 2005). Mathematical analyses have been used to develop a computational model of the plant clock system, which contains the main transcriptional feedback loop (Loop I) and additional loops (Loops II and III) associated with the main loop (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). This system of multiple feedback loops is composed of two gene families, *Pseudo-Response Regulator*s (*PRR*s) and *Late Elongated Hypocotyl/Circadian Clock Associated 1* (*LHY/CCA1*), and two unknown factors ("X" and "Y") (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). The main feedback loop consists of two *LHY/CCA1* genes, the *Pseudo-Response Regulator 1/Timing of CAB2 Expression 1* (*PRR1/TOC1*) gene and the unknown factor "X". In this loop, the feedback regulatory network operates as follows: the evening-acting *PRR1/TOC1* gene induces the morning-acting *LHY* and *CCA1* genes via the unknown factor "X", and is in turn repressed by *LHY/CCA1* (Alabadí *et al.*, 2001). To underpin this loop, PRR1/TOC1 protein interacts with PRR3 protein, the product of a paralogous gene of *PRR1/TOC1*, to inhibit its protein degradation (Fujiwara *et al.*, 2008). The main feedback loop associates with Loop II via the *PRR1/TOC1* gene and with Loop III via *LHY/CCA1* genes (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). Loop II is made up of *PRR1/TOC1* and an unknown factor "Y". It has been proposed that the unknown factor "Y" is *GIGANTEA* and/or *PRR5* (McClung, 2008). Loop III consists of *LHY/CCA1* genes and two *PRR* genes, *PRR7* and *PRR9*. Together, the gene families *PRR*s and *LHY/CCA1*s have key roles and form the complex regulatory network in the plant clock system.

PRR genes are widely conserved among angiosperm evolutionary lineages as are the partner *LHY/CCA1* genes (Murakami *et al.*, 2007; Takata *et al.*, in press). *PRR* genes are characterized by two conserved regions; the pseudo-receiver (PR) domain at the N-terminal and the CONSTANS, CONSTANS-LIKE, and TOC1 (CCT) motifs located at the C-terminal region (Murakami *et al.*, 2007). In eudicot lineages, five copies of *PRR* genes have been identified in *A. thaliana* and *Carica papaya* and seven copies have been found in *Populus trichocarpa* (Matsushika *et al.*, 2000; Ming *et al.*, 2008; Ramírez-Carvajal *et al.*, 2008). In monocot lineages, *Oryza sativa,* like *A. thaliana* and *C. papaya,* has five *PRR* genes (Murakami *et al.*, 2003). The expression patterns of *PRR* genes in *A. thaliana* and *O. sativa* share some common features. The five *PRR* genes in *A. thaliana* show diurnal and sequential expression patterns from dawn to dusk as follows; *PRR9→PRR7→PRR5→PRR3→PRR1* (Matsushika *et al.*, 2000). The same sequential expression pattern is found in orthologous genes of *O. sativa*, which are expressed as follows; *OsPRR73* (*OsPRR37*)*→OsPRR95* (*OsPRR59*)*→OsPRR1* (Murakami *et al.*, 2003). In spite of these similarities in the copy numbers and the gene expression patterns of *PRR* genes, it is still unclear how *PRR* genes have evolved in angiosperm lineages.

Rapid accumulation of genomic sequence data offers new perspectives on the molecular phylogeny of genes in angiosperms (Tang *et al.*, 2008). Completion of genomic sequences for various plant species reveals that angiosperm genomes have undergone several ancient chromosomal or whole genome duplication events (Arabidopsis Genome Initiative, 2000; Jaillon *et al.*, 2007; Ming *et al.*, 2008; Tuskan *et al.*, 2006). In monocotyledonous plants, the ρ polyploidy event occurred before the speciation of *O. sativa* and *Sorghum bicolor* in commelinids (Salse *et al.*, 2008; Yu *et al.*, 2005). On the other hand, four polyploidy events appear to have occurred in eudicotyledonous plants. It is thought that the γ triplication event took place after the speciation of monocots and eudicots and before the radiation of rosids species (*Vitis vinifera*, *P. trichocarpa*, *C. papaya,* and *A. thaliana*) but the timing of this event is still being debated (Jaillon *et al.*, 2007; Ming *et al.*, 2008; Tang *et al.*, 2008; Tuskan *et al.*, 2006). The draft genomic sequence of *C. papaya* reveals that the genome of *A. thaliana* underwent two polyploidy events (α and β) after the speciation of *C. papaya* and *A. thaliana*

in eurosids II (Ming *et al.*, 2008). The β polyploidy event is thought to have occurred before the α event (Tang *et al.*, 2008). In the *Populus* lineage of eurosids I, the Salicoid polyploidy event occurred within the Salicaceae after the speciation between Fabales and Malpighiales (Tuskan *et al.*, 2006). The footprints of these chromosomal duplication events are the conserved order of the genes on the duplicated chromosomes in the present genomic sequences (Adams & Wendel, 2005). Thus comparing the order of genes surrounding duplicated genes provides molecular evolutionary information on their phylogenetic relationships (Bocock *et al.*, 2008; Sampedro *et al.*, 2005).

The aim of the present study was to clarify the phylogenetic relationships among angiosperm *PRR* genes. To do this, I identified *PRR* genes from available genomic databases of eudicots (*V. vinifera*, *P. trichocarpa*, *C. papaya,* and *A. thaliana*) and monocots (*O. sativa* and *S. bicolor*). The evolutionary processes of angiosperm *PRR* genes were then examined by comparison of genomic structures, conventional phylogenetic reconstruction, and examination of syntenic relationships. From the results obtained, I reconstructed the molecular phylogeny of *PRR* genes in angiosperms and showed that gene expansion occurred via whole genome duplication events in eudicots and monocots. From the molecular phylogenies of the two gene families (*PRR*s and *LHY/CCA1*s), I propose the evolutionary process of the multiple feedback loops' system in angiosperms.

Materials and Methods

Retrieving sequences of PRR *genes from draft genome sequences*

PRR genes in *A. thaliana* and *O. sativa* were retrieved from genomic databases for *A. thaliana* (TIGR *Arabidopsis thaliana* Database, http://www.tigr.org/tdb/e2k1/ath1/) and *O. sativa* (The Rice Annotation Project Database, http://rapdb.dna.affrc.go.jp/), respectively. To identify *PRR* genes in *S. bicolor*, *V. vinifera*, *P. trichocarpa,* and *C. papaya*, TBLASTN searches were performed against the genomic databases using amino acid sequences encoded by *PRR* genes of *A. thaliana* or *O. sativa* as queries: JGI *Sorghum bicolor* v1.0 (http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html) for *S. bicolor*; Grape Genome Browser (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/) for *V. vinifera*; JGI *Populus trichocarpa* v1.1 (http://genome.jgi-psf.org/Poptr1_1/

Poptr1_1.home.html) for *P. trichocarpa*; and Papaya Genome Project v0.4 in CoGe (http://synteny.cnr.berkeley.edu/CoGe/) for *C. papaya*. Genes that retained the typical PR-domain at the N-terminal region and the CCT-motif at the C-terminal region were retrieved from these genomic databases. Genes that lacked the PR-domain or CCT motif but showed significantly high similarity were also retrieved (those with E-values lower than 10^{-50} or >90% similarity). The genes retrieved from the genomic databases were aligned with *PRR* genes in *A. thaliana* and *O. sativa* using the TCoffee program (http://www.ebi.ac.uk/t-coffee/). Mispredicted genes, if found, were manually modified as follows: For predicted genes lacking a conserved portion of the *PRR* gene, I searched the database for expressed sequence tags (ESTs) of the gene (TIGR Plant Transcript Assemblies, http://plantta.tigr.org/) and re-annotated by assembling the predicted gene and relevant ESTs. In some cases, the open reading frame (ORF) of the gene was repredicted by the Fgenesh+ program (http://linux1.softberry.com/berry.phtml). When the exon-intron boundary of a gene was mis-demarcated, I improved the boundary based on standard donor/acceptor splice sites without resulting in a frame shift.

Phylogenetic analysis

Amino acid sequences were deduced from nucleotide sequences of the predicted *PRR* and then aligned using the TCoffee program. The number of amino acids substituted between each pair of PRR proteins was estimated by the Jones-Taylor-Thornton (JTT) model (Jones *et al.*, 1992) with the complete-deletion option. From the number of estimated amino acid substitutions, a phylogenetic tree was reconstructed by the Minimum Evolution (ME) method (Rzhetsky & Nei, 1992). Bootstrap values were calculated with 1,000 replications using the ME method (Felsenstein, 1985). These procedures were performed using MEGA4 software (http://www.megasoftware.net/index.html) (Tamura *et al.*, 2007).

Identification of chromosomal synteny

Conservation of chromosomal synteny in *V. vinifera*, *P. trichocarpa*, *C. papaya*, and *A. thaliana* was determined as follows: I reconstructed the ancient gene

organization of the flanking regions of *A. thaliana PRR* genes before the α and β polyploidy events using the chromosomal syntenies reported in previous studies (Bower *et al.*, 2003; Thomas *et al.*, 2006). Then, I compared the syntenic relationships between the ancient gene organization in *A. thaliana* and the flanking regions of *PRR* genes in *V. vinifera*, *P. trichocarpa,* and *C. papaya* using the comparative genomic tool, CoGe (http://synteny.cnr.berkeley.edu/CoGe/) (Lyons & Freeling, 2008; Ming *et al.*, 2008). This process also reconstructed chromosomal syntenies in *P. trichocarpa* that were derived from the Salicoid polyploidy event. To clarify syntenic relationships derived from the γ polyploidy event, I used information on chromosomal syntenies within the genome of *V. vinifera* (Jaillon *et al.*, 2007) and the comparative genomic tool, CoGe.

Syntenic relationships between the flanking regions of *PRR* genes in *O. sativa* and those in *S. bicolor* were reconfirmed using the chromosomal syntenies reported in previous studies (Paterson *et al.*, 2009) and the comparative genomic tool VISTA Browser (http://genome.lbl.gov/vista/index.shtml). To reconstruct chromosomal syntenies of the flanking regions of *PRR* genes derived from the monocotyledonous ρ polyploidy event, syntenic regions were identified according to methods reported previously using the *O. sativa* genomic sequence (Salse *et al.*, 2008).

Results

Identification of PRR *genes in angiosperms*

There are five copies of the *PRR* gene in the genomes of *O. sativa*, *S. bicolor*, *V. vinifera* and *C. papaya*, six copies in *A. thaliana*, and eight copies in *P. trichocarpa* (Table 3-1). *PRR1/TOC1* genes in *C. papaya* were not retrieved from the genomic sequence database because the nucleotide sequence of this gene's C-terminal region has not yet been determined. The angiosperm *PRR* genes retained a highly conserved PR-domain at the N-terminus and a CCT-motif at the C-terminus (Fig. 3-1). However, two *PRR*-like genes (*PRR9b* in *A. thaliana* and *PRR5c* in *P. trichocarpa*) retained the CCT-motif but not the PR-domain (Figs. 3-2 and 3-3). *AtPRR9b* showed high homology with *AtPRR9* at the C-terminal region, whereas *AtPRR9b* lacked one-third of the N-terminal region compared with *AtPRR9* (Fig. 3-2a). The exon-intron structures of *AtPRR9b* were similar to

the corresponding region of *AtPRR9* (Fig. 3-2c). *AtPRR9* and *AtPRR9b* are encoded on opposite DNA strands in chromosome 2, and are separated by a 71-kb genomic region containing 13 intervening genes (Fig. 3-2b). Similarly, *PtPRR5c* shared high homology and similar exon-intron structure with *PtPRR5b*, although *PtPRR5c* lacked two-thirds of its N-terminal portion compared with *PtPRR5b* (Figs. 3-3a and c). *PtPRR5b* and *PtPRR5c* are located within the 18-kb genomic region on chromosome 15 and are separated by two intervening genes (Fig. 3-3b). In the present study, I examined the molecular phylogeny of *PRR* genes, except for *AtPRR9b* and *PtPRR5c*.

Genomic structures of angiosperm PRR *genes*

To investigate the evolutionary footprints of *PRR* genes in angiosperms, I compared exon-intron structures and insertions/deletions (indels) of the genes. I aligned the predicted amino acid sequences of *PRR* genes and compared conservation of intron insertion sites and distribution of indel variation.

Angiosperm PRR proteins showed high alignment quality around the two conserved regions, the PR-domain, and the CCT-motif (Fig. 3-1). Around the PR-domain, three intron insertion sites were highly conserved in all eudicot and monocot *PRR* genes (Fig. 3-4). On the other hand, three nucleic acids were deleted from regions adjacent to the PR-domain in 14 genes (*AtPRR5*, *AtPRR9*, *CpPRR5*, *CpPRR9*, *PtPRR5a*, *PtPRR5b*, *PtPRR9a*, *PtPRR9b*, *VvPRR5*, *VvPRR9*, *OsPRR59*, *OsPRR95*, *SbPRR59,* and *SbPRR95*; Figs. 3-4 and 3-5). An intron-exon boundary was conserved within the CCT motif of most *PRR* genes, except for *AtPRR1/TOC1*, *PtPRR1/TOC1*, *VvPRR1/TOC1*, *OsPRR1/TOC1,* and *SbPRR1/TOC1*. These differences among exon-intron structures and indels indicated that angiosperm *PRR* genes were divided into three groups; the *PRR1/TOC1* clade, the *PRR3* and *7* clade, and the *PRR5* and *9* clade. The *PRR1/TOC1* clade contains *PRR1/TOC1*s of both eudicots and monocots, the *PRR3* and *7* clade contains *PRR3/7*s of eudicots and *PRR37/73*s of monocots, and the *PRR5* and *9* clade contains *PRR5/9*s of eudicots and *PRR59/95*s of monocots.

Phylogenetic analysis of PRR *gene family*

To deduce the evolutionary relationships among *PRR* genes in angiosperms, a

phylogenetic tree was reconstructed using the Minimum Evolution (ME) method. Angiosperm *PRR* genes clearly separated into three clades (*PRR1/TOC1* clade, *PRR3* and *7* clade, and *PRR5* and *9* clade), which are consistent with the categorization of *PRR* genes' genomic structures (Figs. 3-4 and 3-6). Each clade contains genes from both eudicots and monocots. Consequently, these data suggest that ancient *PRR* gene(s) diverged into three clades before the speciation of monocots and eudicots.

In all plant species examined, one copy of the *PRR1/TOC1* gene was retained in the *PRR1/TOC1* clade (Fig. 3-6), whereas at least two copies were found in the *PRR3* and *7* clade and the *PRR5* and *9* clade.

The *PRR3* and *7* clade consisted of two different clusters, each exclusively monocot or eudicot genes (Fig. 3-6). Accordingly, the phylogenetic tree suggested that the gene duplication events producing monocotyledonous *PRR37* and *PRR73* or eudicotyledonous *PRR3* and *PRR7* occurred independently within monocot and eudicot lineages, respectively. After the duplication event in eudicots, orthologs of *P. trichocarpa PRR3* were lost, whereas the *P. trichocarpa PRR7* gene appeared to be duplicated into *PRR7a* and *7b*.

In the *PRR5* and *9* clade, the monocot *PRR59* and *PRR95* genes showed an earlier gene duplication event that may have occurred in a common ancestor of monocots and eudicots. However, the bootstrap value supporting this branch was not very high, 56% (Fig. 3-6). On the other hand, eudicotyledonous *PRR5* and *PRR9* genes formed a gene cluster in the phylogenetic tree. In this cluster, *A. thaliana PRR9* was distantly related to other *PRR5* and *PRR9* genes. This was also observed in the phylogenetic tree reconstructed by the Neighbor-Joining method (Fig. 3-7). In *Populus*, both *PRR5a* and *5b* and *PRR9a* and *9b* were more closely related than other *PRR5* and *PRR9* genes, suggesting that the gene duplication events that produced *PRR5a* and *5b* and *PRR9a* and *9b* occurred in *P. trichocarpa* (Fig. 3-6).

Consequently, although the *PRR3* and *7* clade and the *PRR5* and *9* clade contained at least two copies of *PRR* genes in both eudicots and monocots, *PRR*s in the two clades are assumed to have independently duplicated in eudicot and monocot lineages.

Phylogenetic relationships of PRR *gene family inferred from chromosome syntenies*

To clarify evolutionary events such as gene duplication and gene deletion among angiosperm *PRR* genes, I investigated chromosomal syntenies among the genomes of monocots or eudicots. Because ancient chromosome duplication events result in conserved gene order on the duplicated chromosomes (Adams & Wendel, 2005), comparisons of gene organization and detection of chromosomal syntenies can provide molecular evolutionary information to understand the phylogenetic relationships of the genes (Bocock *et al.*, 2008; Sampedro *et al.*, 2005).

In the genomic sequences of *O. sativa* and *S. bicolor*, the flanking region of the *PRR* gene in *O. sativa* (*OsPRR1/TOC1*, *OsPRR73*, *OsPRR59* and *OsPRR95*) showed conserved synteny with that of orthologous genes in *S. bicolor* (Paterson *et al.*, 2009). On the other hand, only a few syntenic regions were identified between *PRR37* of *O. sativa* and the *S. bicolor* ortholog. In the genome of *O. sativa*, the neighboring region of *OsPRR37* shared synteny with that of *OsPRR73*. This syntenic relationship resulted from the chromosomal duplication that occurred in the ρ polyploidy event (Salse *et al.*, 2008). The ρ polyploidy event also resulted in conserved chromosomal synteny between the flanking region of *OsPRR59* and a partial region of chromosome 8, and between the flanking region of *OsPRR95* and a different partial region of chromosome 8. However, these partial regions of chromosome 8 have lost *PRR* genes. These results indicated that the gene duplication event resulting in *PRR37* and *PRR73* was the monocotyledonous ρ polyploidy event, and that *PRR59* and *PRR95* were duplicated via the ρ polyploidy event but one of the duplicated genes was lost from genomes of *O. sativa* and *S. bicolor* (Fig. 3-9).

In eudicots, the flanking region of each *PRR* gene in *A. thaliana* (*AtPRR1/TOC1*, *AtPRR3*, *5*, *7* and *9*) generally shares a syntenic relationship with those of each orthologous gene in *V. vinifera*, *P. trichocarpa,* and *C. papaya* (Fig. 3-8). The syntenic relationship was not found between the neighboring regions of *P. trichocarpa PRR1/TOC1* and those of other *PRR1/TOC1*s, or between the neighboring region of *C. papaya PRR9* and those of other *PRR9*s. It is not clear whether these observations can be ascribed to chromosomal rearrangements or

fragmentations of genomic sequences in small contigs.

In the genome of *A. thaliana*, the flanking regions of *PRR* genes showed syntenies with one or three partial regions of its genome (Fig. 3-8). These syntenic relationships originated from the chromosomal duplications that arose from the β and α polyploidy events (Bower *et al.*, 2003; Thomas *et al.*, 2006). In the *P. trichocarpa* genome, two copies of each *PRR5*, *PRR7,* and *PRR9* gene are located at the syntenic regions of chromosomes 12 and 15 (Fig. 3-8b), those of chromosomes 8 and 10 (Fig. 3-8d), and those of chromosomes 2 and 14 (Fig. 3-8e), respectively. These syntenic regions were produced via the Salicoid polyploidy event (Tuskan *et al.*, 2006). Although flanking regions of *A. thaliana PRR3* share syntenic relationships with partial regions of *P. trichocarpa* chromosomes 1 and 9, these two partial regions did not retain a *PRR* gene (Fig. 3-8c).

Next, I examined the chromosomal syntenic relationships derived from the γ triplication event. Genomic information from *V. vinifera* was used for this analysis because its genome shows a low level of genomic rearrangement (Jaillon *et al.*, 2007). Comparisons within the genome of *V. vinifera* indicated that ancestral flanking regions of *PRR* genes seemed to be derived from the γ triplication event. There were conserved chromosomal syntenies between the flanking regions of *VvPRR3* and *VvPRR7* and between the flanking regions of *VvPRR5* and *VvPRR9* (Jaillon *et al.*, 2007; see also the comparative genomic tool, CoGe, http://synteny.cnr.berkeley.edu/CoGe/).

Consequently, in the eudicotylodonous γ polyploidy event, the ancestral *PRR3/7* gene was duplicated into *PRR3* and *PRR7,* and *PRR5/9* into *PRR5* and *PRR9* (Fig. 3-9). After the γ polyploidy event, one copy of each *PRR* gene (*PRR1/TOC1*, *PRR3*, *5*, *7* and *9*) has been conserved in the present genomes of *V. vinifera* and *C. papaya*, which have not undergone additional polyploidy events. Although *A. thaliana PRR* genes were repeatedly duplicated by the β and/or α polyploidy events (Fig. 3-8), one copy of each gene remains in the present *A. thaliana* genome, similar to *V. vinifera* and *C. papaya* (Fig. 3-9). In the genome of *P. trichocarpa*, *PRR5*, *7,* and *9* were duplicated in the Salicoid polyploidy event, but the *PRR3* gene seems to be deleted from partial regions of chromosomes 1 and/or 9 (Figs. 3-8 and 3-9).

Discussion

The plant clock system consists of multiple interlocked feedback loops, which are comprised predominantly of two gene families, *LHY/CCA1*s and *PRR*s (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). These gene families are widely conserved among both monocots and eudicots (Murakami *et al.*, 2007). To clarify the evolutionary process of this plant clock system, I recently reported the molecular phylogeny of *LHY/CCA1* genes in angiosperms (Takata *et al.*, in press). However, the evolutionary process of *PRR* genes is more difficult to determine, because of the complexities that arise from the multiple copies of *PRR* genes in each genome. In this study, I reconstructed phylogenetic relationships among *PRR* genes in eudicots and monocots using three approaches: comparison of genomic structures, reconstruction of phylogenetic trees, and examination of syntenic relationships. Together, these phylogenetic analyses of the plant circadian clock related-genes *LHY/CCA1*s and *PRR*s are promising tools to unravel the evolutionary process of the plant clock system among angiosperm evolutionary lineages.

Evolutionary processes of PRR *genes in angiosperms*

PRR genes are widely conserved in angiosperms (monocots; *O. sativa* and *S. bicolor*: eudicots; *V. vinifera*, *P. trichocarpa*, *C. papaya* and *A. thaliana*) and at least five copies *of PRR* genes have been retained in their genomes (Table 3-1). These *PRR* genes are tentatively categorized into three or four clusters (Miwa *et al.*, 2006; Murakami *et al.*, 2003). In the present study, differences in genomic structures (exon-intron structures and indels) and the topology of phylogenetic tree demonstrated that the angiosperm *PRR* genes are grouped into three clades (the *PRR1/TOC1* clade, the *PRR3* and *7* clade, and the *PRR5* and *9* clade) (Figs. 3-4 and 3-6). Each clade contains *PRR* genes of both eudicots and monocots. These results revealed that three major clusters of *PRR* genes already existed in the most recent common ancestor of eudicots and monocots, but the precise timing of the diversification of *PRR* genes could not be determined. Further accumulation of genomic information from lower plants (Bryophyta, Pteridophyta, Lycopodiophyta, Equisetophyta and gymnosperms) will clarify aspects of the ancient divergence of *PRR* genes.

The *PRR3* and *7* clade and the *PRR5* and *9* clade have at least two copies of *PRR* genes in each plant species (Figs. 3-4 and 3-6). The topology of the phylogenetic tree indicates that the *PRR* genes were duplicated independently in monocots and eudicots (Fig. 3-6). Indeed, the syntenic relationships clearly showed that the gene duplication between monocotyledonous *PRR37* and *PRR73* results from the ρ polyploidy event that occurred in monocotyledonous lineages, and that the gene duplication between eudicotyledonous *PRR3* and *PRR7* is derived from the γ polyploidy event that occurred in eudicotyledonous lineages (Fig. 3-9; Jaillon *et al.*, 2007; Salse *et al.*, 2008; Yu *et al.*, 2005). In the *PRR5* and *9* clade, eudicotyledonous *PRR5* and *PRR9* were duplicated in the eudicotyledonous γ polyploidy event. On the other hand, the timing of gene duplication between monocotyledonous *PRR59* and *PRR95* is yet to be determined. *OsPRR59* and *OsPRR95* must have existed before the ρ polyploidy event because the flanking regions of these genes were affected by the ρ polyploidy event. However, it is unclear whether this gene duplication had already occurred before the speciation of monocots and eudicots.

PRR genes in *P. trichocarpa* have expanded more than those in other plant species (Table 3-1; Ramírez-Carvajal *et al.*, 2008). This expansion resulted from the Salicoid polyploidy event that occurred in the *Populus* lineage but not in other eudicots (*V. vinifera, C. papaya,* and *A. thaliana*). The Salicoid polyploidy event affected more than 90% of the genome of *P. trichocarpa* and approximately 8,000 pairs of the Salicoid duplicated genes remain in the present genome (Tuskan *et al.*, 2006). In addition, nearly 5,000 *Populus* genes have undergone tandem duplication. For example, the present study demonstrated that *PRR5b* and *PRR5c* arose from a tandem duplication within a region on chromosome 15 (Fig. 3-3). The gene duplication event between *PRR5b* and *PRR5c* probably occurred after the Salicoid polyploidy event because the flanking region of *PRR5a,* which is a Salicoid duplicated region with *PRR5b,* does not retain the tandem duplication trait (see also JGI *Populus trichocarpa* v1.1, http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). However, the PR-domain of *PRR5c* was deleted after the gene duplication event, which implies pseudogenization of the gene.

In contrast to gene expansion via the Salicoid polyploidy event, *P. trichocarpa* has not retained the *PRR3* gene, which has a role in the posttranslational

degradation of PRR1/TOC1 protein in *A. thaliana* (Fujiwara *et al.*, 2008). Ancient *PRR3*(s) of *P. trichocarpa* are located on the Salicoid duplicated regions of chromosomes 1 and/or 9 (Fig. 3-8). However, *PRR3*(s) have been lost from one or both regions in the genome after speciation of eurosids I and II. The difference in the composition and number of *P. trichocarpa PRR* genes suggest that there are differences between the regulation mechanisms of clock systems in *P. trichocarpa* and *A. thaliana.*

The plant clock system is a complex mechanism involving many regulator genes, which can transmit a photoperiodic signal to a number of downstream genes (reviewed in McClung, 2006; Más, 2005, Michael & McClung, 2003). Recent computational genomic analyses in *A. thaliana* and *O. sativa* have revealed that some signal transduction components and transcriptional factors were over-retained in the genome after whole genome duplication events (Blanc & Wolfe, 2004; Chapman *et al.*, 2006; Freeling & Thomas, 2006; Maere *et al.*, 2005; Thomas *et al.*, 2006). It is likely that factors participating in such a regulatory mechanism would be particularly dosage-sensitive and show a stoichiometric balance, which if upset, would modify regulation of downstream targets (Birchler & Veitia, 2007; Freeling & Thomas, 2006). This study revealed that *A. thaliana*'s *PRR* genes, examples of a transcriptional factor, have been repeatedly duplicated through β and α polyploidy events (Fig. 3-9). In spite of these repetitive gene duplications, the present genome of *A. thaliana* retains only one copy of each *PRR* gene (*PRR1/TOC1*, *PRR3*, *5*, *7*, and *9*). Likewise, *LHY/CCA1* genes, the partner of *PRR* genes, have undergone subsequent gene losses after gene duplications occurred in the α polyploidy event (Takata *et al.*, in press). Collectively, the findings in the present study imply that the regulatory network of the *Arabidopsis* clock system retained a degree of organization throughout the dynamic changes of copy numbers and functions of circadian clock-related genes.

Phylogenetic footprint of the plant clock system in angiosperms

Loops I and III of the *Arabidopsis* clock system are thought to contain four *PRR* genes (*PRR1/TOC1*, *PRR3*, *7* and *9*) and two *LHY/CCA1* genes (*LHY* and *CCA1*) (Fig. 3-10; Locke *et al.*, 2006; Zeilinger *et al.*, 2006). Although the circadian

clock-related genes in *A. thaliana* were duplicated via the β and α polyploidy events, the present genome of *A. thaliana* retains only one pair of *LHY* and *CCA1* genes, which is derived from the β polyploidy event (Takata *et al.*, in press). As the β polyploidy event is assumed to have occurred in the Brassicaceae, *LHY* and *CCA1* genes did not diverge before the speciation of *A. thaliana* and *C. papaya*, which is consistent with the fact that there is only one copy of the *LHY/CCA1* gene in the genome of *C. papaya* (Fig. 3-10; Ming *et al.*, 2008). Similar to the genome of *A. thaliana*, the *C. papaya* genome retained only one copy each of the *PRR1/TOC1*, *PRR3*, *7*, and *9* genes (Table 3-1). These results suggest that only one copy of *LHY/CCA1*, *PRR1/TOC1*, *PRR3*, *7*, and *9* were involved in the plant clock system in the common ancestor of *A. thaliana* and *C. papaya*.

The similarities between *A. thaliana* and *C. papaya* raise the question as to when the Brassicales-type clock system arose. Phylogenies of the circadian clock-related genes showed that a set of the genes, one copy each of *LHY/CCA1*, *PRR1/TOC1*, *PRR3*, *7* and *9*, are conserved in the genome of *V. vinifera* (Table 3-1; Fig. 3-10; Jaillon *et al.*, 2007; Takata *et al.*, in press). Although it remains to be determined whether functional divergences between *PRR3* and *7* and between *PRR5* and *9* exist in *V. vinifera*, conservation of the set of clock-related genes suggests that the fundamental mechanism of the Brassicales-type clock system was organized before the speciation of *V. vinifera* and other eurosids species (*P. trichocarpa*, *C. papaya*, and *A. thaliana*; Fig. 3-10).

Intriguingly, the clock system of *P. trichocarpa* might differ from the Brassicales-type clock system, because the *P. trichocarpa PRR3* gene was lost and *LHY/CCA1* and *PRR7* and *9* were duplicated via the Salicoid polyploidy event that occurred after the speciation of eurosids I and II (Figs. 3-9 and 3-10). *Populus LHY*s show typical diurnal expressions similar to *LHY/CCA1* genes in other plant species (Izawa *et al.*, 2002; Kaldis *et al.*, 2003; Ramos *et al.*, 2005; Takata *et al.*, in press; Wang & Tobin, 1998). On the other hand, the lack of *PRR3* might influence the posttranslational regulation of PRR1/TOC1 protein in *Populus*. Furthermore, duplication of *LHY/CCA1* and *PRR7* and *9*, but not *PRR1/TOC1*, could also affect the regulation mechanism of the *Populus* clock system. Further functional studies are needed to clarify the regulation network of the *Populus* clock system.

The monocots *O. sativa* and *S. bicolor* retain one *LHY/CCA1* gene and five *PRR* genes in their genomes (Table 3-1; Takata *et al.*, in press). Phylogenetic analysis showed that the gene duplication events that produced monocots' *PRR37* and *73*, and *PRR59* and *95* occurred at different times to the events that produced eudicots' *PRR3* and *7*, and *PRR5* and *9* (Fig. 3-9). This finding raises a complex question; what are the roles of these paralogous genes in the monocotyledonous clock system? It was reported that *PRR59* and *95* in *O. sativa* had peak expressions just before dusk and were not induced by light signals. These aspects of expression are more similar to the regulation of *A. thaliana PRR5* expression than to that of *A. thaliana PRR9* expression (Ito *et al.*, 2003; Matsushika *et al.*, 2000; Murakami *et al.*, 2003). Furthermore, peak expressions of *O. sativa PRR37* and *73* are followed by expressions of *PRP59* and *95*, patterns that are similar to the sequential expressions of *A. thaliana PRR7* and *PRR5* (Matsushika *et al.*, 2000; Murakami *et al.*, 2003). The expression patterns of *PRR37*, *73*, *59,* and *95* of *O. sativa* are similar to those of orthologous genes in the monocot *Lemna* (Arales) (Miwa *et al.*, 2006). These results suggest that monocots' paralogous gene pairs *PRR59/95* and *PRR37/73* genes share functional roles with *PRR5* and *PRR7*, respectively, in *A. thaliana* (Fig. 3-10). Together, these data suggest that a common ancestor of eudicots and monocots may have had a main feedback loop (*LHY/CCA1-PRR1/TOC1*) that was not posttranslationally regulated by *PRR3,* and may have decreased activity of the morning loop (*LHY/CCA1-PRR3/7*) without *PRR5/9.*

In conclusion, the present study shows the molecular phylogeny of angiosperm *PRR* genes that have key roles in the plant clock system. *PRR* genes diverged into three clades before the speciation of monocots and eudicots and, in addition, *PRR3/7* and *PRR5/9* underwent independent expansion in eudicots and monocots, respectively (Fig. 3-9). Furthermore, reconsidering the molecular phylogenies of *PRR* genes and *LHY/CCA1* genes led to a hypothesis on the evolutionary process of the angiosperm plant clock system (Fig. 3-10). Additional functional analyses and accumulation of genomic information from other plant species will clarify details of evolutionary and developmental processes of plants' clock systems.

Table 3-1. *PRR* genes in angiosperms used in the present study.

Classification[a]		Species	Gene name	Accession number / Gene ID	Chromosome Number	Location
monocots	commelinids	*Oryza sativa*	*OsPRR1/TOC1*	Os02g0618200[b], Os02g40510[g]	2	24564051-24566250
			OsPRR37	Os07g0695100[b], Os07g49460[g]	7	29615737-29626907
			OsPRR59	Os11g0157600[b], Os11g05930[g]	11	2784911-2789386
			OsPRR73	Os03g0284100[b], Os03g17570[g]	3	9739149-9746770
			OsPRR95	Os09g0532400[b], Os09g36220[g]	9	20885108-20889176
		Sorghum bicolor	*SbPRR1/TOC1*	Sb04g026190[c]	4	55958216-55960718
			*SbPRR37**	Sb06g014570[c]	6	40280414-40290602
			SbPRR59	Sb05g003660[c]	5	4254506-4259308
			SbPRR73	Sb01g038820[c]	1	62307206-62313798
			*SbPRR95**	Sb02g030870[c], TA24276_4558[d]	2	65861609-65865883
eudicots	rosids	*Vitis vinifera*	*VvPRR1/TOC1*	GSVIVT00016296001[e]	17	7198529-7204195
			*VvPRR3**	GSVIVT00024444001[e]	6	4584374-4607119
			*VvPRR5**	GSVIVT00037169001[e]	16	7347191-7353600
			VvPRR7	GSVIVT00029486001[e]	13	1829516-1861245
			VvPRR9	GSVIVT00026171001[e]	15	7285024-7290530
	eurosids I	*Populus trichocarpa*	*PtPRR1/TOC1*	fgenesh4_pg.C_scaffold_129000038[f]	scaffold_129	470210-475633
			PtPRR5a	fgenesh4_pm.C_LG_XII000134[f]	12	3179830-3184499
			PtPRR5b	eugene3.00150024[f]	15	134835-139611
			*PtPRR5c***	estExt_fgenesh4_pg.C_LG_XV0023[f]	15	149645-153132
			PtPRR7a	estExt_fgenesh4_pm.C_LG_VIII0151[f]	8	2476118-2483078
			PtPRR7b	fgenesh4_pg.C_LG_X001956[f]	10	18604746-18611132
			PtPRR9a	fgenesh4_pg.C_LG_II001656[f]	2	14066628-14071402

Table 3-1. continued

		PtPRR9b	estExt_fgenesh4_pg.C_LG_XIV0468[f]	14	4273730-4278465
eurosids II	*Arabidopsis thaliana*	*AtPRR1/TOC1*	At5g61380[g]	5	24692290-24695776
		AtPRR3	At5g60100[g]	5	24215225-24218590
		AtPRR5	At5g24470[g]	5	8355954-8358876
		AtPRR7	At5g02810[g]	5	637895-641975
		AtPRR9	At2g46790[g]	2	19239718-19242156
		*AtPRR9b***	At2g46670[g]	2	19171484-19172303
	Carica papaya	*CpPRR1/TOC1***	EVM prediction supercontig_13.294[h]	supercontig_13	2714773-2715852
		CpPRR3	EVM prediction supercontig_139.32[h]	supercontig_139	194903-200919
		CpPRR5	EVM prediction supercontig_3.152[h]	supercontig_3	1172848-1176846
		CpPRR7	EVM prediction supercontig_1.291[h]	supercontig_1	4024079-4030318
		CpPRR9	EVM prediction supercontig_193.20[h]	supercontig_193	363399-369853

[a]Plant classification refers to APGII (http://www.mobot.org/MOBOT/Research/APweb/welcome.html).

[b]Gene ID corresponds to the name obtained from RAP-DB (http://rapdb.dna.affrc.go.jp/).

[c]Gene ID corresponds to the name obtained from JGI (http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html).

[d]Gene IDs correspond to the names obtained from TIGR Plant Transcript Assemblies (http://plantta.tigr.org/).

[e]Gene ID corresponds to the name obtained from Grape Genome Browser (http://www.genoscope.cns.fr/externe/English/Projets/Projet_ML/index.html).

[f]Gene IDs correspond to the names obtained from JGI (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html).

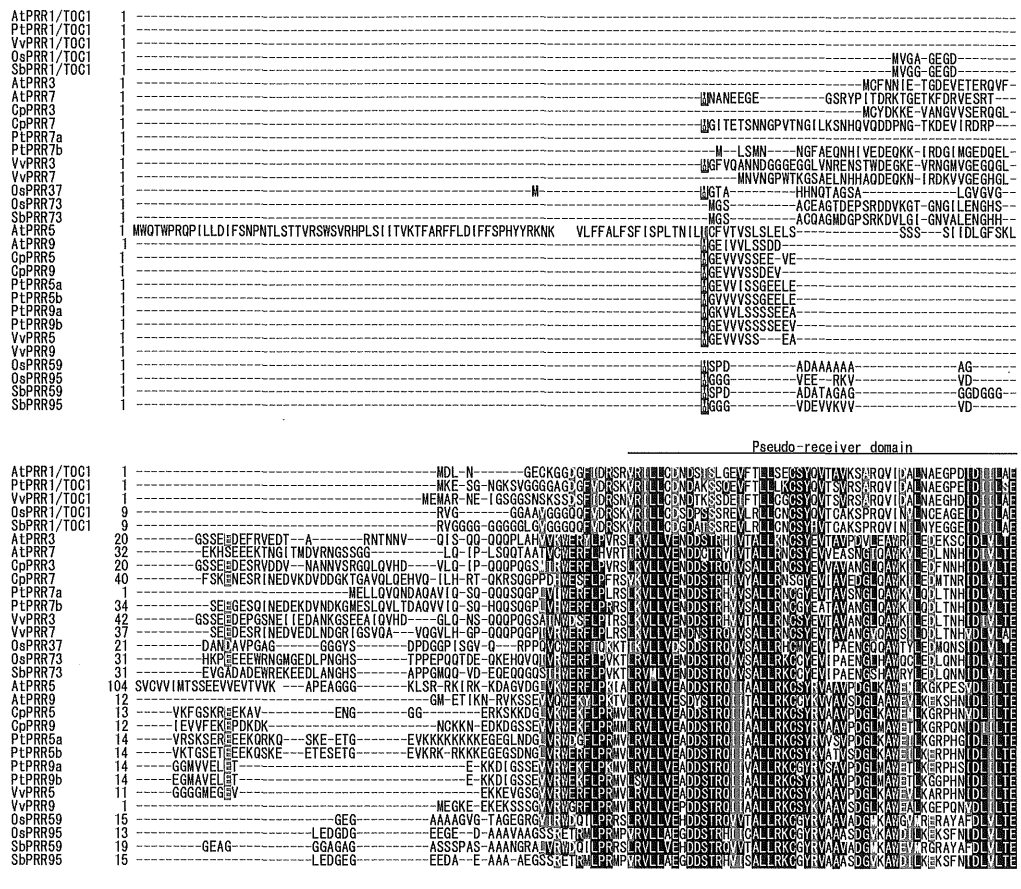
[g]Gene IDs correspond to the names obtained from TIGR (http://plantgenomics.tigr.org/).

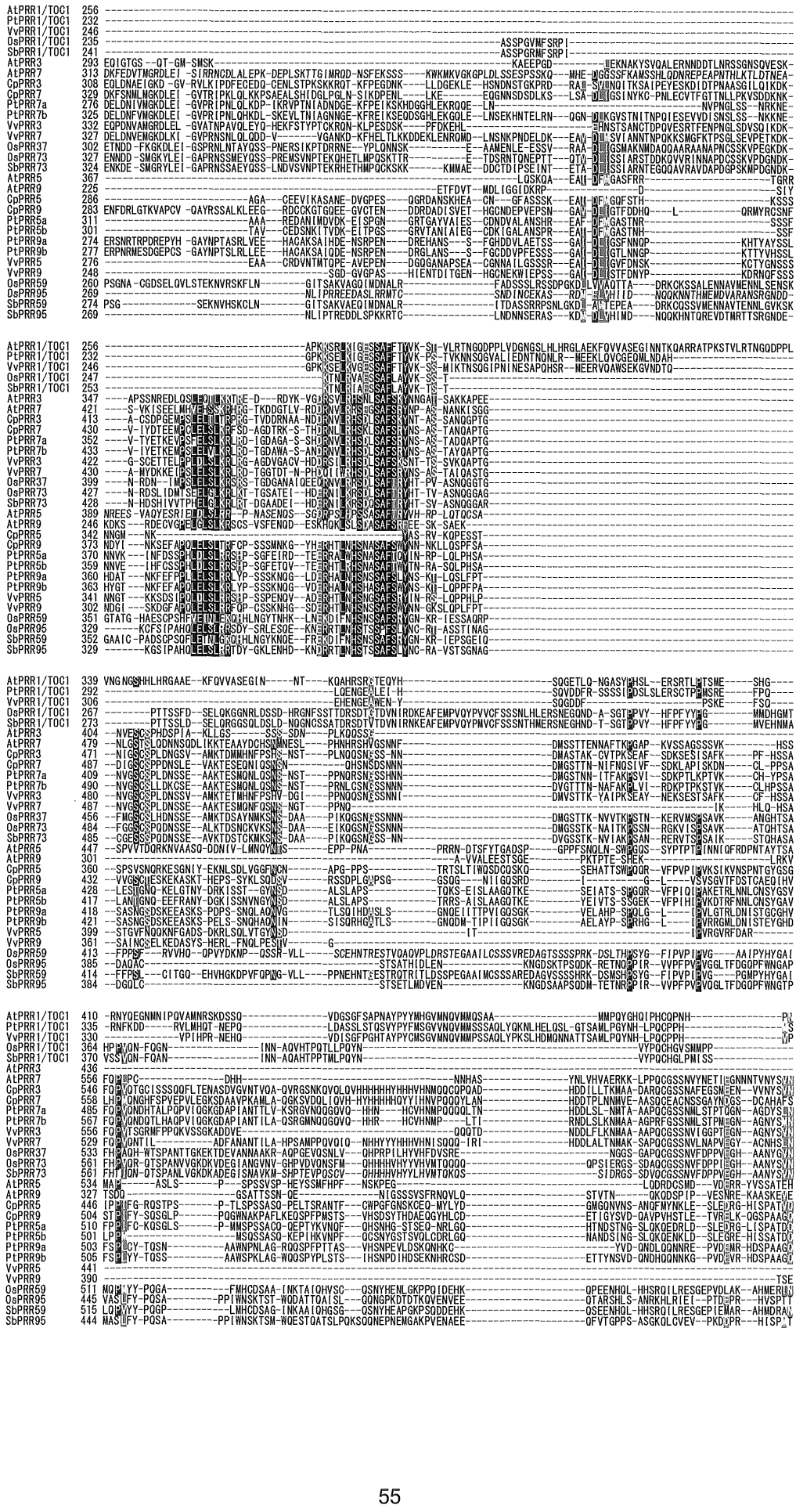
[h]Gene IDs correspond to the names obtained from Comparative Genomics Homepage (http://synteny.cnr.berkeley.edu/CoGe/).

*Genes, which appeared to be misprediced, were manually modified in this study.

**Gene is misannotated.

***Genes lack the PR-domain.


Pseudo-receiver domain



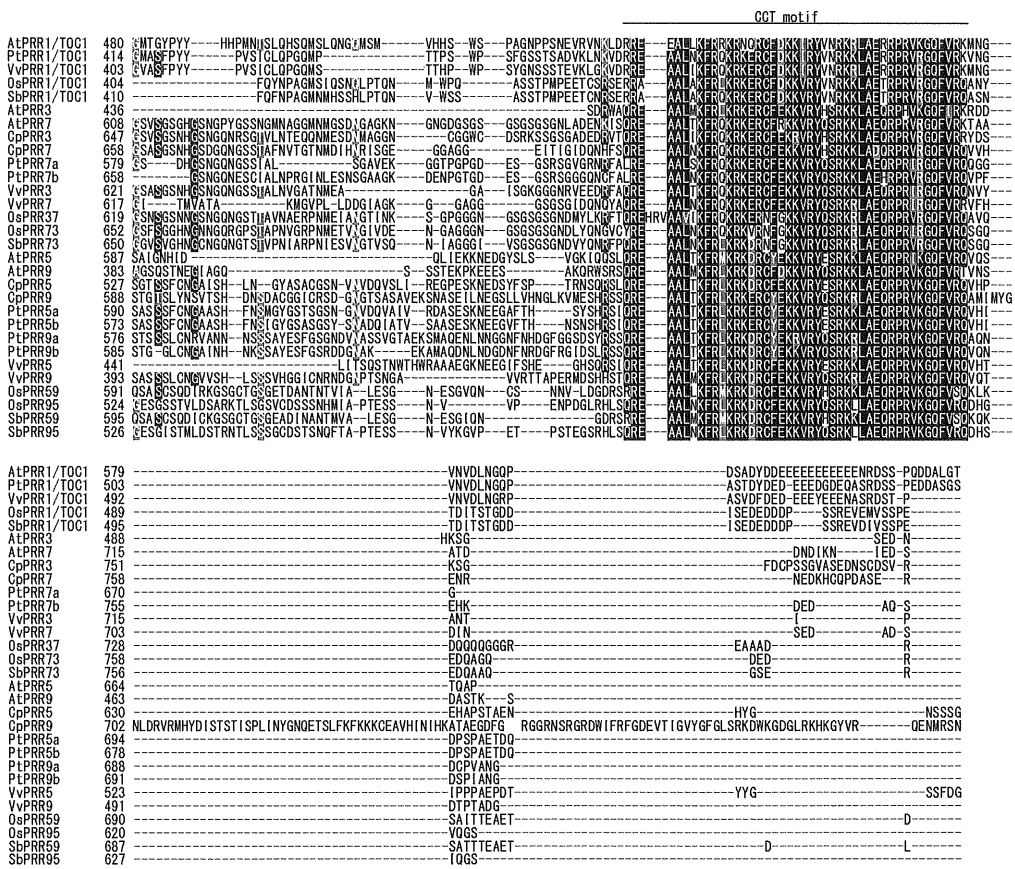


Figure 3-1. Alignments of the amino acid sequences encoded by angiosperm *PRR* genes. Amino acid sequences were aligned using TCoffee program (http://www.ebi.ac.uk/t-coffee/). Amino acid conservation was highlighted by the boxshade program (http://www.ch.embnet.org/software/BOX_form.html). Identical and similar amino acid residues are highlighted with black and gray background, respectively.

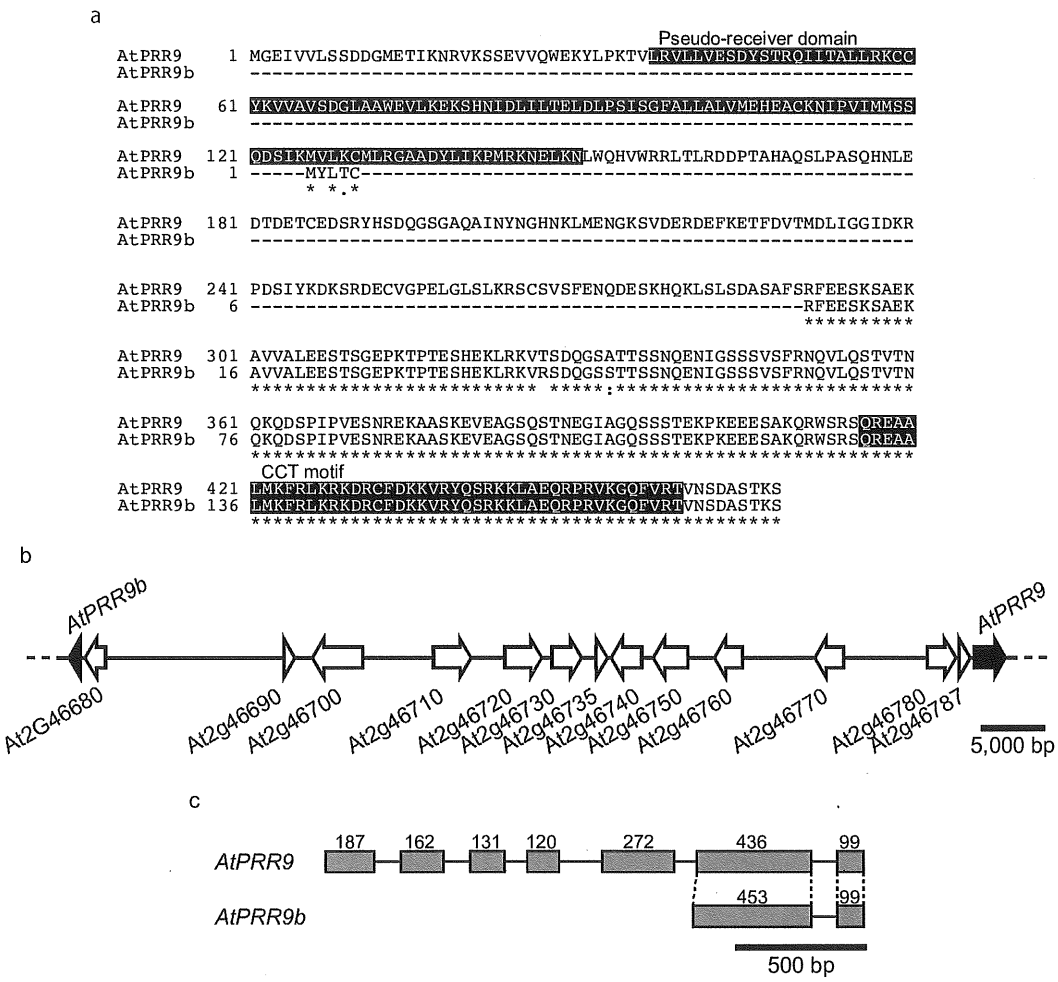


Figure 3-2. Characterization of *Arabidopsis PRR9* and *PRR9b*. (a) Alignment the amino acid sequences encoded by *AtPRR9* and *AtPRR9b*. Sequence similarity is indicated below the alignment using the symbols "asterisk," "colon," and "dot" for identical, highly similar, and weakly similar residues, respectively. Black shading indicates the PR-domain and the CCT-motif. (b) Flanking region of *AtPRR9* and *AtPRR9b*. The arrow indicates the gene with the direction. Black and white arrows indicate *AtPRR9* and *AtPRR9b* and intervening genes, respectively. (c) Schematic diagram of the exon-intron structures of *AtPRR9* and *AtPRR9b*. Gray boxes indicate ORFs. Lengths of nucleotide sequences of each ORF are indicated above the boxes.

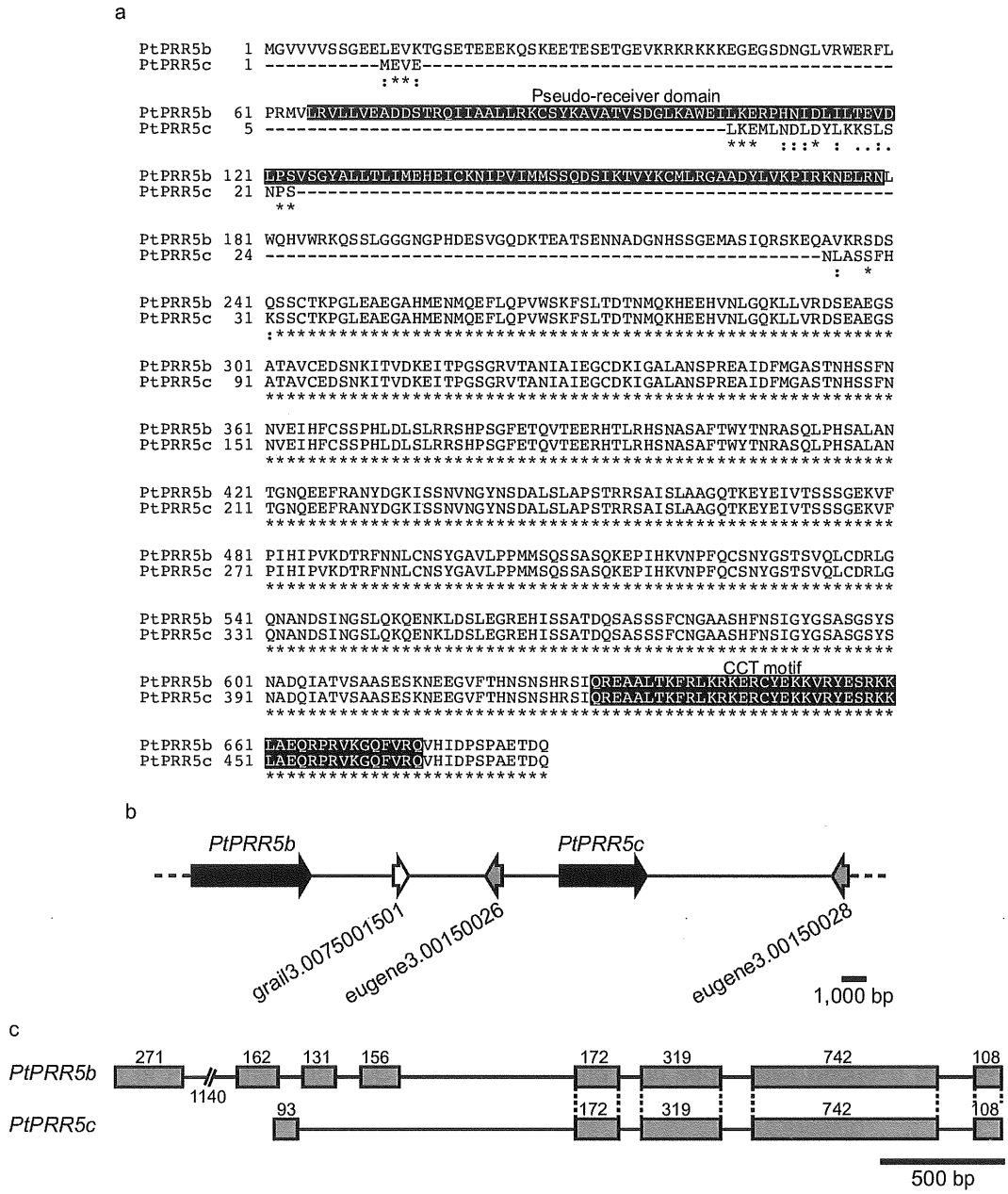


Figure 3-3. Characterization of *Populus PRR5b* and *PRR5c*. (a) Alignment of the amino acid sequences encoded by *PtPRR5b* and *PtPRR5c*. Sequence similarity is indicated below the alignment using the symbols "asterisk," "colon," and "dot" for identical, highly similar, and weakly similar residues, respectively. Black shading indicates the PR-domain and the CCT-motif. (b) Flanking region of *PtPRR5b* and *PtPRR5c*. The arrow indicates the gene with the direction. Black, gray and white arrows indicate *PtPRR5b* and *PtPRR5c*, paralogous genes and the intervening gene, respectively. (c) Schematic diagram of the exon-intron structures of *PtPRR5b* and *PtPRR5c*. Gray boxes indicate ORFs. Lengths of nucleotide sequences of each ORF are indicated above the boxes.

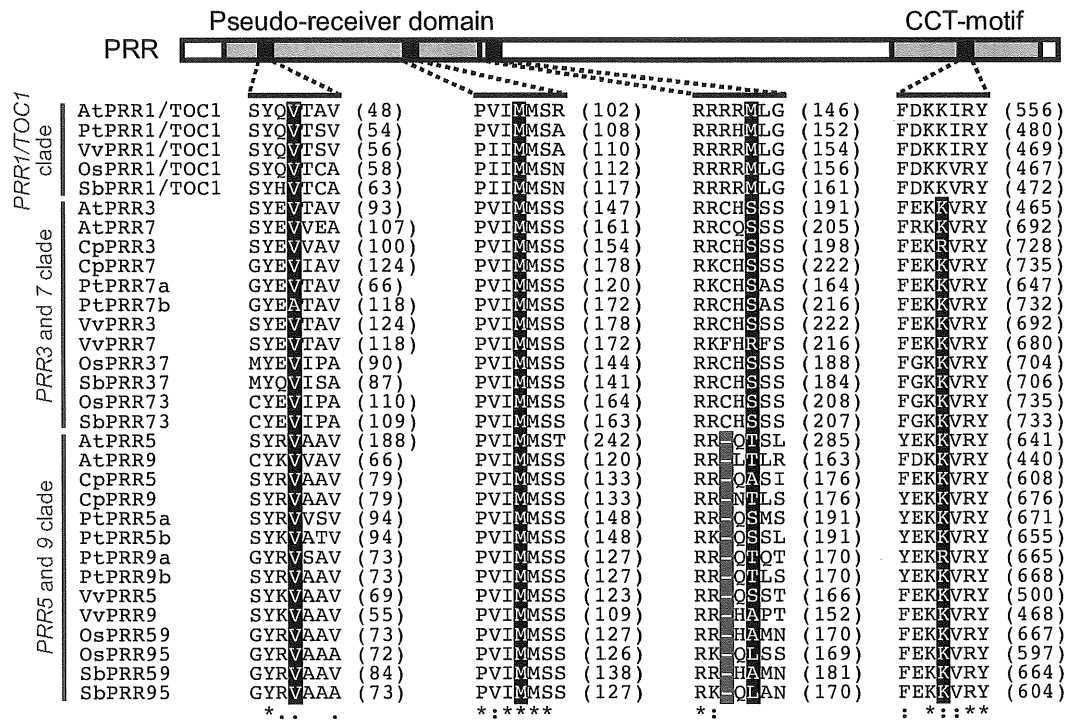


Figure 3-4. Comparison of the exon-intron structures of angiosperm *PRR* genes around the region of PR-domain and CCT-motif. The amino acid sequences encoded by *PRR* genes were aligned using TCoffee program (http://www.ebi.ac.uk/t-coffee/). The numerals on the right indicate the numbers of amino acid residues. Sequence similarity is indicated below the alignment using the symbols "asterisk," "colon," and "dot" for identical, highly similar, and weakly similar residues, respectively. Black and gray shadings on the alignments indicate a site of exon-intron boundary and one-amino acid deletion, respectively.

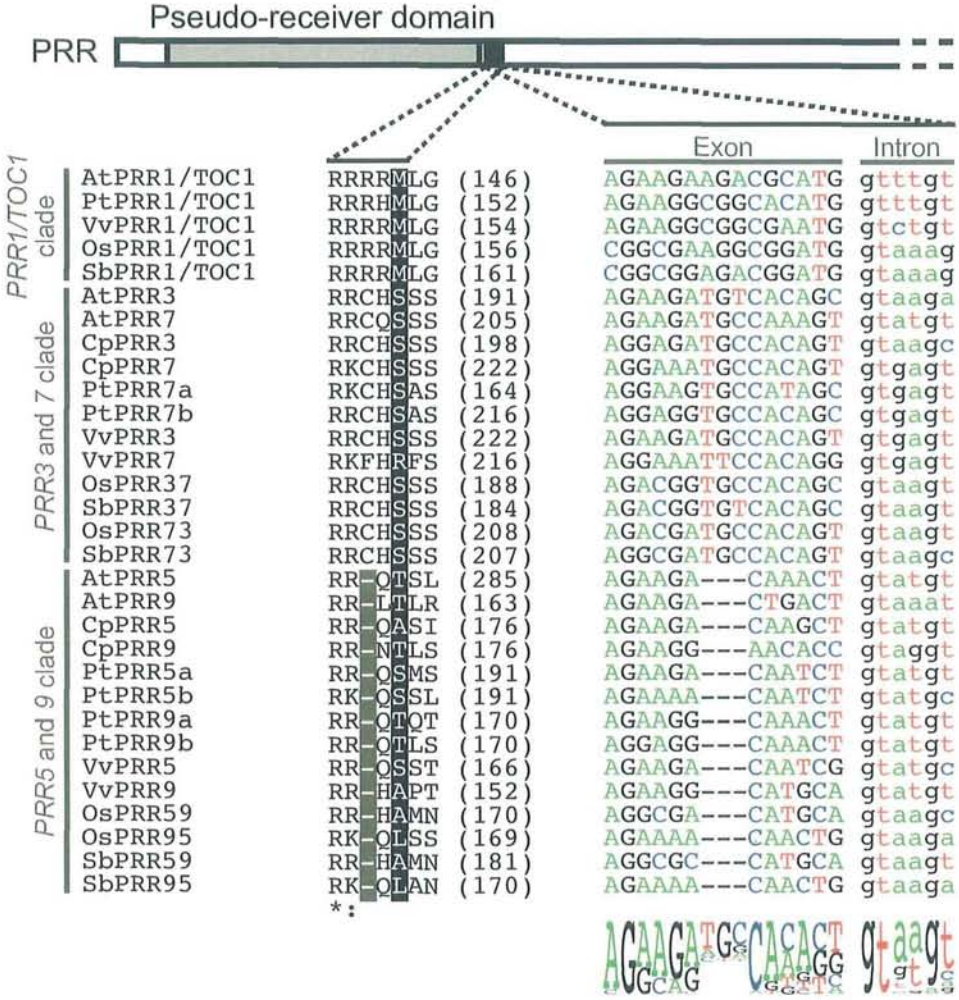


Figure 3-5. Nucleotide sequences around the region of the exon-intron boundaries of angiosperm *PRR* genes at the flanking region of PR-domain. Black and gray shadings on the alignments indicate a site of exon-intron boundary and one-amino acid deletion, respectively. Higher conservation of nucleotide sequence is shown by the bigger size of letters.

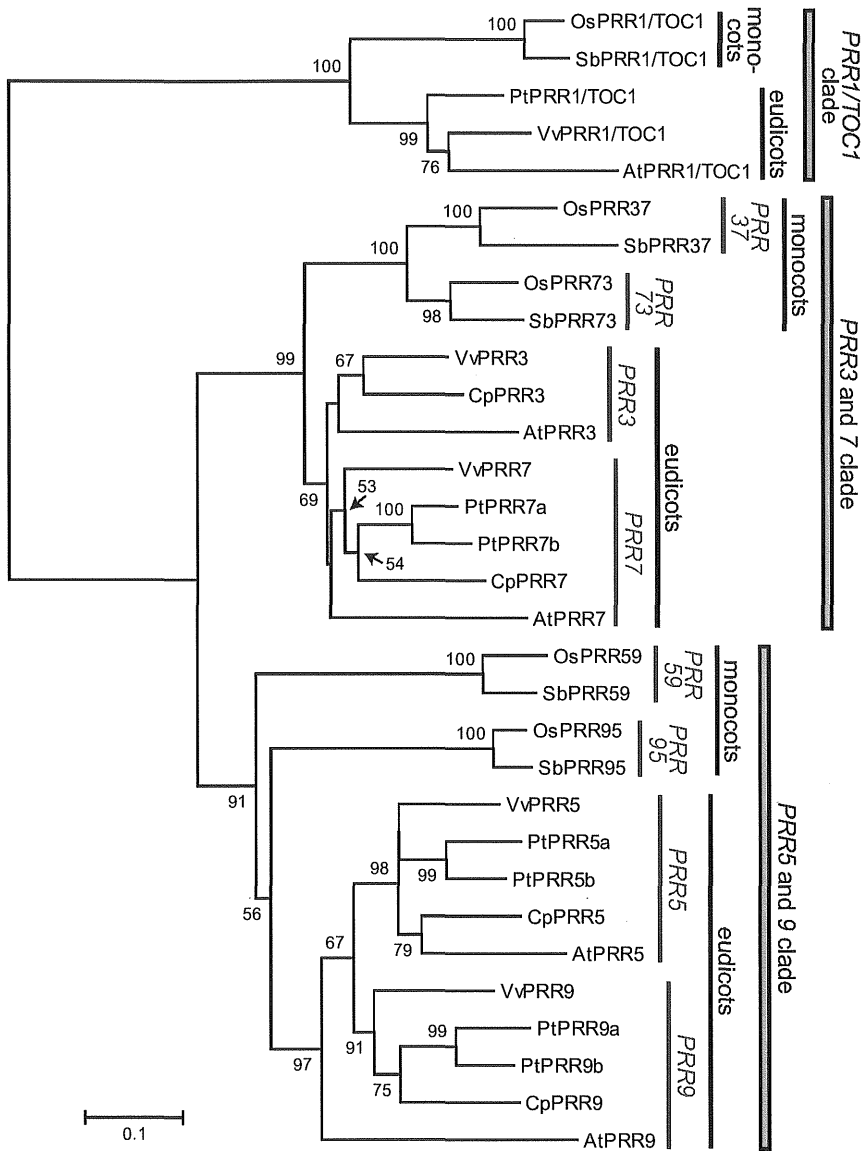


Figure 3-6. Phylogenetic tree of angiosperm *PRR* genes. Amino acid sequences were aligned using TCoffee program (http://www.ebi.ac.uk/t-coffee/). The phylogenetic tree was reconstructed by the ME method from the numbers of amino acid substitutions estimated by the JTT model. The numerals at the branch indicate bootstrap values calculated by the ME method with 1,000 replications. Bootstrap values >50% are shown.

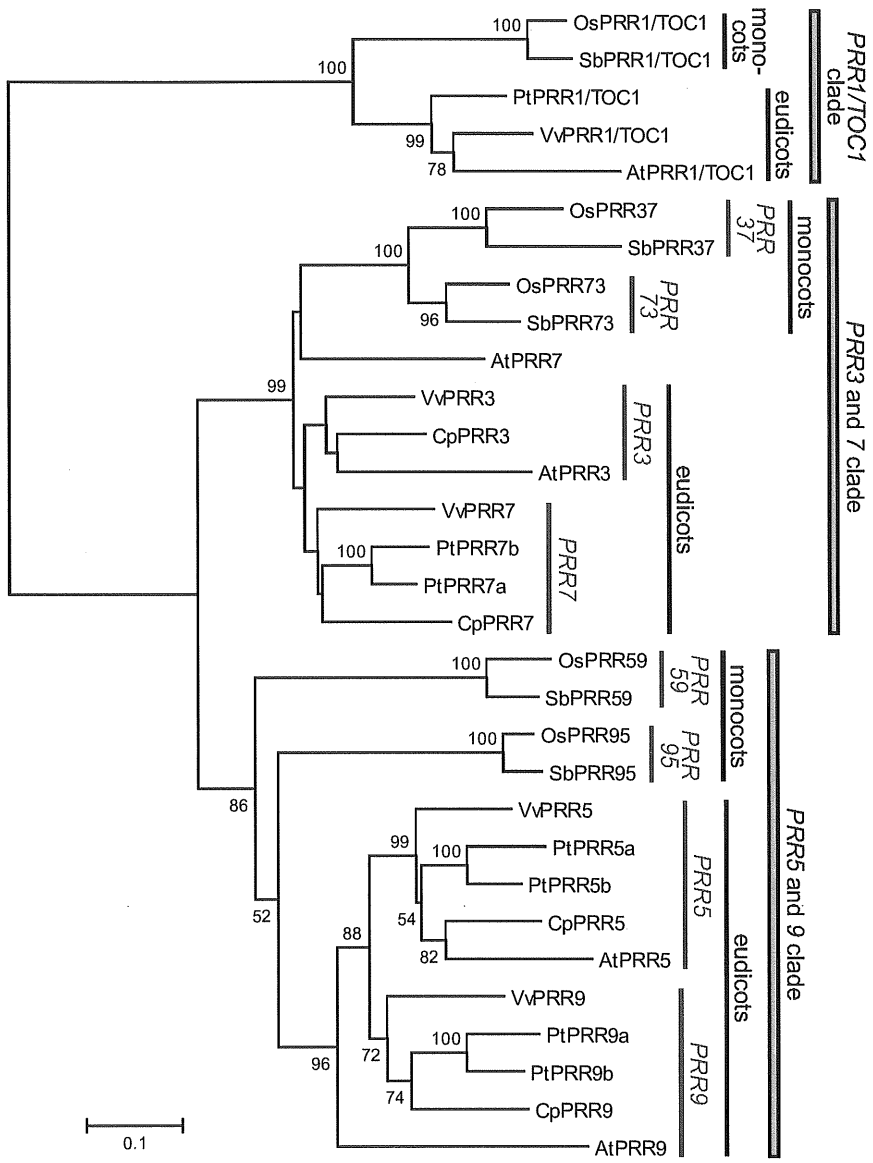


Figure 3-7. Phylogenetic tree of *PRR* genes reconstructed by the Neighbor-Joining (NJ) method. Full-length amino acid sequences were aligned using TCoffee program. The phylogenetic tree was reconstructed by the NJ method from the numbers of amino acid substitutions estimated by applying the JTT model. The numerals at the branch indicate bootstrap values calculated by the NJ method with 1,000 replications. Bootstrap values >50% are shown.

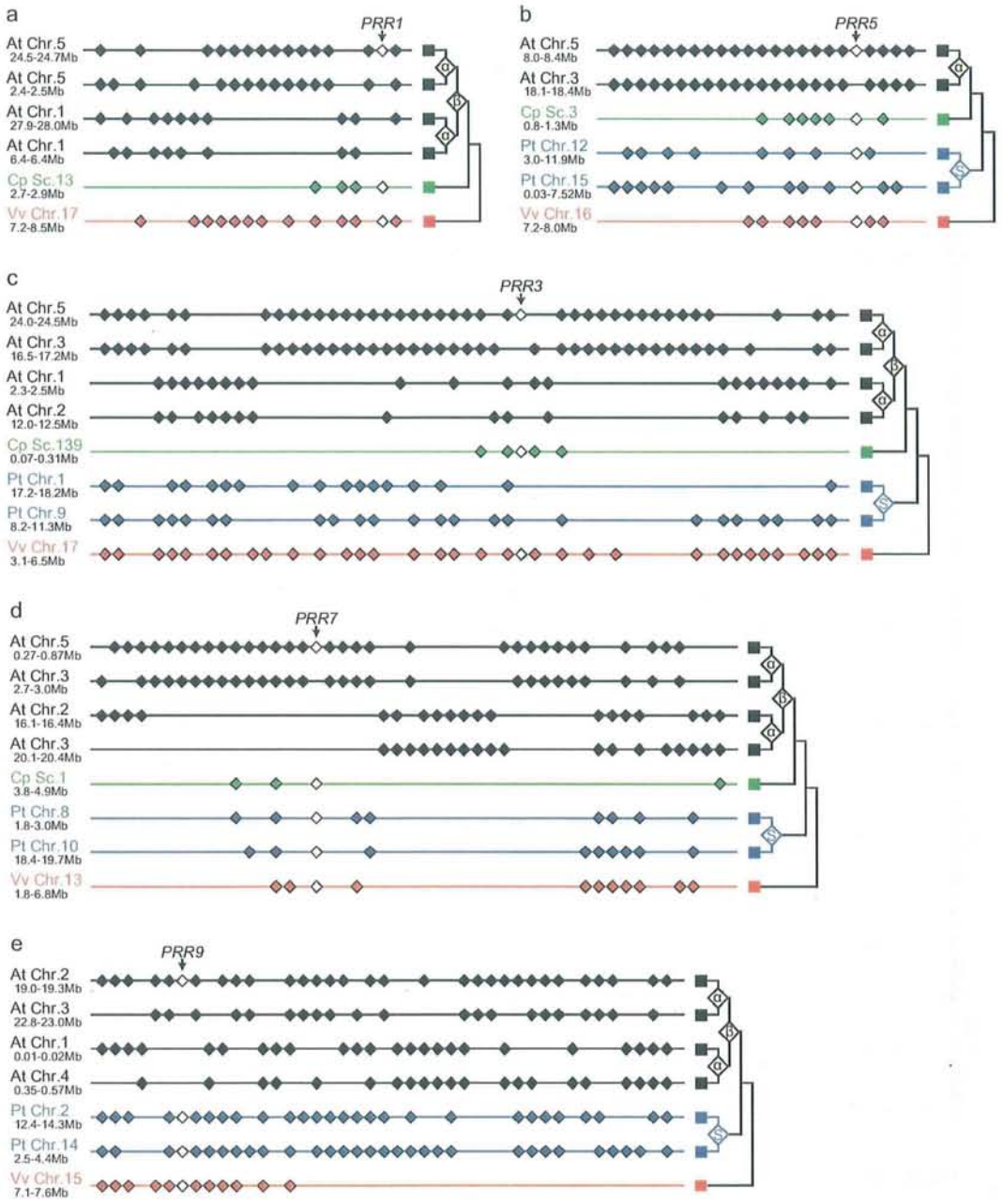


Figure 3-8. Chromosomal syntenies of flanking regions of eudicotyledonous (a) *PRR1/TOC1*, (b) *PRR5*, (c) *PRR3*, (d) *PRR7* and (e) *PRR9*. Chromosomal syntenies among eudicots were examined using the comparative genomic tool, CoGe (http://synteny.cnr.berkeley.edu/CoGe/). Syntenic relationships within each plant species were analyzed by comparative genomic tool, CoGe, and/or according to previous studies (Bower *et al.*, 2003; Jaillon *et al.*, 2007; Lyons and Freeling, 2008; Ming *et al.*, 2008; Salse *et al.*, 2008; Thomas *et al.*, 2006; Tuskan *et al.*, 2006). Boxes colored with black (*A. thaliana*), green (*C. papaya*), blue (*P. trichocarpa*) and red (*V. vinifera*) indicate individual genes. White boxes marked with arrows indicate *PRR* genes. Genes with no syntenic matches on the selected regions are not plotted. Diamonds with characters on the right side of strands indicate angiosperm polyploidy event (α, β, γ, Salicoid and ρ). The lengths of the genomic regions are shown on the left.

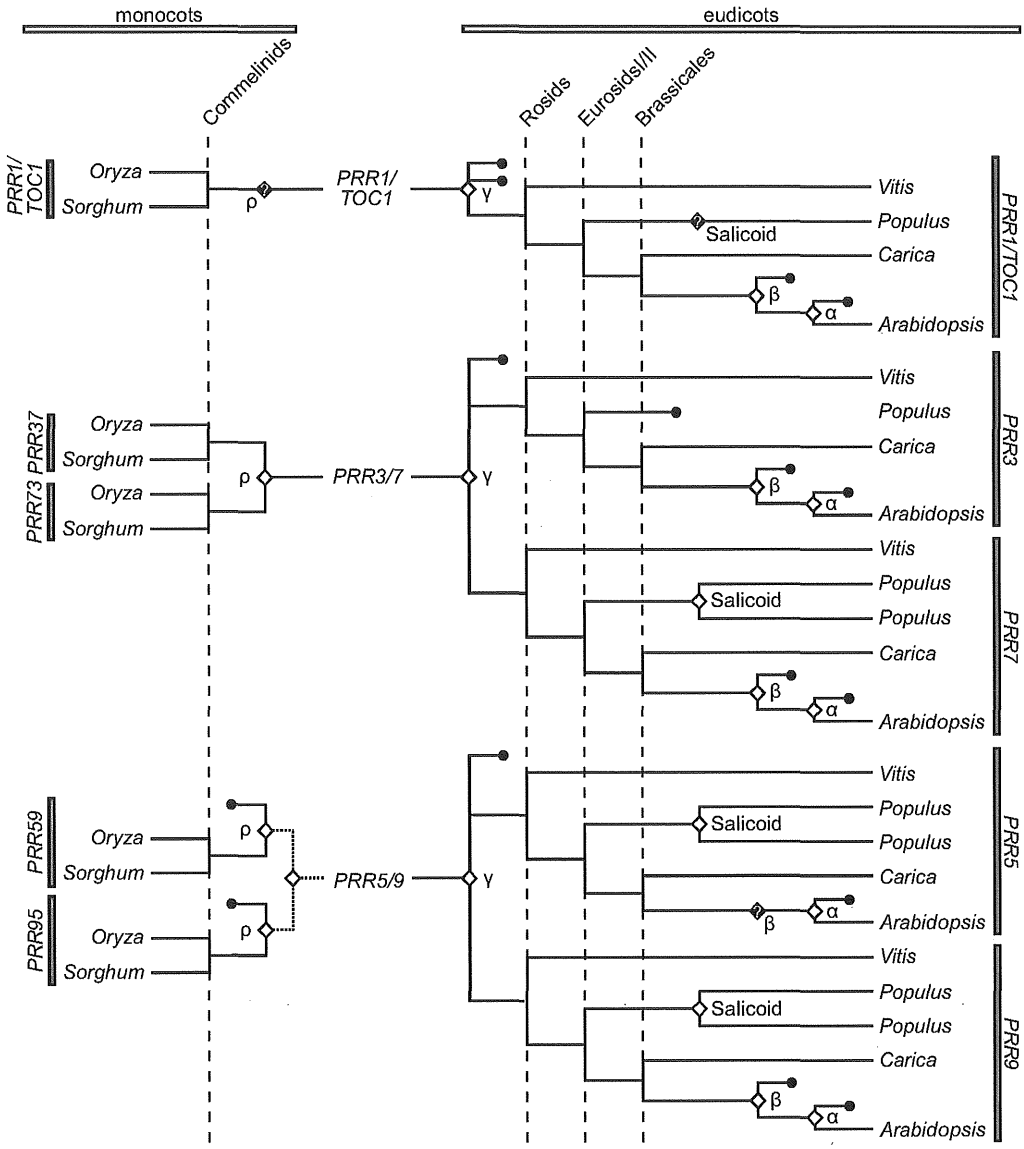


Figure 3-9. Evolutionary processes of *PRR* genes in angiosperms reconstructed by phylogenetic analysis and syntenic relationships. Chromosomal syntenies among eudicots or monocots were examined using the comparative genomic tool, CoGe (http://synteny.cnr.berkeley.edu/CoGe/) or VISTA Browser (http://genome.lbl.gov/vista/index.shtml). Chromosomal syntenies within each plant species were analyzed by comparative genomic tool, CoGe, and/or according to previous studies (Bower *et al.*, 2003; Jaillon *et al.*, 2007; Lyons and Freeling, 2008; Ming *et al.*, 2008; Salse *et al.*, 2008; Thomas *et al.*, 2006; Tuskan *et al.*, 2006). Diamonds and circles indicate gene duplication and gene loss event, respectively. The timing of a gene duplication event that is not clear in the previous studies is shown by dotted line. Black diamonds with question mark indicate that a gene duplication event derived from a polyploidy event is not resolved. Timings of plant speciation are described by broken lines.

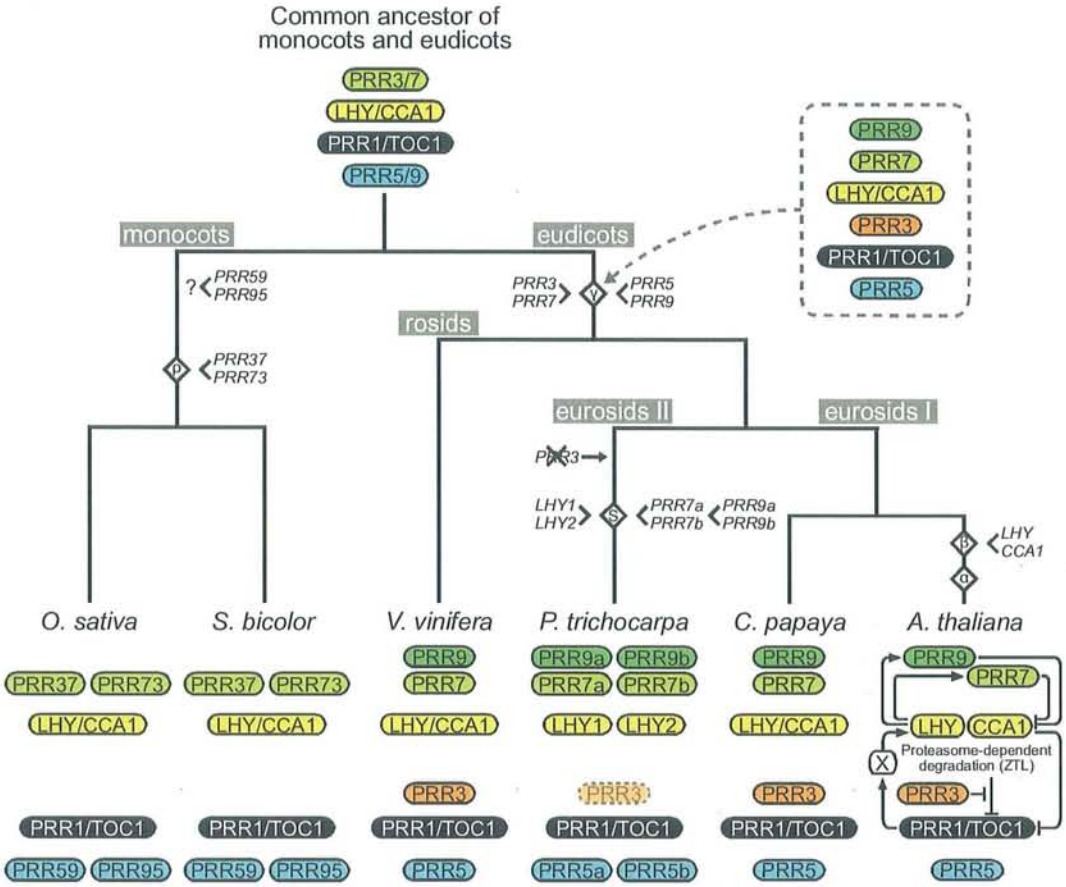


Figure 3-10. Schematic diagram of the evolutionary process of the plant circadian clock system. For the plant clock system model, loop I and III are described in this diagram. Diamonds with characters indicate angiosperm polyploidy event (α, β, γ, Salicoid and ρ).