# The roles of DNA repair and epigenetic regulation in plant longevity: Systematic comparisons of copy number variation of genes and seasonal gene expression dynamics

青柳, 優太

# The roles of DNA repair and epigenetic regulation in plant longevity: Systematic comparisons of copy number variation of genes and seasonal gene expression dynamics

（植物の寿命における DNA 修復とエピジェネティック制御の役割：
遺伝子コピー数の網羅的な比較解析、及び野外に生育する樹木の
遺伝子発現動態の解析）

**Yuta Aoyagi**

Submitted to the faculty of the Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in Science Kyushu University

January 2022

## Contents

# Preface

**Longevity: Why do trees have the capacity to live for an extraordinary long time?**

Lifespans of organisms are widely diverse across species, and some organisms can live for hundreds or even thousands of years. For example, in animals, the longest lifespan of Greenland shark (*Somniosus microcephalus*) is estimated to be 392 years from a chronology obtained from eye lens nuclei (Nielsen et al., 2016), and the oldest recorded age of Aldabra giant tortoise (*Aldabrachelys gigantea*) is 152 years old (Castanet, 1994). Human also have long maximum lifespan and the longest lifespan is that of Jeanne Calment, a French who lived to age 122 years (Allard, 1998). In plants, trees generally have a long lifespan, and some trees can live for an extraordinary long time. Japanese Jomon cedar, a famous long-lived tree in Yakushima, Japan, is estimated to live for 2170 years (Yakusugi Musium, http://www.yakusugi-museum.com/), and the age of longest-lived bristlecone pine (*Pinus longaeva*), one of the longest-lived trees, is estimated to be 4713 years (Lanner and Connor, 2001). Why do trees have the capacity to live for an extraordinary long time? What are the mechanisms underlying great longevity is a central question in life science.

**The theory of aging and longevity**

Various theories about aging and longevity of organisms have been proposed and examined (Review in Hayflick 1985; Semsei 2000; Weinert and Timiras 2003; Jin 2010). Most of the theories can be categorized into two categories: program theories and error theories. According to the program theories, the aging process and longevity are genetically controlled, just as the development and growth. The program theories

include endocrine theory, which proposed that the biological clocks controlled the regulation of genes involved in hormones, development, and immune system, controlling the pace of aging; Limited number of proliferation theory (Hayflick, 1965), which proposed that there is a specific limitation on the number of cell divisions and the organismal lifespans are determined by the number of cell divisions. The error theories of aging imply the aging is caused by the accumulation of errors (damages) at various levels (e.g., DNA damage, oxygen radicals accumulation, cross-linking in protein) through the lifespan. The error theories include free radical theory (Gerschman, 1954; Harman, 1955) , which proposed that free radicals and reactive oxygen species damages molecular components such as DNA and proteins and the accumulation of such damages causes cellular dysfunctions, resulting aging; Cross-linking theory (Bjorksten, 1942; Kohn, 1978) , which proposed that the accumulation of cross-linking proteins damages cells, slowing down bodily processes resulting in aging; DNA damage/mutation accumulation theory (Failla 1958; Szilard 1959; Gensler and Bernstein 1981), which proposed that the accumulation of DNA damages and mutations causes the functional decline and the disruption of homeostasis, resulting in aging. In this thesis, I focused on the error theories of aging and longevity and how long-lived trees deal with errors/damages and survive for a long time.

**Accumulation of damage and aging and longevity**

Living organisms are exposed to many endogenous and exogenous stresses on a daily basis. For example, ultraviolet (UV) radiation, high/low temperature and drought as an abiotic stress and pathogen infection and herbivory as a biotic stress. Such stresses can cause damage at various levels (e.g., DNA damage, alteration of epigenetic state) (Pal &

Tyler, 2016; Yousefzadeh et al., 2021). Accumulation of DNA damage and somatic mutations disrupts genome integrity and causes genetic and cellular dysfunctions, enhancing aging. There are many types of diseases that show signs of accelerated aging and short lifespan, such as Werner syndrome and ataxia telangiectasia (Martin & Oshima, 2000). Werner syndrome is caused by mutations in *WRN* gene encoding RecQ DNA helicase protein (Yu et al., 1996), which involved in several biological processes, such as DNA replication (Sidorova et al., 2008) and recombination (Hu et al., 2007). Patients of Werner syndrome show accumulation of DNA double-strand breaks in cells (Ariyoshi et al., 2007). In addition, alteration of epigenetic states disrupts homeostasis. Epigenetic regulation is involved in vital biological processes, such as the regulation of gene expression (Busslinger, 1983; Grunstein, 1997), DNA replication (Zhang et al., 2000), DNA repair (Shim et al., 2005), and the inhibition of exogenous genetic elements (Al-Kaff et al., 1998). Loss of DNA methylation leads to activation of silenced DNA sequences, resulting in the activation of transposable elements and abnormal expression of genes (Pal & Tyler, 2016). These suggest that accumulation of DNA damage and alteration of epigenetic states due to stresses relates to aging and longevity. Therefore, it is supposed that functions to suppress such damage more developed in long-lived organisms, such as long-lived trees, than in short-lived organisms.

**DNA repair and epigenetic regulation in longevity**

A growing number of studies have explored the relationships between DNA repair and epigenetic regulation and longevity, especially in animals. The naked mole-rat, the longest-lived rodent, has higher copy numbers of genes for CCAAT/enhancer binding protein-g (*CEBPG*), a regulator of DNA repair, compared to more short-lived species

(MacRae et al., 2015). Analyses of genomes of other long-lived species, the bowhead whale and bat, showed the signature of positive selection of multiple DNA repair and DNA damage signaling genes (Zhang et al., 2013; Keane et al., 2015). Sirtuins, $NAD^+$-dependent histone deacetylases, are involved in the regulation of many metabolic functions, including DNA repair, genome stability, inflammatory responses, apoptosis, the cell cycle, and mitochondrial functions (Wątroba & Szukiewicz, 2016). Overexpression or activation of Sir2 homologs extends the lifespan of worms (*Caenorhabditis elegans*) (Tissenbaum & Guarente, 2001) and fruit flies (*Drosophila melanogaster*) (Rogina & Helfand, 2004). These studies in animals suggest the importance of DNA repair and epigenetic regulation for longevity.

**Aim of this study**

Despite the wealth of studies in animals, systematic comparisons to explore DAN repair and epigenetic regulation associated with longevity across species with different lifespans are not sufficiently represented in plants. To understand the relationships between DNA repair and epigenetic regulation and longevity in organisms, it is also necessary to analyze plants, which include diverse species with a wide range of lifespans, from annual herbs with short lifespans less than one year to perennial herbs and trees with long lifespans. Therefore, in this thesis, we focused on the copy number variation in genes and gene expression to response to environmental stress in plants, and performed systematic comparative analyses of copy number variation of genes associated with DNA repair and epigenetic regulation using a genome database (chapter 1 and chapter 2) and seasonal expression dynamics of DNA repair and epigenetic

regulatory genes among trees under natural conditions (chapter 3). I summarize the contents for each chapter as follows.

**Chapter 1: Copy number analyses of DNA repair genes reveal the role of poly(ADP-ribose) polymerase (PARP) in tree longevity**

Using the recent accumulation of the complete genome sequences of diverse plant species, we performed systematic comparative analyses of the copy number variations of DNA repair gene families in 61 plant species with different lifespans. Among 121 DNA repair gene families, *PARP* gene family was identified as a unique gene that exhibits significant expansion in trees compared to annual and perennial herbs. Among three paralogs of plant *PARP*s, *PARP1* showed a close association with growth rate. PARPs catalyze poly(ADP-ribosyl)ation and play pivotal roles in DNA repair and antipathogen defense. Our study suggests the conserved role of PARPs in longevity between plants and animals.

**Chapter 2: Analyses of copy number variation in epigenetic regulatory genes across plants: Increased copy numbers of *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) and *SILENCING DEFECTIVE 3* (*SDE3*) in long-lived trees**

To identify the epigenetic regulatory genes with increased copy number in long-lived tree species than in short-lived annual and perennial herb species, we conducted systematic comparisons of copy number variation in 121 gene families involved in various epigenetic regulatory pathways across 85 plant species with different lifespans using a genome database. Among 121 epigenetic regulatory gene families, the gene family encoding *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) and that

encoding *SILENCING DEFECTIVE 3* (*SDE3*) were found to exhibit significantly higher copy number of genes in tree species than in both perennial and annual herb species. BRU1/TSK/MGO3 is involved in chromatin modifications and plays an important role in the maintenance of meristems, genome integrity, and the inheritance of chromatin states. SDE3 is involved in RNA silencing and has an important role in antiviral defense through posttranscriptional gene silencing. Increasing copy numbers of *BRU1/TSK/MGO3* and *SDE3* genes are likely to be favored in the maintenance of meristems, genome integrity, the inheritance of chromatin states, and antiviral defense in long-lived trees, and these factors would contribute to survival over a long lifespan.

## Chapter 3: Seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation in *Quercus glauca* and *Lithocarpus edulis* under natural conditions

Living organisms are exposed many types of stresses including biotic and abiotic stresses. To suppress damage due to stresses and maintain to survive for a long time, it is necessary to respond appropriately to stresses that change over time. In the present study, to examine and compare the seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation, we analyzed time-series transcriptome data collected throughout about two years from individuals of different tree species, *Quercus glauca* and *Lithocarpus edulis*, growing in natural environments. The present study demonstrated similar and different seasonal expression dynamics of DNA repair genes and epigenetic regulatory genes among species. Results of the present study suggest that a large number of genes associated with DNA repair and epigenetic regulation exhibit similar seasonal expression patterns among species. In addition, genes with different seasonal expression

dynamics are associated with multiple functions and involved in plant development, growth, and reproduction, which is likely to reflect the difference in vegetative and reproductive schedules among species.

**References**

Al-Kaff, N.S., Covey, S.N., Kreike, M.M., Page, A.M., Pinder, R., Dale, P.J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science*, 279, 2113 2115.

Allard, M., Lèbre, V. and Robine, J.M. (1998). Jeanne Calment. From Van Gogh's time to ours. New York: W.H. Freeman Press.

Ariyoshi, K., Suzuki, K., Goto, M., Watanabe, M., & Kodama, S. (2007). Increased chromosome instability and accumulation of DNA double-strand breaks in Werner syndrome cells. *Journal of Radiation Research*, *48*(3), 219–231. https://doi.org/10.1269/jrr.07017

Bjorksten, J. (1942). Chemistry of duplication. *Chem. Industries*, 49, p. 2

Busslinger, M., Hurst, J., & Flavell, R. A. (1983). DNA methylation and the regulation of globin gene expression. *Cell*, 34(1), 197–206. https://doi.org/https://doi.org/10.1016/0092-8674(83)90150-2

Castanet, J. (1994). Age estimation and longevity in reptiles. *Gerontology*, 40(2–4), 174–192. https://doi.org/10.1159/000213586

Failla, G. (1958), THE AGING PROCESS AND CANCEROGENESIS. *Annals of the New York Academy of Sciences*, 71: 1124-1140. https://doi.org/10.1111/j.1749-6632.1958.tb54674.x

Gensler HL, Bernstein H. (1981). DNA damage as the primary cause of aging. *The Quarterly Review of Biology*. Sep;56(3):279-303. DOI: 10.1086/412317. PMID: 7031747.

Gerschman, R., Gilbert, D. L., Nye, S. W., Dwyer, P., & Fenn, W. O. (1954). Oxygen poisoning and X-irradiation: A mechanism in common. *Science*, 119(3097), 623–626. https://doi.org/10.1126/science.119.3097.623

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649), 349–352. https://doi.org/10.1038/38664

Harman, D. (1955). Lawrence Berkeley National Laboratory Recent Work Title AGING: A THEORY BASED ON FREE RADICAL AND RADIATION CHEMISTRY. *Lawrence Berkeley National Laboratory*. https://escholarship.org/uc/item/3w86c4g7

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, 37(3), 614–636. https://doi.org/10.1016/0014-4827(65)90211-9

Hayflick, L. (1985). Review Article Theories of Biological Aging. *Experimental Gerontology*, 20, 145–159.

Hu, Y., Raynard, S., Sehorn, M. G., Lu, X., Bussen, W., Zheng, L., Stark, J. M., Barnes, E. L., Chi, P., Janscak, P., Jasin, M., Vogel, H., Sung, P., & Luo, G. (2007). RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes and Development*, *21*(23), 3073–3084. https://doi.org/10.1101/gad.1609107

Jin, K. (2010). Modern biological theories of aging. *Aging and Disease*, 1(2), 72–74. https://doi.org/10.1093/jn/119.6.952

Keane, M. *et al.* (2015) Insights into the evolution of longevity from the bowhead whale genome, *Cell Reports*, 10(1), pp. 112–122. doi: 10.1016/j.celrep.2014.12.008.

Kohn, R.R.; ed. (1978) Principals of mammalian aging. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Lanner, R. M., & Connor, K. F. (2001). Does bristlecone pine senesce? *Experimental Gerontology*, 36(4–6), 675–685. https://doi.org/10.1016/S0531-5565(00)00234-5

MacRae, S. L. *et al.* (2015) Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human, *Aging Cell*, 14(2), pp. 288–291. doi: 10.1111/acel.12314.

Martin, G. M., & Oshima, J. (2000). Lessons from human progeroid syndromes. *Nature*, *408*(6809), 263–266. https://doi.org/10.1038/35041705

Nielsen, J., Hedeholm, R. B., Heinemeier, J., Bushnell, P. G., Christiansen, J. S., Olsen, J., Ramsey, C. B., Brill, R. W., Simon, M., Steffensen, K. F. et al. (2016). Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (Somniosus microcephalus). *Science*, 353(6300), 702–704. https://doi.org/10.1126/science.aaf1703

Pal, S., & Tyler, J. K. (2016). Epigenetics and aging. *Science advances*, 2(7), e1600584. https://doi.org/10.1126/sciadv.1600584

Rogina, B., & Helfand, S. L. (2004). Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), 15998–16003. https://doi.org/10.1073/pnas.0404184101

Semsei, I. (2000). On the nature of aging. *Mechanisms of Ageing and Development*, 117(1–3), 93–108. https://doi.org/10.1016/S0047-6374(00)00147-0

Shim, E. Y., Ma, J.-L., Oum, J.-H., Yanez, Y., & Lee, S. E. (2005). The Yeast Chromatin Remodeler RSC Complex Facilitates End Joining Repair of DNA Double-Strand Breaks. *Molecular and Cellular Biology*, 25(10), 3934–3944. https://doi.org/10.1128/mcb.25.10.3934-3944.2005

Sidorova, J. M., Li, N., Folch, A., & Monnat, R. J. (2008). The RecQ helicase WRN is required for normal replication fork progression after DNA damage or replication fork arrest. *Cell Cycle*, *7*(6), 796–807. https://doi.org/10.4161/cc.7.6.5566

Szilard L. (1959). ON THE NATURE OF THE AGING PROCESS. *Proceedings of the National Academy of Sciences of the United States of America*, 45(1), 30–45. https://doi.org/10.1073/pnas.45.1.30

Tissenbaum, H. A., & Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. *Nature*, 410(6825), 227–230. https://doi.org/10.1038/35065638

Wątroba, M., & Szukiewicz, D. (2016). The role of sirtuins in aging and age-related diseases. *Advances in Medical Sciences*, 61(1), 52–62. https://doi.org/10.1016/j.advms.2015.09.003

Weinert, B. T., & Timiras, P. S. (2003). Invited review: Theories of aging. *Journal of Applied Physiology*, 95(4), 1706–1716. https://doi.org/10.1152/japplphysiol.00288.2003

屋久杉自然館, http://www.yakusugi-museum.com/

Yousefzadeh, M., & Henpita, C. (2021). DNA damage-how and why we age? *ELife*, *10*, 1–17. https://pubmed.ncbi.nlm.nih.gov/33512317/

Yu, C. E., Oshima, J., Fu, Y. H., Wijsman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan, J., & Schellenberg, G. D. (1996). Positional cloning of the Werner's syndrome gene. *Science*, 272(5259), 258–262. https://doi.org/10.1126/science.272.5259.258

Zhang, Z., Shibahara, K. I., & Stillman, B. (2000). PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature*, 408(6809), 221–225. https://doi.org/10.1038/35041601

Zhang, G. *et al.* (2013) Comparative Analysis of Bat Genomes, *Science*, 339(January), pp. 456–460.

## Acknowledgments

I would like to thank my supervisor Professor Akiko Satake, in Mathematical Biology Laboratory, Kyushu University, for her great support and encouragement on my research activities. I also thank her for advising me on my carrier and life. I appreciate for helpful comments and discussions on my work given by Associate Professor Eriko Sasaki, in Mathematical Biology Laboratory, Kyushu University. For the study in Chapter 1, I would like to thank Associate Professor Junko Kusumi, in Department of Environmental Changes, Kyushu University, for help with analysis and helpful comments and discussion as a research collaborator; Assistant Professor Mizuki Ohno, in Department of Medical Biophysics and Radiation Biology, Kyushu University, for many insightful comments. I am also thankful to the members and alumni of Mathematical Biology Laboratory at Kyushu University, for stimulating discussion from the points of view of their various interests. Especially, I am grateful to Dr. Quenta Araye, Dr. Ryosuke Imai and Dr. Akane Hara for discussing interesting ideas with me and advising me on my future carrier. I would like to thank my family for giving me an opportunity and a lot of support to study; I could not have continued my research without their support. Finally, this work was financially supported by JSPS KAKENHI (JP26251042; JP17H06478; JP17H01449) to Professor Akiko Satake.

# Chapter 1: Copy number analyses of DNA repair genes reveal the role of poly(ADP-ribose) polymerase (PARP) in tree longevity

The study in this chapter, done in collaboration with Professor Akiko Satake and Associate Professor Junko Kusumi, was published in *iScience*, Volume 24, Issue 7 on 23 July 2021

**ABSTRACT**

Long-lived organisms are exposed to the risk of accumulating mutations due to DNA damage. Previous studies in animals have revealed the positive relationship between the copy number of DNA repair genes and longevity. However, the role of DNA repair in the lifespan of plants remains poorly understood. Using the recent accumulation of the complete genome sequences of diverse plant species, we performed systematic comparative analyses of the copy number variations of DNA repair genes in 61 plant species with different lifespans. Among 121 DNA repair gene families, *PARP* gene family was identified as a unique gene that exhibits significant expansion in trees compared to annual and perennial herbs. Among three paralogs of plant *PARP*s, *PARP1* showed a close association with growth rate. PARPs catalyze poly(ADP-ribosyl)ation and play pivotal roles in DNA repair and antipathogen defense. Our study suggests the conserved role of PARPs in longevity between plants and animals.

**INTRODUCTION**

Organisms accumulate DNA damage via exogenous environmental factors (e.g., ionizing radiation and UV light) and constant threats to the endogenous metabolic process (e.g., production of reactive oxygen species and errors in DNA metabolism). DNA lesions commonly include oxidized or alkylated base damage, single- and double-strand breaks, intra- or inter-strand crosslinks, and base loss. The resulting alteration of the DNA structure leads to genomic instability, apoptosis, or senescence, which can affect the organism's development and aging process. To reverse the potentially deleterious damage, life in all its forms has evolved sophisticated machinery, involving hundreds of proteins, to efficiently recognize and properly repair DNA damage.

Depending on the type of DNA lesion, organisms have developed diverse functional pathways for DNA repair (Sancar et al., 2004; Ciccia and Elledge, 2010). The base excision repair (BER) and direct damage reversal/repair (DR) pathways repair DNA base damage, whereas mismatch repair (MMR) corrects base mispairs and small loops often found in repetitive sequence DNA. More complex lesions, such as pyrimidine dimers and intrastrand crosslinks, are corrected by nucleotide excision repair (NER). Double-strand breaks (DSBs) are repaired either by non-homologous end-joining (NHEJ) or homologous recombination (HR). These major functional pathways for DNA repair have been identified in virtually all organisms, including bacteria, archaea, and eukaryotes, reflecting the universal need to counter DNA damage in living organisms (Aravind et al., 1999; Eisen and Hanawalt, 1999).

With the recent accumulation of the complete genome sequences of diverse organisms, it has become possible to systematically compare the DNA repair systems of the respective organisms and identify the origins of the different repair genes and functional pathways. A global comparative analysis of DNA repair proteins based upon the available complete genome sequences of bacteria, archaea, and eukaryotes has shown that repair machinery shows considerable diversity in terms of the presence and absence of genes. Eisen and Hanawalt (1999) showed that only DR pathways are highly homologous between species (they make use of homologous genes in all species), whereas other pathways are not homologous, with the use of genes of different origins between species despite performing the same functions.

The diversity of repair machinery among species can be formed by frequent gene duplication and gene loss. Members of the *recA/RAD51* gene family, which is associated with HR, are suggested to be generated by multiple duplication events (one

before the archaea/eukaryote split and another in the early stage of eukaryotic evolution), gene loss, and endosymbiotic gene transfer (Lin et al., 2006). A study based on angiosperm genomes reported the strong selection pressure to preserve many of the DNA repair genes as singletons in *Arabidopsis thalian*a, regardless of repeated whole genome or single gene duplication events in flowering plants (De Smet et al., 2013). The species-specific history of gene duplication and loss will result in copy number variations of DNA repair genes among species, which can have profound effect on organismal phenotypes, including mutation rates (Baer et al., 2007), lifespan (Lorenzini et al., 2009; Freitas and De Magalhães, 2011), and adaptation to extreme environments (Matic et al., 1995; White et al., 1999).

Previous studies focused on aging have highlighted the positive correlation of an increased copy number of DNA repair genes and longevity in mammals (Tian et al., 2017). The naked mole-rat, the longest-lived rodent, has higher copy numbers of genes for CCAAT/enhancer binding protein-g (*CEBPG*), a regulator of DNA repair, and TERF1-interacting nuclear factor 2 (*TINF2*), a protector of telomere integrity compared to more short-lived species (MacRae et al., 2015a). Another long-living mammal, the African elephant, encodes 20 copies of the tumor suppressor gene, *TP53*, which induce apoptosis or senescence programs in response to DNA damage (Sulak et al., 2016). Analyses of genomes of other long-lived species, the bowhead whale and bat, showed the signature of positive selection of multiple DNA repair and DNA damage signaling genes (Zhang et al., 2013a; Keane et al., 2015). These studies in mammals suggest the importance of genome maintenance mechanisms for longevity.

Despite the wealth of studies in animals, there are no studies that employ comparative genome analyses to identify the DNA repair genes associated with the

evolution of longevity in plants (Umeda et al., 2021). Plants exhibit a wide range of lifespans, from a few weeks in monocarpic annuals to as long as millennia in long-lived perennials. Plant development fundamentally differs from that of animals. Plant lifespan is characterized by rudimentary body plan, modular growth, and disparity between cell death and death of the organism (Watson and Riha, 2010), allowing high plasticity and indeterminate growth of vegetative meristems that are unique to plants. In perennials, meristematic cells may undergo thousands of divisions. In addition, being sessile organisms, environmental stress may result in increased DNA damage. It is a major interest, therefore, to determine the efficiency of the DNA repair mechanisms in long-lived plant species.

Thanks to the significant progress in the elucidation of the DNA damage repair systems in *A. thaliana* as a model (Hays, 2002; Manova and Gruszka, 2015; Bray and West, 2005; Yoshiyama et al., 2013), all major DNA repair pathways have been reported to be conserved between plants and other organisms. Moreover, a growing number of sequenced genomes in non-model plant species are available. In this study, using more than 60 species of plants, including long-lived trees, perennial herbs, annual herbs, and algae, we performed systematic comparative analyses of the copy number variations of genes that encode proteins involved in DNA repair in diverse plant species with different life forms.

## MATERIALS AND METHODS

### Experimental model and subject details

To collect the information regarding copy number of DNA repair genes in plant species, we used the PLAZA database, the genomic database of diverse plant species. We used Dicots PLAZA 4.0 (Van Bel et al., 2018) and Gymno PLAZA 1.0 (Proost et al., 2009) in order to cover both angiosperms and gymnosperms. These databases also include bryophytes (*Marchantia polymorpha* and *Physcomitrella patens*) and algae (*Chlamydomonas reinhardtii* and *Micromonas commoda*). We categorized each species included in the database into five groups according to life form: alga, annual herb, perennial herb, shrub, and tree based on the information from the databases (the PLANTS Database, Plants of the World Online, Plants For A Future, the University and Jepson Herbaria, the University of Massachusetts Weed Herbarium, the Angiosperm Phylogeny Website, and the Gymnosperm Database) and in the literature (Takasaki et al., 1994; Gotmare et al., 2000; Inan et al., 2004; Zhang et al., 2013b; Tivoli et al., 2006; Merchant et al., 2007; van Baren et al., 2016; Cove, 2005). The species name and number of species of each life form are listed in Appendix Table S5. We eliminated four shrub species (*Actinidia chinensis*, *Gossypium raimondii*, *Manihot esculenta*, and *Vitis vinifera*) from the analyses of life form comparison due to their intermediate life forms, which are tree-like, small sized (< 5 m), and have a relatively short lifespan. Thus, 61 species, including 23 tree species, 15 perennial herb species, 21 annual herb species, and two algae species were used for our analyses (Table 1).

**Methods details**

**Selecting genes associated with DNA repair**

From the Dicots PLAZA 4.0 and Gymno 1.0 PLAZA databases, we selected 171 genes associated with DNA repair systems within *Arabidopsis thaliana* and categorized these genes into 11 functional groups depending on the pathways for DNA repair following Singh et al. (2010) (Appendix Table S1). We used the orthologous groups predicted by the OrthoMCL method from the PLAZA database (Van Bel et al., 2012) as the gene families and grouped 171 DNA repair genes of *A. thaliana* into 121 gene families.

**The index of the copy number of genes for analyses**

To compare the copy number of DNA repair genes between species, we needed to normalize the copy number of genes within each gene family in the focal species by the total number of genes in the species because the species with a large total number of genes would have a large number of DNA repair genes. The PLAZA database provided the copy number ratio rather than the actual copy number. The copy number ratio of the gene family *j* in species *i* was calculated based on four values: the sum of genes included in gene family *j* over all species ($N_j$), the total number of genes included in gene family *j* in species *i* ($L_{ij}$), the sum of the total number of genes over all species ($N_{total}$), and the total number of gene in species *i* ($L_{total,i}$). Using these four values, the copy number ratio is given as follows:

$$\text{The copy number ratio of gene family } j \text{ in species } i = \frac{L_{ij}}{L_{total,i}} \Big/ \frac{N_j}{N_{total}}.$$

The numerator, $\frac{L_{ij}}{L_{total,i}}$, indicates the normalized copy number of the gene family *j* in species *i* that have total number of genes, $L_{total,i}$. The denominator represents the fraction

of the gene family *j* in the total number of genes. We can estimate whether the normalized copy number of the gene family *j* in species *i* is relatively higher compared to the average normalized copy number of the gene family *j* over all species of the dataset using this copy number ratio. By this normalization, the mean of the copy number ratio becomes one.

**Construction of the phylogenetic tree**

To adopt statistical methods to consider the phylogenetic relatedness of target traits, we first drew a phylogenetic tree using species included in database. We constructed a phylogenetic tree using the National Center for Biotechnology Information (NCBI) Taxonomy Browser (Appendix Figure S4). Then, to calculate the branch length of the phylogenetic tree, we collected the DNA sequences of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and maturaseK (*matK*) from the NCBI. Because the sequence data of *rbcL* of *Citrus clementina* was not found, the sequence of *rbcL* of *C. sinensis*, a closely related species of *C. clementina*, was used as an alternative. Because no sequence data for two algae species, *Chlamydomonas reinhardtii* and *Micromonas commoda*, was available, we eliminated these two algae species from the analysis. Thus, we used 59 species for the analyses that considered the phylogenetic relationships (Appendix Table S6). We aligned the sequences using the ClustalW algorithm in the program Molecular Evolutionary Genetics Analysis (MEGA) X (v. 10.1.5); Kumar et al., 2018). After alignment, we calculated the branch lengths of the phylogenetic tree using RAxML (v. 8.0.0; Stamatakis, 2014).

## Quantification and statistical analysis

### The similarity of the copy number ratio between species

To assess the similarity of the copy number ratio between species, we performed hierarchical clustering based on the Euclidean distance of the copy number ratio of each species using the Ward's method. To test the enrichment or dilution of each life form in each of the significantly different clusters, Fisher's exact tests (two-sided) were performed. After the clustering, we tested whether the species in the cluster had a higher or lower copy number ratio than the mean of all species. The mean copy number ratio of 121 gene families within each species was calculated. Then, we tested whether the average of the mean copy number ratio of 121 gene families within species included in the cluster was significantly higher or lower than one (that is, mean copy number ratio of all species) by *t*-test. After the *t*-tests, to control for false discovery rate, we used the method of Storey's Q-value (Storey, 2002), and the Q-value of each test was estimated using the q-value package (ver. 2.16.0); Storey et al. (2015) in R.

### PGLS to investigate the relationship between the copy number ratio and life forms

Next, we explored the relationship between the copy number ratio and life forms in each gene family using phylogenetic generalized least squares (PGLS) regression (Grafen, 1989) in the phylolm package (ver. 2.6); Tung Ho and Ané (2014) in R. For this analysis, we estimated Pagel's lambda (Pagel, 1997) to evaluate the influence of phylogenetic relationships on the data and tested whether the regression coefficients differed from zero. After the PGLS analyses, we controlled the false discovery rate and estimated the Q-value using the method explained above.

**The evolutionary history of the PARP gene family**

Because our analyses revealed the potential role of *PARP* (Poly(ADP-ribose) polymerase) genes in longevity in tree species, we investigated the evolutionary history of the *PARP* gene family in plant species. First, to assess and compare the domain structures of *PARP* genes, we constructed a phylogenetic tree of 189 *PARP* genes from 53 dicot species (Appendix Table S7) using the tree explore tool in Dicots PLAZA 4.0 (note that this function in PLAZA is available only for dicot species). Based on the method provided by PLAZA, genes with low sequence similarity were removed from the phylogenetic trees as partial or outlier genes.

Second, we constructed a phylogenetic tree of *PARP* genes from diverse plant species, including angiosperms, gymnosperms, lycophytes, and bryophytes. There were 332 *PARP* genes in the original data set (Appendix Table S8). Of 332 genes, 131 were selected by increasing gap-free sites using MaxAlign with a heuristic algorithm (Gouveia-Oliveira et al., 2007) and aligned using the MAFFT online service (Katoh et al., 2018). Then, the phylogenetic tree was constructed using the neighbor-joining method with the Jones–Taylor–Thornton (JTT) substitution model (Jones et al., 1992) and bootstrapping over 1000 trees.

We categorized each *PARP* gene into three groups: *PARP1*, *PARP2*, and *PARP3*, based on different methods in the Dicot and Gymno PLAZA databases. *PARP* genes included in the Dicots PLAZA 4.0 database were categorized following the annotation given in the PLAZA database. We removed the genes categorized as unknown, or genes without detailed annotation, in Dicots PLAZA 4.0.

The *PARP* genes included in the Gymno PLAZA 1.0 database were categorized into three different paralogs based on the clustering information in the gymnosperm phylogenetic tree because most of the *PARP* genes included in the Gymno PLAZA 1.0 database showed no annotation.

We constructed the phylogenetic tree for gymnosperms using the same method explained above by extracting 24 gymnosperm *PARP* genes (Appendix Table S9 and Appendix Figure S5). 88 *PARP* genes in gymnosperms were removed for the phylogenetic tree construction using MaxAlign due to the existence of long gaps in their sequences. These genes were annotated using the Basic Local Alignment Search Tool (BLAST+) (Camacho et al., 2009) against the database, which included 24 sequences of gymnosperm species *PARP* genes used to construct the phylogenetic tree of *PARP* in gymnosperm species. After the annotations, each gene was categorized according to the "best hit" in BLAST. For each paralog of *PARP* gene family, we conducted PGLS regressions and compared the copy number ratios among life forms using the method explained above.

**The relationship between copy number ratio of *PARP* and growth rate**

Our analyses revealed that *PARP* gene family and especially *PARP1* and *PARP2* genes showed significant higher copy number ratios in tree species that generally live longer than herb species. To assess the possibility that the increased copy number of *PARP* is associated with longevity, it is useful to investigate the relationship between copy number ratio of *PARP* and plant lifespan. Because reliable estimation of plant lifespan is very difficult and published maximum tree lifespans are not always supported by scientific evidence (Piovesan and Biondi, 2020), we used growth rate that is inversely

related to lifespan of many plant species (Johnson and Abrams, 2009; Black et al., 2008). It has been discussed that long-lived, late successional species typically grow more slowly, invest more resources for defensive compounds and structural support, and maintain lower rates of photosynthesis and respiration than shorter-lived, early successional species (Loehle, 1988).

We successfully collected the data regarding the individual ages and heights in 11 tree species including angiosperms and gymnosperms from the literature (Köstler, 1956; Burns and Honkala, 1990a; 1990b; Liebhard et al., 2003; Bravo-Oviedo et al., 2004) (Appendix Table S4). Then, we calculated the average growth rate (the rate of height increment per year) for each species. We collected the data sampled in similar regions (e.g., North America and Switzerland) to align the environmental conditions for tree growth. Because inverse relationship between growth rate and longevity has been argued mainly in tree species, and height growth rate is difficult to obtain in herbs, we applied this analysis only for tree species.

Next, we constructed the phylogenetic tree of 11 tree species for the analysis considering the phylogenetic relationships. We constructed the phylogenetic tree based on amino acid sequences of rbcL and matK using the neighbor-joining method with the JTT substitution model and bootstrapping over 1000 trees by MEGA X (Appendix Figure S6).

Finally, to investigate the relationship between the copy number ratio of each type of *PARP* and the growth rate in 11 tree species, we performed PGLS regression using the method explained above. After the regression analyses, we controlled the false discovery rate and estimated the Q-value using the method explained above.

To perform all statistical analyses, we used R ver. 3.6.3 (the R project, http://www.r-project.org/).

## RESULTS

### Interspecies comparison of copy number ratio of 121 DNA repair gene families

To compare the copy number variations of DNA repair genes between diverse species, we used the PLAZA database (Dicots PLAZA 4.0; Van Bel et al., 2018 and Gymno PLAZA 1.0; Proost et al., 2009), the genomic database of diverse plant species. We used 61 plant species, including 23 tree species, 15 perennial herb species, 21 annual herb species, and two algae species for our analyses (Table 1), thereby covering both angiosperms and gymnosperms. Because the species with large genome sizes would have a large number of DNA repair genes, the PLAZA database provided the normalized index, namely the copy number ratio, by dividing the actual copy number of genes within each gene family in the focal species by the total number of genes in the species (see Methods section). We selected 171 genes involved in DNA repair within *A. thaliana* (Appendix Table S1). We used the orthologous groups predicted by the OrthoMCL method from the PLAZA database (Van Bel et al., 2012) as the gene family and grouped 171 DNA repair genes of *A. thaliana* into 121 gene families.

Hierarchical clustering based on the similarity of the copy number ratio between species showed that 61 species were divided into four clusters (Figure 1A). Cluster1 consisted of three species, which were two algae species and one perennial herb (lycophyte) species, revealing significant enrichment of algae species (Fisher's exact test; Q-value = 0.0262) (Appendix Table S2). The average of the mean copy number ratio over 121 DNA repair gene families was higher, but not significantly

different from the mean of all species and other clusters (*t*-test; Q-value = 0.145) (Figure 1B). Cluster2 consisted of only five species, all of which were trees (one angiosperm and four gymnosperms), revealing significant enrichment of tree species (Fisher's exact test; Q-value = 0.0452) (Appendix Table S2). The average of the mean copy number ratio in Cluster2 was significantly larger than the mean of all species and other clusters (*t*-test; *t*-value = 12.55, P-value = $2.32\times10^{-4}$, Q-value = $4.64\times10^{-4}$) (Figure 1B). In Cluster3, which consisted of 17 species, the average of the mean copy number ratio was significantly lower than the mean of all species (t-test; *t*-value = –3.83, P-value = 0.00147, Q- value = 0.00197) (Figure 1B). Cluster3 included eight tree, five perennial herb, and four annual herb species, revealing no significant enrichment or dilution of a certain type of life form (Appendix Table S2). Cluster4 included the largest number of species, in which the average of the mean copy number ratio was significantly larger than the mean of all species (*t*-test; *t*-value = 5.80, P-value = $1.42\times10^{-6}$, Q-value = $5.69\times10^{-6}$) (Figure 1B). Among the 36 species in Cluster4, ten species were trees, nine were perennial herbs, and 17 species were annual herbs. There was no significant enrichment or dilution of a certain type of life form (Appendix Table S2).

An alga, *Micromonas commoda*, is a unique species with low similarity of copy number ratio compared to the other species studied here (Figure 1A). In *M. commoda*, the copy number ratio was greater than the mean of all species in 105 gene families, whereas it was zero in 16 gene families (Figure 1A). Such a clear contrast of high and low copy number ratios among gene families was also found in another alga species, *Chlamydomonas reinhardtii*, and gymnosperm tree species, such as *Ginkgo biloba* and *Picea sitchensis*, but the pattern of the gene families with a high copy

number ratio or a zero copy number ratio varied among species. This result suggests that each gene family has a species-specific history of gene loss and gene duplication.

The mean of actual copy number over species in each gene family was smaller than five and variance among species was low in most of the gene families (Figure 1A). However, in several gene families, the mean and variance of actual copy number was extremely large. For example, in the gene family involved in protein kinase production, including checkpoint kinase 2 (*CHEK2*), which participates in the DNA damage response in many cell types (Cybulski et al., 2004), and the cullin family, including cullin 4 (*CLU4*), which is involved in repair of UV-induced DNA lesions (Molinier et al., 2008), the means of the actual copy number were 47.07 and 15.61, and the variances of the actual copy number were 499.04 and 513.52, respectively (Figures 1A and Appendix Figure S1). The phylogenetic signals in these gene families that had large mean copy numbers and large variance among species were weak (e.g., the estimated Pagel's lambda in the protein kinase gene family was $7.55\times10^{-5}$; and 0.077 in the cullin family). In addition, there was no significant relationship between the copy number and the life forms. Conversely, these gene families showed a positive correlation between the copy number and the total number of genes in a species (e.g., Spearman's rank correlation coefficient in the protein kinase gene family was 0.77; and 0.61 in the cullin family). This suggests that the family sizes of protein kinase and cullin increased with the genome size expansion.

**Extracting the DNA repair gene family that showed a high copy number ratio in tree species**

Next, we investigated whether copy number ratios are significantly different among tree, perennial, and annual herb species for each gene family using phylogenetic generalized least squares (PGLS). The phylogenetic signals in the copy number ratio varied depending on the gene family (Table S3). The estimated values of Pagel's lambdas were smaller than 0.1 in 60 gene families (e.g., poly(ADP-ribose) polymerase [*PARP*], breast cancer 2 [*BRCA2*], and DNA damage-binding protein [*DDB*]), and were greater than 0.1 in 61 gene families (e.g., DNA glycosylase superfamily protein [Tag], replication protein A2 [RPA2] and structural maintenance of chromosomes 6 [*SMC6*]) (Appendix Table S3).

Among the 121 gene families, only one showed a significantly higher copy number ratio in tree species than in perennial and annual herb species, which was poly(ADP-ribose) polymerase (*PARP*s) (Figure 2A). Another gene family (*Tag*) showed a significantly higher copy number ratio in tree species than in perennial herb species, but the difference between tree and annual herb species was not significant in this gene family (Appendix Figure S2). The three species with the highest copy number ratio of *PARP*s were *Pseudotsuga menziesii* (Douglas-fir), *Pinus sylvestris* (Scots pine), and *Malus domestica* (apple) (Figure 2B). Douglas-fir and Scots pine are known as long-lived conifers and can live for over 1000 years (Franklin and Dyrness, 1973). Apple trees live between 60 and 100 years (Pereira-Lorenzo et al., 2009). Although the longevity of the apple tree is not as long as that of conifers, it is significantly longer than that of herb species.

PARPs are key enzymes associated with poly(ADP-ribosyl)ation. Poly(ADP-ribosyl)ation is a covalent posttranslational modification process of proteins via the synthesis and transfer of poly ADP-ribose from NAD+ to target proteins (Rissel and

Peiter, 2019). The ADP-ribose polymer formed by the sequential attachment of ADP-ribosyl moieties attracts enzymes for DNA repair, particularly those associated with BER and other types of ssDNA repair. *PARP*s are found in all eukaryotic supergroups (Citarelli et al., 2010) and *A. thaliana* encodes three canonical PARP proteins (AtPARP1, AtPARP2, and AtPARP3).

Poly(ADP-ribosyl)ation is reversible and the covalently attached poly(ADP-ribose) from acceptor proteins are removed by poly(ADP-ribose) glycohydrolase (PARG) enzymes (Briggs and Bent, 2011; Vainonen et al., 2016). PARP and PARG proteins interact with each other, and the cellular pools of ADP- ribose are regulated. Because plant PARGs are also involved in DNA repair and biotic/abiotic stress responses (Li et al., 2011; Zhang et al., 2015; Song et al., 2015), we compared the copy number ratio of *PARG* genes among lifeforms. We found there was no significant difference in the copy number ratio of *PARG* gene family between tree species and perennial herb species (Q-value was 0.732) and between tree species and annual herb species (Q-value was 0.286), although the copy number ratio in tree species was lower than those in herb species. This result suggests that increased copy number of *PARG*s is not essential for DNA repair and the longevity in plants.

**The *PARP* gene family was divided into three functional groups**

189 *PARP* genes in dicot species were divided into four distinct clades based on sequences and protein domain structures using the tree explore tool in Dicots PLAZA 4.0 (Figure 3). One clade consisted of 59 genes from 52 species and was named as the *PARP1* clade because almost all members were characterized by a highly conserved domain structure of *Arabidopsis* PARP1 (Figure 3A). *Arabidopsis* PARP1 possesses an

N-terminal DNA interaction domain (Zinc-finger), a C-terminal catalytic domain (PARP catalytic; Rissel and Peiter, 2019), a PARP regulatory domain (PARP regulatory), and a WGR domain, named after its repeating amino acid motif (W-G-R), located in the central region. The *PARP2* clade consisted of 66 genes in 48 species, including *Arabidopsis PARP2* (Figure 3B). Almost all members of the *PARP2* clade lack the zinc-finger domains but possess SAF-A/B, acinus, and PIAS (SAP) domains in the N-terminus, consistent with the previous characterization of *Arabidopsis* PARP2 (Lamb et al., 2012). The SAP domain has been shown to bind to nucleic acids (Okubo et al., 2004), suggesting the ability of DNA binding for PARP2 protein. Another clade, named as the *PARP3* clade, consisted of 56 genes in 48 species, including *Arabidopsis PARP3* (Figure 3C). The domain structure of the *PARP3* clade members resembles those of the *PARP1* clade, but members of the *PARP3* clade lack the zinc-finger domains, consistent with the finding of previous study based on *A. thaliana* (Vainonen et al., 2016).

Members in a minor clade (named ‘‘Other’’), consisted of eight genes and had only zinc-finger domains, implying no catalytic or regulatory functions (Figure 3D). BLAST search against human genome showed that the sequences of these genes are the most similar to human *PARP1* gene rather than other human *PARP* genes. In addition, the sequences of these genes were more similar to plant *PARP2* gene rather than radical-induced cell death 1 (*RCD1*) gene and Similar to RCD one (*SRO*s) genes, which encode proteins containing PARP-like domains (Jaspers et al., 2010).

The phylogenetic tree constructed from the plant species, including angiosperms, gymnosperms, lycophytes, and bryophytes, also showed that plant *PARP* genes were divided into three distinct clades of *PARP1*, *PARP2*, and *PARP3* (Appendix

Figure S3), suggesting that three paralogs of the *PARP* genes were present in the common ancestor of angiosperms, gymnosperms, lycophytes, and bryophytes.

**Tree species have higher copy number ratios in *PARP1* and *PARP2* but not in *PARP3***

The copy number ratios of the members of *PARP1* and *PARP2* clades were significantly higher in tree species than in annual and perennial herb species (Figures 4A and 4B and Table 2), but there was no significant difference between life forms for *PARP3* (Figure 4C and Table 2). The tree species that showed the highest copy number ratio of each *PARP* gene were different: *Pinus sylvestris*, *Ziziphus jujuba*, and *Pseudotsuga menziesii* showed the highest copy number ratios of *PARP1*, *PARP2*, and *PARP3*, respectively (Figure 4D).

The actual copy number of *PARP* genes was also large in tree species, especially in gymnosperms (*P. sylvestris*, *Pinus taeda*, *Pinus pinaster*, and *P. menziesii*: Figure 4E). *P. taeda* had eight *PARP1* genes, the largest number of *PARP1* genes among all species. *P. menziesii* had 44 *PARP3* genes, the largest number of *PARP3* genes among all species. All tree species had at least one *PARP1* gene, but some gymnosperms had lost the *PARP2* and/or *PARP3* genes (Figure 4E), suggesting that *PARP1* is the most essential gene for long-lived trees.

**An inverse relationship between copy number ratios in *PARP*s and growth rate in tree species**

Next, we tested whether there is a significant association between copy number ratio of *PARP*s and longevity. Because reliable estimation of plant lifespan is difficult and

maximum tree lifespans published in prestigious scientific journals are not always supported by scientific evidence (Piovesan and Biondi, 2020), we used growth rate (the rate of height increment) instead of lifespan. In the field of forest ecology, there is a longstanding argument that slow-growing trees live longer than fast-growing trees (Johnson and Abrams, 2009; Black et al., 2008). Because the data for growth rate can be more easily available than those for longevity, we collected the growth data in 11 tree species including angiosperms and gymnosperms from previous studies (Ko¨stler, 1956; Burns and Honkala, 1990a; 1990b; Liebhard et al., 2003; Bravo-Oviedo et al., 2004) (Appendix Table S4) and investigated the relationship between the growth rate and the copy number ratio of *PARP*s using phylogenetic generalized least squares (PGLS) regression analyses. Because inverse relationship between growth rate and longevity has been argued mainly in tree species, and height growth rate is difficult to obtain in herbs, we applied this analysis only for tree species.

There was significantly negative correlation between log growth rate (m/year) and the copy number ratio in *PARP* gene family (Figure 5) (Table 3). Among three *PARP* family members, the significantly negative correlation between log growth rate and the copy number ratio was shown only in *PARP1* (Figure 5) (Table 3). This result strongly suggests the important role of PAPR1 for slow growth and longevity in tree species.

## DISCUSSION

To examine the role of DNA repair in plant longevity, we systematically compared the copy number variations of 121 DNA repair gene families in 61 plant species, including trees, annual/perennial herbs, and algae. Among the diverse DNA repair gene families

studied here, the *PARP* gene family was identified as the only one that revealed significant expansion in tree species relative to annual/perennial herb species. The long-lived conifers, Douglas-fir and Scots pine, as well as fruit tree (apple tree) were found to be the species with highest copy number ratios of *PARP*s. These results suggest that selection probably promotes convergent evolution of increased copy numbers of *PARP*s in tree species.

As key enzymes associated with poly(ADP-ribosyl)ation, PARPs have been extensively studied in animals. The *PARP* gene family is considerably larger in vertebrates than in plants. In humans, there are 17 family members that share the PARP catalytic domain of PARP1 (Amé et al., 2004; Hottiger et al., 2010). Our analyses showed that 59 plant species, including angiosperms, gymnosperms, lycophytes, and bryophytes have only two or three *PARP* family members (Figures 3 and 4E). PARP proteins in *A. thaliana* (AtPARP1, 2, and 3), *Zea mays* (maize) and *Glycine max* (soybean) have confirmed or predicted poly ADP-ribosylation activity (Jaspers et al., 2010; Babiychuk et al., 1998; Amor et al., 1998), and AtPARP1 and AtPARP3 are structurally the most similar to human PARP1, whereas AtPARP2 is similar to human PARP2, indicating the functional similarities between *Arabidopsis* and human PARPs (Rissel and Peiter, 2019).

Among three *PARP* family members in plants, only the copy number ratios of the two members, *PARP1* and *PARP2*, were significantly higher in tree species than those in annual and perennial herb species (Figure 4). In *A. thaliana*, AtPARP1 and AtPARP2 play the predominant role in poly(ADP-ribose) polymerase activity and DNA damage response (Song et al., 2015; Gu et al., 2019). In contrast to *AtPARP1* and *AtPARP2*, the expression of *AtPARP3* is restricted to seed tissues (Rissel et al., 2014).

Moreover, a recent study reported that AtPARP3 does not have poly(ADP-ribose) polymerase activity (Gu et al., 2019). Together with these previous reports, our results suggest that increased copy numbers of *PARP*s that are capable of adding ADP-ribose units onto protein substrates are likely to be evolutionary favored in long-lived tree.

The best-studied PARPs, including the founding member PARP1, catalyze the formation of long, branched chains of ADP-ribose, known as poly (ADP-ribose) (PAR) (Hassa and Hottiger, 2008; Gibson and Kraus, 2012). These PAR-forming enzymes perform functions such as nucleation of DNA-damage foci (PARP1 and 2) and proper chromosome segregation during mitosis (PARP5a in human) (Schreiber et al., 2006; Hassa and Hottiger, 2008). Although historically PARP1 in animals has been studied with the focus on DNA damage detection and repair, more recently it has been understood that in the absence of DNA damage, PARP1 also plays an important role in regulating chromatin structure and gene expression by biding near the promoters of transcriptionally active genes (Krishnakumar and Kraus, 2010). Cell survival after genotoxic stress is determined by a counterbalance of pro- and anti-death factors. Sirtuins (SIRTs) are deacetylases that promote cell survival, whereas poly(ADP-ribose) polymerases (PARPs) can act both as survival and death inducing factor. The two protein families are strictly dependent on the oxidized form of nicotinamide adenine dinucleotide (NAD+) for their activities. Previous studies have reported that increased activity of PARP1, but not overexpression, is associated with longevity of mammalian species (Grube and Bürkle, 1992). Furthermore, increased amounts sirtuins are associated with improved health and longevity in mammals (Mouchiroud et al., 2013). Although less is known about the functions of plant PARPs in contrast to their mammalian counterparts, AtPARP1 and AtPARP2 have been shown to be associated

with DNA repair (Doucet-Chabeaud et al., 2001; De Block et al., 2005) and transcriptional regulation (Babiychuk et al., 2001; Storozhenko et al., 2001; Vanderauwera et al., 2007). Our findings that long-lived trees have higher copy number ratio of *PARP*s than herbs will lead to the intriguing hypothesis that PARPs play an important role on aging and longevity both in plants and animals.

The pharmacological and genetic inhibition of PARP in *A. thaliana* results in an increased stress tolerance and increased growth by preventing cell death (De Block et al., 2005) but it also leads to reduced defense because of the reduced accumulation of protective molecules, especially anthocyanin and ascorbate (Schulz et al., 2012). The antagonistic relationship between increased growth and decreased defense by inhibition of PARP provides an important insight into the long-standing ecological argument that slow-growing trees live longer than fast-growing trees (Johnson and Abrams, 2009; Black et al., 2008). Long-lived, late successional species typically grow more slowly, invest more resources for defensive compounds and structural support, and maintain lower rates of photosynthesis and respiration than shorter-lived, early successional species (Loehle, 1988). Although the underlying molecular mechanism for long-lived and short-lived tree species remained completely unknown, our finding provides the new testable hypothesis that increasing copy number of *PARP*s enhance allocation to defensive compounds that leads to slow growth and great longevity. Indeed, the plot of growth rate against the copy number ratio of *PARP1*s showed a significant negative correlation (Figure 5).

In mammals, there is a clear positive correlation between activity of PARPs and longevity (Grube and Bürkle, 1992), although the copy number of *PARP* genes are not so different among species with different life spans (MacRae et al., 2015a). Given

these previous reports in mammals, we speculate that the enhanced activity of PARPs could contribute to the longevity in animals, while an increased copy number of *PARP*s is more likely to occur in long-lived plants. The difference between animals and plants may be originated from the different history of genome evolution. In plants, whole genome duplication and polyploidization events occurred more frequently than those in animals (Murat et al., 2012). Because frequent duplication and polyploidization would lead to dynamic and faster genome evolution, the copy number of *PARP*s could change more flexible in plants than in mammalian genomes that are conserved and stable.

Another important function of PARPs is to regulate viral infectivity and pathogenesis (Kuny and Sullivan, 2016). In humans, PARP13 has been reported to reveal broad antiviral activity through direct biding of viral RNA by PARP13, followed by recruitment of the exosome and specific degradation of viral RNA (Gao et al., 2002; Müller et al., 2007; Bick et al., 2003; Mao et al., 2013). Daugherty et al. (2014) demonstrated that nearly one-third of primate *PARP* genes, including *PARP13*, are evolving under strong recurrent positive selection, implicating the essential role of PARPs in antiviral defense in mammalian genomes. The role of PARPs in antipathogen defense can also be identified in plants. In *A. thaliana,* AtPARP2 has been demonstrated to regulate the response to pathogen infection and repair of pathogen-induced DNA damage (Song et al., 2015). Because long-lived trees are exposed to the continuous risk of pathogen-induced DNA damage, protection of the plant host genome against pathogen invasion is essential (Song and Bent, 2014). A recent comparative genomics study showed the clear expansion of plant resistance genes (R-genes) and orthologs related to plant immunity in trees relative to herbs (Tobias and Guest, 2014; Plomion et

al., 2018). An increased copy number of *PARP*s could provide another mechanism of antipathogen defense that is necessary for the success of long-lived trees.

In addition to the *PARP* gene family, our hierarchical clustering analysis results (Figure 1) showed that the increased copy number ratio of various DNA repair gene families may contribute to the longevity of some tree species, including *Citrus clementina*, *Cycas micholitzii*, *Ginkgo biloba*, *Gnetum montanum*, and *Taxus baccata*. DNA damages varied from basal lesions to DNA double-strand breaks (DSBs) due to various genotoxic stresses, and such DNA damages can be repaired by various DNA repair. Previous studies showed the positive correlation between the activities of DNA repair in multiple pathway and longevity in animal species. Humans and naked mole-rats, which have long lifespans, have significantly higher expression levels of DNA repair genes including genes involved in DNA damage sensing, mismatch repair (MMR), non-homologous end-joining (NHEJ) repair and base excision repair (BER) than mouse (MacRae et al., 2015b). DNA repair genes involved in BER and repair of DNA DSBs are more highly expressed in long-lived bat species than in short-lived bat species (Huang et al., 2020). Thus, the coevolution of copy number variations of DNA repair genes in multiple pathways may provide a strategy for efficient DNA repair, contributing to the success of long-lived organisms. Comparison of expression profiles of DNA repair genes including *PARP*s among plant species with different lifespans will be extremely interesting in future studies.

Among 121 DNA repair gene families studied, only one gene family, *PARP* gene family, was identified as the gene family that revealed significant expansion in tree species relative to annual and perennial herb species. Although some gene families also had an important role in DNA repair, significant expansion in tree species relative to

herb species was not found in most gene families. This is because the number of species in the dataset was not large enough and species were limited. In spite of the limitation of data, *PARP* gene family was found to have significantly higher copy number ratio in tree species than annual and perennial herb species. This suggests that *PARP* gene family is a strong candidate gene family associated with tree longevity.

Overall, systematic comparative analyses of the copy number variations in DNA repair genes in diverse species demonstrates that *PARP*s, especially *PARP1* and *PARP2*, are strong candidate genes associated with tree longevity. PARPs have pivotal roles in the response to and repair of DNA damage, including basal and bulky lesions and single- and double-strand breaks due to endogenous and exogenous stresses. The result of our study can be a foundation for research to elucidate the relationships of DNA repair and the evolution of species longevity in plants. As genome sequences of more diverse plant species become available, systematic comparative genome analyses will provide important clues to reveal the relationships of DNA repair and the evolution of longevity in diverse organisms.

**Limitations of the study**

We collected the information regarding the copy number of DNA repair genes in plant species from the PLAZA database, the genomic database of diverse plant species. We used Dicots PLAZA 4.0 and Gymno PLAZA 1.0 so that we could cover both angiosperms and gymnosperms. The predicted copy number of DNA repair genes are largely derived from newly sequenced plant genomes using homologous sequences. The estimates may therefore not accurately represent true biological gene numbers and should be interpreted with caution. We also acknowledge that 61 species used for our

analyses may not be sufficient. Thanks to the advances in DNA sequencing technology, genomes from increasingly large number of species will be available in the near future. Applying our analyses to the larger set of data will uncover new DNA gene families that could be involved in tree longevity.

## ACKNOWLEDGMENTS

We would like to thank Dr. Mizuki Ohno and Dr. Eriko Sasaki for helpful discussions and comments on this study. This study was funded by JSPS KAKENHI (JP26251042; JP17H06478) to A.S.

## REFERENCES

Amé, J. C., Spenlehauer, C., & de Murcia, G. (2004). The PARP superfamily. *Bioessays, 26*(8), 882-893.

Amor, Y., Babiychuk, E., Inzé, D., & Levine, A. (1998) The involvement of poly(ADP-ribose) polymerase in the oxidative stress responses in plants, *FEBS Letters*, 440(1–2), pp. 1–7. doi: 10.1016/S0014-5793(98)01408-2.

Aravind, L., Walker, D. R. and Koonin, E. V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems, *Nucleic Acids Research*, 27(5), pp. 1223–1242. doi: 10.1093/nar/27.5.1223.

Babiychuk, E., Cottrill, P. B., Storozhenko, S., Fuangthong, M., Chen, Y., O'Farrell, M. K., Van Montagu, M., Inzé, D., & Kushnir, S. (1998) Higher plants possess two

structurally different poly(ADP-ribose) polymerases, *Plant Journal*, 15(5), pp. 635–645. doi: 10.1046/j.1365-313X.1998.00240.x.

Babiychuk, E., Van Montagu, M., & Kushnir, S. (2001). N-terminal domains of plant poly(ADP-ribose) polymerases define their association with mitotic chromosomes. *Plant Journal*, 28(3), 245–255. https://doi.org/10.1046/j.1365-313X.2001.01143.x

Baer, C. F., Miyamoto, M. M. and Denver, D. R. (2007) Mutation rate variation in multicellular eukaryotes: Causes and consequences, *Nature Reviews Genetics*. doi: 10.1038/nrg2158.

Bick, M. J., Carroll, J.-W. N., Gao, G., Goff, S. P., Rice, C. M., & MacDonald, M. R. (2003) Expression of the Zinc-Finger Antiviral Protein Inhibits Alphavirus Replication, *Journal of Virology*, 77(21), pp. 11555–11562. doi: 10.1128/jvi.77.21.11555-11562.2003.

Black, B. A., Colbert, J. J., & Pederson, N. (2008). Relationships between radial growth rates and lifespan within North American tree species. *Écoscience*, 15(3), 349–357. https://doi.org/10.2980/15-3-3149

Bravo-Oviedo, A., Río, M. Del, & Montero, G. (2004). Site index curves and growth model for Mediterranean maritime pine (Pinus pinaster Ait.) in Spain. *Forest Ecology and Management*, 201(2–3), 187–197. https://doi.org/10.1016/j.foreco.2004.06.031

Bray, C. M. and West, C. E. (2005) DNA repair mechanisms in plants: Crucial sensors and effectors for the maintenance of genome integrity, *New Phytologist*, 168(3), pp. 511–528. doi: 10.1111/j.1469-8137.2005.01548.x.

Briggs, A. G. and Bent, A. F. (2011) Poly(ADP-ribosyl)ation in plants, *Trends in Plant Science*. Elsevier Ltd, 16(7), pp. 372–380. doi: 10.1016/j.tplants.2011.03.008.

Burns, R.M. & Honkala, B.H. (1990a) Silvics of North America: Volume 1. Conifers. Agriculture Handbook 654, U.S. Department of Agriculture, Forest Service, Washington, DC.

Burns, R.M. & Honkala, B.H. (1990b) Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654, U.S. Department of Agriculture, Forest Service, Washington, DC.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009) BLAST+: Architecture and applications, *BMC Bioinformatics*. doi: 10.1186/1471-2105-10-421.

Ciccia, A. and Elledge, S. J. (2010) The DNA Damage Response: Making It Safe to Play with Knives, *Molecular Cell*. doi: 10.1016/j.molcel.2010.09.019.

Citarelli, M., Teotia, S. and Lamb, R. S. (2010) Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes, *BMC Evolutionary Biology*, 10(1), pp. 1–26. doi: 10.1186/1471-2148-10-308.

Cove, D. (2005) The Moss Physcomitrella patens, *Annual Review of Genetics*, 39(1), pp. 339–358. doi: 10.1146/annurev.genet.39.073003.110214.

Cybulski, C. *et al.* (2004) CHEK2 is a multiorgan cancer susceptibility gene, *American Journal of Human Genetics*, 75(6), pp. 1131–1135. doi: 10.1086/426403.

De Block, M., Verduyn, C., De Brouwer, D., & Cornelissen, M. (2005). Poly(ADP-ribose) polymerase in plants affects energy homeotasis, cell death and stress tolerance. *Plant Journal*, 41(1), 95–106. https://doi.org/10.1111/j.1365-313X.2004.02277.x

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C. E., Maere, S., & Van De Peer, Y. (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants, *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), pp. 2898–2903. doi: 10.1073/pnas.1300127110.

Daugherty, M. D., Young, J. M., Kerns, J. A., & Malik, H. S. (2014) Rapid Evolution of PARP Genes Suggests a Broad Role for ADP-Ribosylation in Host-Virus Conflicts, *PLoS Genetics*, 10(5). doi: 10.1371/journal.pgen.1004403.

Doucet-Chabeaud, G., Godon, C., Brutesco, C., De Murcia, G., & Kazmaier, M. (2001). Ionising radiation induces the expression of PARP-1 and PARP-2 genes in Arabidopsis. *Molecular Genetics and Genomics*, 265(6), 954–963. https://doi.org/10.1007/s004380100506

Eisen, J. A. and Hanawalt, P. C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes, *Mutation Research - DNA Repair*. doi: 10.1016/S0921-8777(99)00050-6.

Franklin, J. F. and Dyrness, C. T. (1973) Natural vegetation of Oregon and Washington. USDA For. Serv, USDA Forest Service - General Technical Report GTR-PNW-8.

Freitas, A. A. and De Magalhães, J. P. (2011) A review and appraisal of the DNA damage theory of ageing, *Mutation Research - Reviews in Mutation Research*, 728(1–2), pp. 12–22. doi: 10.1016/j.mrrev.2011.05.001.

Gao, G., Guo, X. and Goff, S. P. (2002) Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein, *Science*. doi: 10.1126/science.1074276.

Gibson, B. A., & Kraus, W. L. (2012). New insights into the molecular and cellular functions of poly (ADP-ribose) and PARPs. *Nature reviews Molecular cell biology*, *13*(7), 411-424.

Gotmare, V. P., P. Singh, and B. N. Tule. (2000) Wild and Cultivated Species of Cotton. Technical Bulletin No. **5**. CICR, Nagpur, India

Gouveia-Oliveira, R., Sackett, P. W. and Pedersen, A. G. (2007) MaxAlign: Maximizing usable data in an alignment, *BMC Bioinformatics*, 8, pp. 1–8. doi: 10.1186/1471-2105-8-312.

Grafen, A. (1989) The phylogenetic regression., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: 10.1098/rstb.1989.0106.

Grube, K. and Bürkle, A. (1992) Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span, *Proceedings of the National Academy of Sciences of the United States of America*, 89(24), pp. 11759–11763. doi: 10.1073/pnas.89.24.11759.

Gu, Z., Pan, W., Chen, W., Lian, Q., Wu, Q., Lv, Z., Cheng, X., & Ge, X. (2019) New perspectives on the plant PARP family: Arabidopsis PARP3 is inactive, and PARP1 exhibits predominant poly (ADP-ribose) polymerase activity in response to DNA damage, *BMC Plant Biology*, 19(1), pp. 1–18. doi: 10.1186/s12870-019-1958-9.

Hassa, P. O., & Hottiger, M. O. (2008). The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Front Biosci*, *13*(13), 3046-3082.

Hays, J. B. (2002) Arabidopsis thaliana, a versatile model system for study of eukaryotic genome-maintenance functions, *DNA Repair*, 1(8), pp. 579–600. doi: 10.1016/S1568-7864(02)00093-9.

Hottiger, M. O., Hassa, P. O., Lüscher, B., Schüler, H., & Koch-Nolte, F. (2010). Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends in biochemical sciences*, *35*(4), 208-219.

Huang, Z., Whelan, C. V., Dechmann, D., & Teeling, E. C. (2020) Genetic variation between long-lived versus short-lived bats illuminates the molecular signatures of longevity, *Aging*, 12(16), pp. 15962–15977. doi: 10.18632/aging.103725.

Jaspers, P., Overmyer, K., Wrzaczek, M., Vainonen, J. P., Blomster, T., Salojärvi, J., ... & Kangasjärvi, J. (2010). The RST and PARP-like domain containing SRO protein family: analysis of protein structure, function and conservation in land plants. *BMC genomics*, *11*(1), 170.

Johnson, S. E., & Abrams, M. D. (2009). Age class, longevity and growth rate relationships: Protracted growth increases in old trees in the eastern United States. *Tree Physiology*, 29(11), 1317–1328. https://doi.org/10.1093/treephys/tpp068

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences, *Bioinformatics*. doi: 10.1093/bioinformatics/8.3.275.

Katoh, K., Rozewicki, J. and Yamada, K. D. (2018) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization, *Briefings in Bioinformatics*, 20(4), pp. 1160–1166. doi: 10.1093/bib/bbx108.

Keane, M. *et al.* (2015) Insights into the evolution of longevity from the bowhead whale genome, *Cell Reports*, 10(1), pp. 112–122. doi: 10.1016/j.celrep.2014.12.008.

Köstler, Josef. (1956) Silviculture. Edinburgh: Oliver and Boyd. 416 p.

Krishnakumar, R. and Kraus, W. L. (2010) The PARP Side of the Nucleus: Molecular Actions, Physiological Outcomes, and Clinical Targets, *Molecular Cell*. Elsevier Inc., 39(1), pp. 8–24. doi: 10.1016/j.molcel.2010.06.017.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms, *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kuny, C. V. and Sullivan, C. S. (2016) Virus–Host Interactions and the ARTD/PARP Family of Enzymes, *PLoS Pathogens*, 12(3), pp. 1–7. doi: 10.1371/journal.ppat.1005453.

Lamb, R. S., Citarelli, M. and Teotia, S. (2012) Functions of the poly(ADP-ribose) polymerase superfamily in plants, *Cellular and Molecular Life Sciences*, 69(2), pp. 175–189. doi: 10.1007/s00018-011-0793-4.

Li, G., Nasar, V., Yang, Y., Li, W., Liu, B., Sun, L., Li, D., & Song, F. (2011) Arabidopsis poly(ADP-ribose) glycohydrolase 1 is required for drought, osmotic and oxidative stress responses, *Plant Science*. Elsevier Ireland Ltd, 180(2), pp. 283–291. doi: 10.1016/j.plantsci.2010.09.002.

Liebhard, R., Kellerhals, M., Pfammatter, W., Jertmini, M., & Gessler, C. (2003) Mapping quantitative physiological traits in apple (Malus x domestica Borkh.), *Plant Molecular Biology*, 52(3), pp. 511–526. doi: 10.1023/A:1024886500979.

Lin, Z., Kong, H., Nei, M., & Ma, H. (2006) Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer, *Proceedings of the National Academy of Sciences of the United States of America*, 103(27), pp. 10328–10333. doi: 10.1073/pnas.0604232103.

Loehle, C. (1988) Tree life history strategies: the role of defenses. *Canadian Journal of Forest Research*. 18(2): 209-222. https://doi.org/10.1139/x88-032

Lorenzini, A., Johnson, F. B., Oliver, A., Tresini, M., Smith, J. S., Hdeib, M., Sell, C., Cristofalo, V. J., & Stamato, T. D. (2009) Significant correlation of species longevity

with DNA double strand break recognition but not with telomere length, *Mechanisms of Ageing and Development*. doi: 10.1016/j.mad.2009.10.004.

MacRae, S. L. *et al.* (2015a) Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human, *Aging Cell*, 14(2), pp. 288–291. doi: 10.1111/acel.12314.

MacRae, L. *et al.* (2015b) DNA repair in species with extreme lifespan differences, 7(12), pp. 1171–1182.

Manova, V. and Gruszka, D. (2015) DNA damage and repair in plants – From models to crops, *Frontiers in Plant Science*, 6(OCTOBER), pp. 1–26. doi: 10.3389/fpls.2015.00885.

Mao, R., Nie, H., Cai, D., Zhang, J., Liu, H., Yan, R., Cuconati, A., Block, T. M., Guo, J. T., & Guo, H. (2013) Inhibition of Hepatitis B Virus Replication by the Host Zinc Finger Antiviral Protein, *PLoS Pathogens*, 9(7). doi: 10.1371/journal.ppat.1003494.

Matic, I., Rayssiguier, C. and Radman, M. (1995) Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species, *Cell*, 80(3), pp. 507–515. doi: 10.1016/0092-8674(95)90501-4.

Molinier, J., Lechner, E., Dumbliauskas, E., & Genschik, P. (2008) Regulation and role of arabidopsis CUL4-DDB1A-DDB2 in maintaining genome integrity upon UV stress, *PLoS Genetics*, 4(6). doi: 10.1371/journal.pgen.1000093.

Mouchiroud, L. et al. (2013) XThe NAD+/sirtuin pathway modulates longevity through activation of mitochondrial UPR and FOXO signaling, *Cell*, 154(2), p. 430. doi: 10.1016/j.cell.2013.06.016.

Müller, S., Möller, P., Bick, M. J., Wurr, S., Becker, S., Günther, S., & Kümmerer, B. M. (2007) Inhibition of Filovirus Replication by the Zinc Finger Antiviral Protein, *Journal of Virology*, 81(5), pp. 2391–2400. doi: 10.1128/jvi.01601-06.

Murat, F., Van De Peer, Y., & Salse, J. (2012). Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biology and Evolution*, 4(9), 917–928. https://doi.org/10.1093/gbe/evs066

Okubo, S., Hara, F., Tsuchida, Y., Shimotakahara, S., Suzuki, S., Hatanaka, H., Yokoyama, S., Tanaka, H., Yasuda, H., & Shindo, H. (2004) NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers, *Journal of Biological Chemistry*, 279(30), pp. 31455–31461. doi: 10.1074/jbc.M403561200.

Pagel, M. (1997) Inferring evolutionary processes from phylogenies, *Zoologica Scripta*, 26(4), pp. 331–348. doi: 10.1111/j.1463-6409.1997.tb00423.x.

Pereira-Lorenzo, S., Ramos-Cabrer, A. M. and Fischer, M. (2009) Breeding apple (Malus × Domestica Borkh), in *Breeding Plantation Tree Crops: Temperate Species*. doi: 10.1007/978-0-387-71203-1_2.

Piovesan, G. and Biondi, F. (2020) On tree longevity, *New Phytologist*, (Box 1). doi: 10.1111/nph.17148.

Proost, S., van Bel, M., Sterck, L., Billiau, K., van Parys, T., van de Peer, Y., & Vandepoele, K. (2009) PLAZA: A comparative genomics resource to study gene and genome evolution in plants, *Plant Cell*. doi: 10.1105/tpc.109.071506.

Rissel, D., Losch, J. and Peiter, E. (2014) The nuclear protein Poly(ADP-ribose) polymerase 3 (AtPARP3) is required for seed storability in Arabidopsis thaliana, *Plant Biology*, 16(6), pp. 1058–1064. doi: 10.1111/plb.12167.

Rissel, D. and Peiter, E. (2019) Poly(ADP-ribose) polymerases in plants and their human counterparts: Parallels and peculiarities, *International Journal of Molecular Sciences*, 20(7), pp. 1–29. doi: 10.3390/ijms20071638.

Plomion, C. *et al.* (2018) Oak genome reveals facets of long lifespan, *Nature Plants*, 4(7), pp. 440–452. doi: 10.1038/s41477-018-0172-3.

Sabeeha S. Merchant et al. (2007) The Chlamydomonas Genome Reveals the Evolution of Key, *National institutes of health*, 318(5848), pp. 245–250.

Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K., & Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Biochemistry*, 73, pp. 39–85. https://doi.org/10.1146/annurev.biochem.73.011303.073723

Schreiber, V., Dantzer, F., Ame, J. C., & De Murcia, G. (2006). Poly (ADP-ribose): novel functions for an old molecule. *Nature reviews Molecular cell biology*, *7*(7), 517-528.

Schulz, P., Neukermans, J., van der Kelen, K., Mühlenbock, P., van Breusegem, F., Noctor, G., Teige, M., Metzlaff, M., & Hannah, M. A. (2012). Chemical PARP inhibition enhances growth of arabidopsis and reduces anthocyanin accumulation and the activation of stress protective mechanisms. *PLoS ONE*, 7(5). https://doi.org/10.1371/journal.pone.0037287

Singh, S. K., Roy, S., Choudhury, S. R., & Sengupta, D. N. (2010) DNA repair and recombination in higher plants: Insights from comparative genomics of arabidopsis and rice, *BMC Genomics*. doi: 10.1186/1471-2164-11-443.

Song, J. and Bent, A. F. (2014) Microbial Pathogens Trigger Host DNA Double-Strand Breaks Whose Abundance Is Reduced by Plant Defense Responses, *PLoS Pathogens*, 10(4). doi: 10.1371/journal.ppat.1004030.

Song, J., Keppler, B. D., Wise, R. R. and Bent, A. F. (2015) PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in Arabidopsis DNA Damage and Immune Responses, *PLoS Genetics*, 11(5), pp. 1–24. doi: 10.1371/journal.pgen.1005200.

Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*, 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Storey, J. D. (2002) A direct approach to false discovery rates, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(3), pp. 479–498. doi: 10.1111/1467-9868.00346.

J. Storey, A.J. Bass, A. Dabney, D. Robinson. 2015. qvalue: Q-value estimation for false discovery rate control. R Package. http://github.com/jdstorey/qvalue

Storozhenko, S., Inzé, D., Van Montagu, M., & Kushnir, S. (2001). Arabidopsis coactivator ALY-like proteins, DIP1 and DIP2, interact physically with the DNA-binding domain of the Zn-finger poly(ADP-ribose) polymerase. *Journal of Experimental Botany*, 52(359), 1375–1380. https://doi.org/10.1093/jxb/52.359.1375

Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., Emes, R. D., & Lynch, V. J. (2016) TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants, *eLife*, 5(September2016), pp. 1–30. doi: 10.7554/eLife.11994.

Takasaki, Y., Seki, Y., Nojima, H. & Isoda, A. (1994) Growth of an Annual Strain of Oryza glaberrima Steud. and a Perennial Cultivar of Oryza sativa L. after Heading, *Japanese Journal of Crop Science*. doi: 10.1626/jcs.63.632.

Tian, X., Seluanov, A. and Gorbunova, V. (2017) Molecular Mechanisms Determining Lifespan in Short- and Long-Lived Species, *Trends in Endocrinology and Metabolism*. Elsevier Ltd, 28(10), pp. 722–734. doi: 10.1016/j.tem.2017.07.004.

Tivoli, B., Baranger, A., Sivasithamparam, K. & Barbetti, M. J. (2006) Annual Medicago: From a model crop challenged by a spectrum of necrotrophic pathogens to a model plant to explore the nature of disease resistance, *Annals of Botany*, 98(6), pp. 1117–1128. doi: 10.1093/aob/mcl132.

Tobias, P. A. and Guest, D. I. (2014) Tree immunity: Growing old without antibodies, *Trends in Plant Science*. Elsevier Ltd, 19(6), pp. 367–370. doi: 10.1016/j.tplants.2014.01.011.

Tung Ho, L. S. & Ané, C. (2014) A linear-time algorithm for gaussian and non-gaussian trait evolution models, *Systematic Biology*, 63(3), pp. 397–408. doi: 10.1093/sysbio/syu005.

Umeda, M., Ikeuchi, M., Ishikawa, M., Ito, T., Nishihama, R., Kyozuka, J., Torii, K. U., Satake, A., Goshima, G., & Sakakibara, H. (2021). Plant stem cell research is uncovering the secrets of longevity and persistent growth. *Plant Journal*, 326–335. https://doi.org/10.1111/tpj.15184

Vainonen, J. P., Shapiguzov, A., Vaattovaara, A., & Kangasjärvi, J. (2016) Plant PARPs, PARGs and PARP-like Proteins, *Current Protein & Peptide Science*, 17(7), pp. 713–723. doi: 10.2174/1389203717666160419144721.

Van Baren, M. J. *et al.* (2016) Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants, *BMC Genomics*, 17(1), pp. 1–22. doi: 10.1186/s12864-016-2585-6.

Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., & Vandepoele, K. (2012) Dissecting plant genomes with the PLAZA comparative genomics platform, *Plant Physiology*. doi: 10.1104/pp.111.189514.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F., & Vandepoele, K. (2018) PLAZA 4.0: An integrative resource for functional,

evolutionary and comparative plant genomics, *Nucleic Acids Research*. Oxford University Press, 46(D1), pp. D1190–D1196. doi: 10.1093/nar/gkx1002.

Vanderauwera, S., De Block, M., Van De Steene, N., Van De Cotte, B., Metzlaff, M., & Van Breusegem, F. (2007). Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*, 104(38), 15150–15155. https://doi.org/10.1073/pnas.0706668104

Watson, J. M. and Riha, K. (2010) Comparative biology of telomeres: Where plants stand, *FEBS Letters*. Federation of European Biochemical Societies, 584(17), pp. 3752–3759. doi: 10.1016/j.febslet.2010.06.017.

White, O. *et al.* (1999) Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1, *Science*, 286(5444), pp. 1571–1577. doi: 10.1126/science.286.5444.1571.

Yoshiyama, K. O., Sakaguchi, K. and Kimura, S. (2013) DNA damage response in plants: Conserved and variable response compared to animals, *Biology*, 2(4), pp. 1338–1356. doi: 10.3390/biology2041338.

Zhang, G. *et al.* (2013) Comparative Analysis of Bat Genomes, *Science*, 339(January), pp. 456–460.

Zhang, H., Gu, Z., Wu, Q., Yang, L., Liu, C., Ma, H., Xia, Y., & Ge, X. (2015). Arabidopsis PARG1 is the key factor promoting cell survival among the enzymes regulating post-translational poly(ADP-ribosyl)ation, *Scientific Reports*. Nature Publishing Group, 5(October 2014), pp. 1–13. doi: 10.1038/srep15892.

Zhang, Q. *et al.* (2010) Salt Cress. A Halophyte and Cryophyte Arabidopsis Relative Model System and Its Applicability to Molecular Genetic Analyses of Growth and Development of Extremophiles 1, *Plant Physiol.*, 135(July 2004), pp. 1718–1737. doi: 10.1104/pp.104.041723.1718.

Zhang, X. C., H. P. Nooteboom & M. Kato. 2013. (2013) 1. SELAGINELLA P. Beauvois, Prodr. Aethéogam. 101. 1805, nom. cons., 2(June), pp. 37–66.

## TABLES

**Table 1.** List of plant species in the dataset. 61 plant species including trees, perennial herbs, annual herbs and algae were used for analyses. Two alga species (*Chlamydomonas reinhardtii* and *Micromonas commoda*) were eliminated from the analyses considering the phylogenetic relationships (PGLS analyses) because the no sequence data of these species were available.

| Life form | Species name |
|---|---|
| Tree: 23 species | |
| Angiosperm | *Amborella trichopoda* |
| | *Carica papaya* |
| | *Citrus clementina* |
| | *Coffea canephora* |
| | *Eucalyptus grandis* |
| | *Hevea brasiliensis* |
| | *Malus domestica* |
| | *Populus trichocarpa* |
| | *Prunus persica* |
| | *Pyrus bretschneideri* |
| | *Theobroma cacao* |
| | *Ziziphus jujuba* |
| Gymnosperm | *Cycas micholitzii* |
| | *Ginkgo biloba* |
| | *Gnetum montanum* |
| | *Picea abies* |
| | *Picea glauca* |
| | *Picea sitchensis* |
| | *Pinus pinaster* |
| | *Pinus sylvestris* |
| | *Pinus taeda* |
| | *Pseudotsuga menziesii* |
| | *Taxus baccata* |
| Perennial herb: 15 species | |
| | *Arabidopsis lyrata* |

| | |
|---|---|
| | *Brassica oleracea* |
| | *Cajanus cajan* |
| | *Capsicum annuum* |
| | *Erythranthe guttata* |
| | *Fragaria vesca* |
| | *Marchantia polymorpha* |
| | *Nelumbo nucifera* |
| | *Oryza sativa* ssp. *japonica* |
| | *Ricinus communis* |
| | *Selaginella moellendorffii* |
| | *Solanum lycopersicum* |
| | *Solanum tuberosum* |
| | *Trifolium pratense* |
| | *Utricularia gibba* |
| Annual herb: 21 species | |
| | *Amaranthus hypochondriacus* |
| | *Arabidopsis thaliana* |
| | *Arachis ipaensis* |
| | *Beta vulgaris* |
| | *Brassica rapa* |
| | *Capsella rubella* |
| | *Chenopodium quinoa* |
| | *Cicer arietinum* |
| | *Citrullus lanatus* |
| | *Corchorus olitorius* |
| | *Cucumis melo* |
| | *Cucumis sativus L.* |
| | *Daucus carota* |
| | *Glycine max* |
| | *Medicago truncatula* |
| | *Petunia axillaris* |
| | *Physcomitrella patens* |
| | *Schrenkiella parvula* |
| | *Tarenaya hassleriana* |
| | *Vigna radiata* var. *radiata* |
| | *Zea mays* |

| Alga: 2 species | |
|---|---|
| | *Chlamydomonas reinhardtii* |
| | *Micromonas commoda* |

**Table 2.** The result of PGLS regressions to compare the copy number ratios among life forms for each paralog of *PARP* gene family.

| Gene | Trees versus annual herbs | | | | Trees versus perennial herbs | | | | Pagel's lambda |
|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard error | *t*-value | Q-value | Coefficient | Standard error | *t*-value | Q-value | |
| *PARP1* | –0.280 | 0.131 | –2.135 | 0.0557 | –0.541 | 0.161 | –3.361 | 0.00285 | $7.54\times10^{-9}$ |
| *PARP2* | –0.973 | 0.209 | –4.655 | $6.1\times10^{-5}$ | –0.686 | 0.21 | –3.263 | 0.00285 | 0.878 |
| *PARP3* | –0.316 | 0.175 | –1.80 | 0.0765 | –0.2.07 | 0.215 | –0.964 | 0.339 | $8.04\times10^{-9}$ |

**Table 3.** The result of regressions to investigate the relationships between growth rate and copy number ratio of *PARP* in 11 tree species by phylogenetic generalized least squares (PGLS) regressions.

| | Coefficient | Standard error | *t*-value | Q-value | Pagel's lambda |
|---|---|---|---|---|---|
| *PARP1* | –0.698 | 0.194 | –3.599 | 0.0173 | 1 |
| *PARP2* | –0.231 | 0.441 | –0.524 | 0.613 | 0.323 |
| *PARP3* | –0.181 | 0.0770 | –2.349 | 0.0651 | 0.110 |
| All *PARP* | –0.618 | 0.173 | –3.571 | 0.00601 | 0.399 |

**FIGURES**

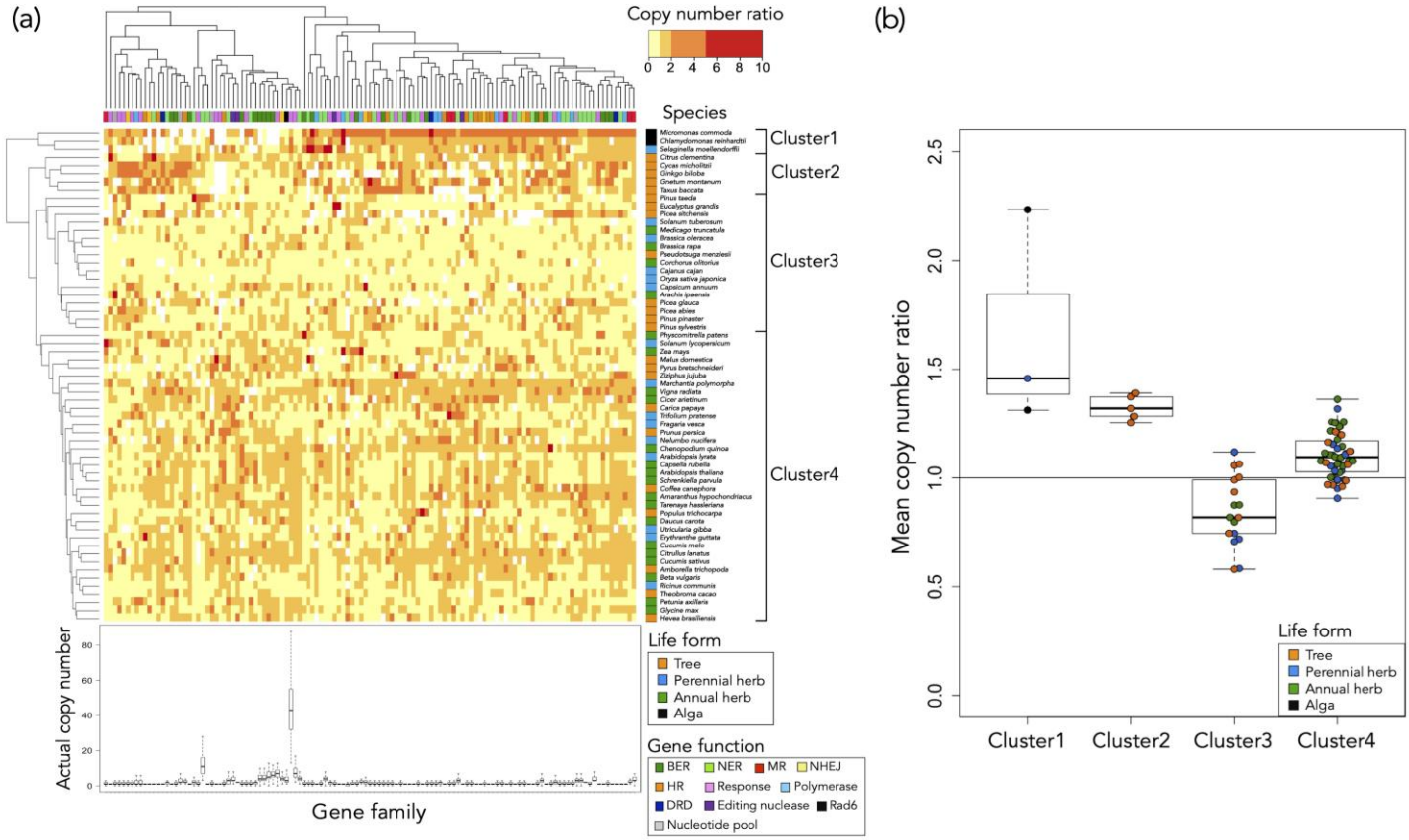



**Figure 1.** Interspecies comparison of copy number ratio of 121 DNA repair gene families. (a) Clustered heatmap of the copy number ratio of 121 DNA repair gene families. Hierarchical clustering was performed based on the Euclidian distance of the copy number ratio of each species using the Ward's method. 23 tree species, 15 perennial herb species, 21 annual herb species, and two alga species were included, and the life form of each species was in colored. Each gene family was categorized into one of 11 groups, and the function of each gene family was in colored: BER, base excision repair; NER, nucleotide excision repair; MR, mismatch repair; NHEJ, nonhomologous end-joining repair; HR, homologous recombination repair; Response, DNA damage response; Polymerase, DNA polymerase; DRD, direct reversal of damage; Editing

nuclease, editing and processing nuclease; Rad6, Rad6 pathway; Nucleotide pool, modulation of nucleotide pool. The actual copy number within each gene family is shown at the bottom of the figure. (b) Mean copy number ratios of 121 DNA repair gene families of species in the cluster. The horizontal line inside the box showed the median and the length of box showed the interquartile range (range between the 25th to 75th percentiles). The whiskers indicated points within 1.5 times the interquartile rage. The colors of the points correspond to the life form of the species.

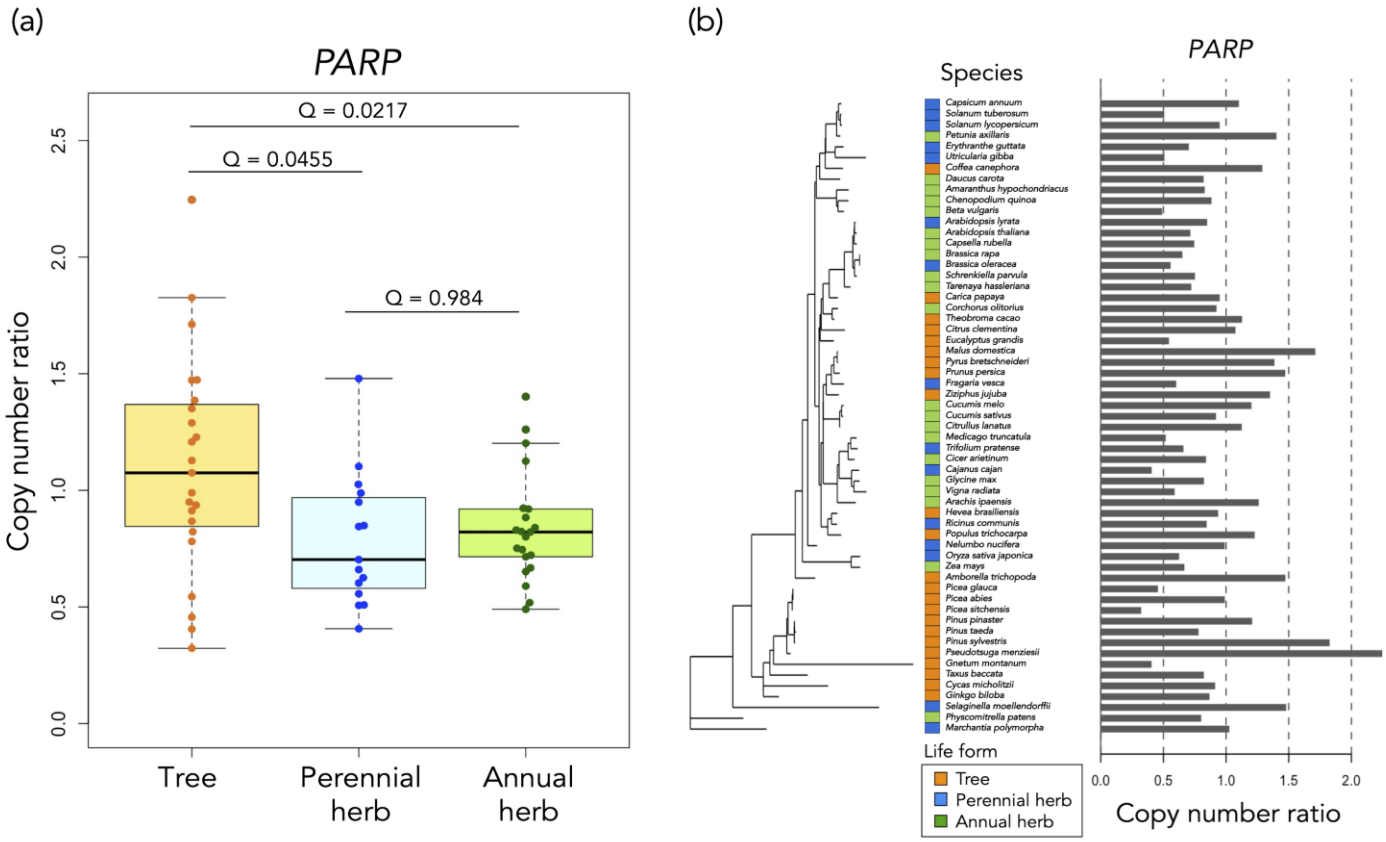



**Figure 2.** The result of phylogenetic generalized least squares regressions. (a) The copy number ratio of *PARP* in each life form. The result of the PGLS regressions showed that tree species had significantly higher copy number ratios in the *PARP* gene family compared to perennial herb species (coefficient = −0.395, standard error = 0.111, *t*-value = −3.560, P-value = $7.659 \times 10^{-4}$, Q-value = 0.0455) and annual herb species (coefficient = −0.363, standard error = 0.090, *t*-value = −4.014, P-value = $1.794 \times 10^{-4}$, Q-value = 0.0217). The horizontal line inside the box showed the median and the length of box showed the interquartile range (range between the 25th to 75th percentiles). The whiskers indicated points within 1.5 times the interquartile rage. The points beyond the whisker range indicated the outliers. (b)The phylogenetic relationships in the copy number ratio of *PARP*. The estimated Pagel's lambda was $4.97 \times 10^{-9}$.

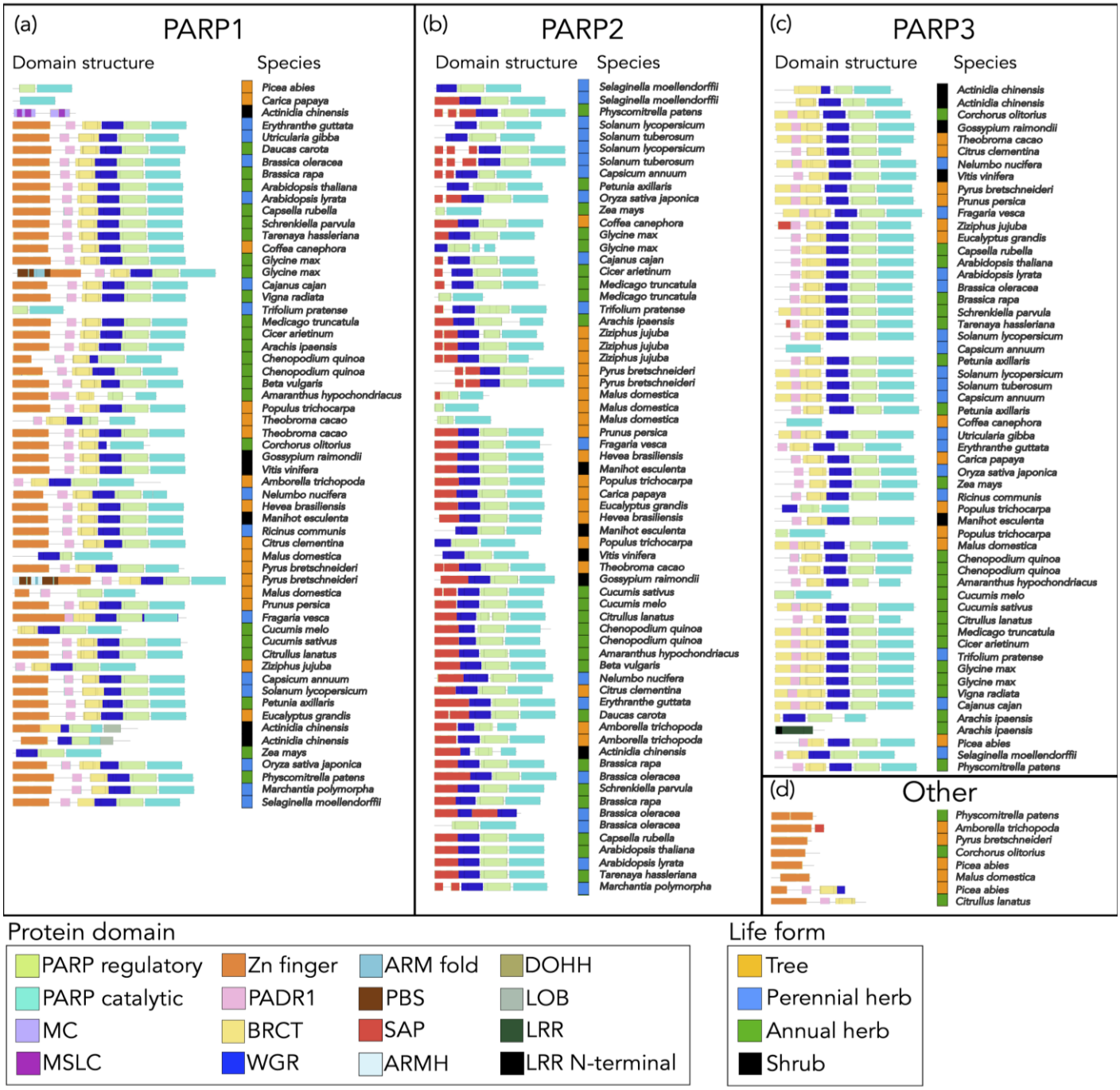



**Figure 3.** The protein domain structures of PARPs of species in Dicots PLAZA 4.0 dataset. Each PARP was categorized into four groups (a) PARP1, (b) PARP2, (c) PARP3 and (d) Other based on the annotations in Dicots PLAZA 4.0 and the phylogenetic tree constructed by the tree explore tool in Dicots PLAZA 4.0. Protein

domains are illustrated by colored. PARP regulatory: Poly(ADP-ribose) polymerase regulatory domain, PARP catalytic: Poly(ADP-ribose) polymerase catalytic domain, MC: Mitochondrial carrier domain, MSLC: Mitochondrial substrate/solute carrier domain, Zn finger: zinc-finger domain, PADR1: PADR1 domain, BRCT: BRCA1 C terminus domain, WGR: tryptophan-glycine–arginine-rich domain, ARM fold: Armadillo-type fold domain, PBS: PBS lyase HEAT-like repeat domain, SAP: SAF-A/B, Acinus and PIAS domain, ARMH: Armadillo-like helical domain, DOHH: Deoxyhypusine hydroxylase domain, LOB: Lateral organ boundaries domain, LRR: Leucine-rich repeat domain, LRR N-terminal: Leucine-rich repeat-containing N-terminal domain.

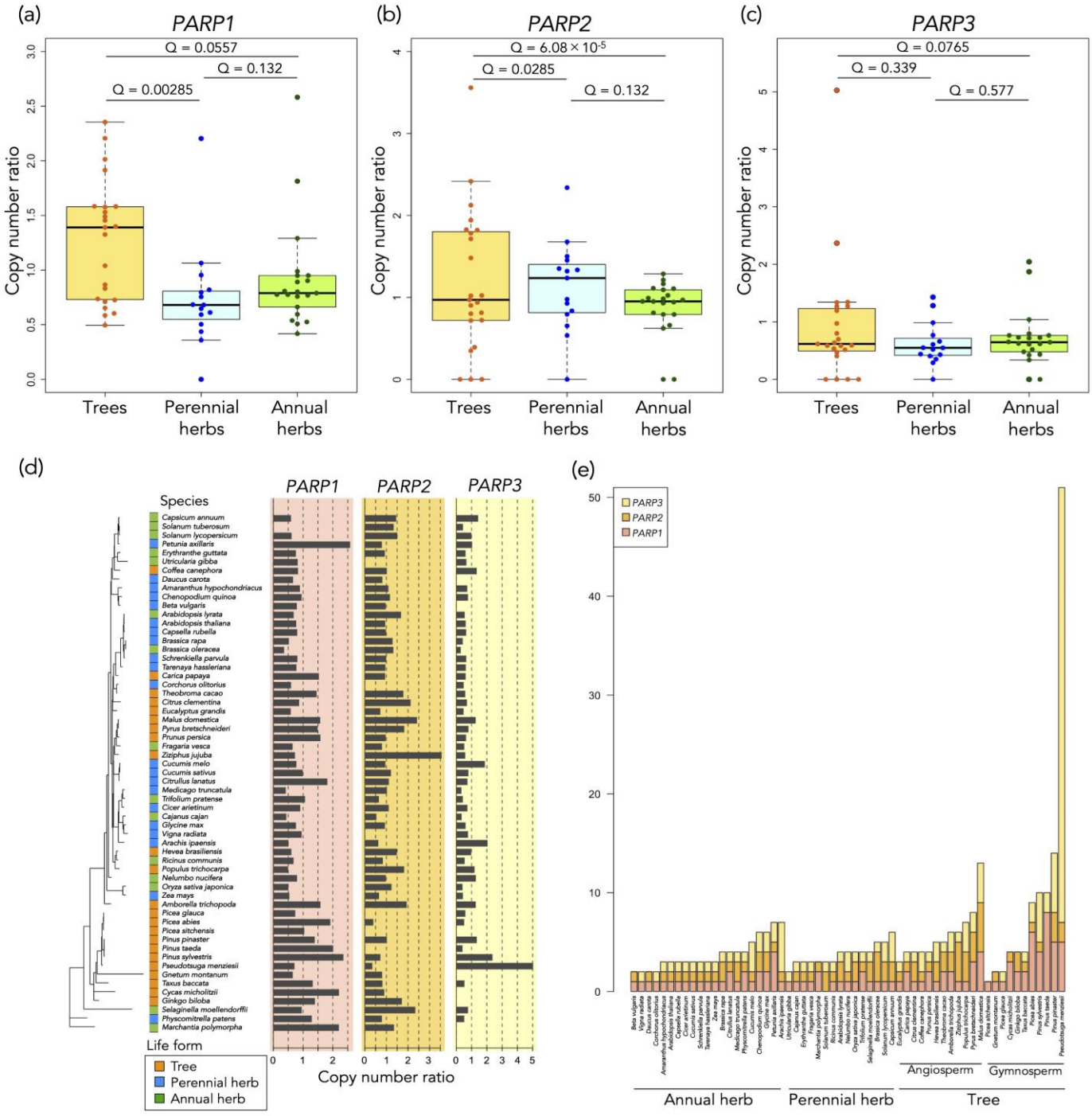



**Figure 4.** Comparison analyses for each type of *PARP*. Comparison of copy number ratios in *PARP1* (a), *PARP2* (b) and *PARP3* (c) among life forms by PGLS regressions. Tree species had significantly higher copy number ratio than perennial herb

species and annual herb species in *PARP1*. Also, tree species had significantly higher copy number ratios in *PARP2* than perennial herb species and annual herb species. The copy number ratios of *PARP3* in tree species were not significantly different compared to perennial herb species and annual herb species. The horizontal line inside the box showed the median and the length of box showed the interquartile range (range between the 25th to 75th percentiles). The whiskers indicated points within 1.5 times the interquartile rage. The points beyond the whisker range indicated the outliers. (d) The phylogenetic relationships of copy number ratios in *PARP1*, *PARP2*, and *PARP3*. (e) The actual copy number of *PARP* genes in the species.

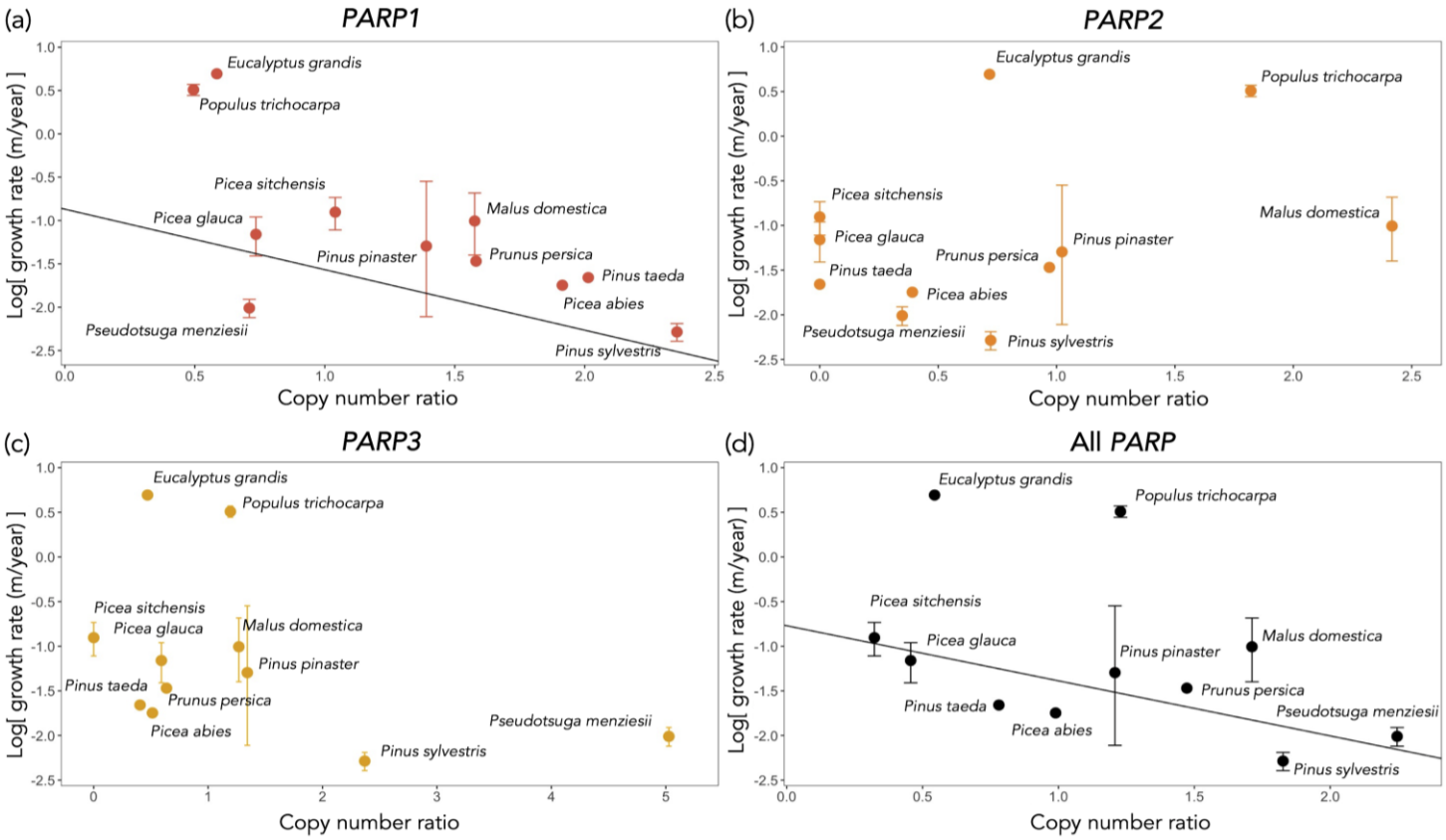



**Figure 5.** The relationships between growth rate and copy number ratio of each *PARP* in 11 tree species. Plots showed the average height growth rate (m/year) and vertical bar showed the highest and lowest growth rate of the species. There were significantly negative correlations between the copy number ratio and log growth rate in *PARP1* (Q-value was 0.0173 in PGLS) (a) and all type of *PARP* including *PARP1*, *PARP2* and *PARP3* (Q-value was 0.00601 in PGLS) (d). There were no significant relationships between the copy number ratio and log growth rate in *PARP2* (b) and *PARP3* (c).

## APPENDIXES

**Appendix Table S1.** The list of DNA repair genes used for the analyses. The accession number in Arabidopsis thaliana and ID of gene family in the Dicots PLAZA 4.0 and Gymno PLAZA 1.0 were shown. Each gene was categorized into 11 functional groups.

| Group of gene function | Symbol of gene | AGI Accession number in *Arabidopsis thaliana* | ID of gene family in Dicots PLAZA 4.0 | ID of gene family in Gymno PLAZA 1.0 |
|---|---|---|---|---|
| Base excision repair | *APE1* | AT2G41460 | HOM04D004383 | HOM03D004400 |
| | *APE1L* | AT3G48425 | HOM04D006817 | HOM03D005832 |
| | *APE2* | AT4G36050 | HOM04D004425 | HOM03D006661 |
| | *APTX* | AT5G01310 | HOM04D004756 | HOM03D002833 |
| | *DML1* | AT2G36490 | HOM04D001046 | HOM03D001428 |
| | *DML2* | AT3G10010 | HOM04D001046 | HOM03D001428 |
| | *DML3* | AT4G34060 | HOM04D001046 | HOM03D001428 |
| | *FPG* | AT1G52500 | HOM04D005473 | HOM03D004609 |
| | *HMGB1* | AT3G51880 | HOM04D000711 | HOM03D000500 |
| | *MAGLP/AlkA* | AT1G19480 | HOM04D002929 | HOM03D004685 |
| | *MAGLP/AlkA* | AT1G75230 | HOM04D002929 | HOM03D004685 |
| | *MAGLP/AlkA* | AT3G50880 | HOM04D002929 | HOM03D004685 |
| | *MBD4* | AT3G07930 | HOM04D004958 | HOM03D003502 |
| | *MPG/MAG* | AT3G12040 | HOM04D007180 | HOM03D007182 |
| | *MUTY* | AT4G12740 | HOM04D005552 | HOM03D004454 |
| | *NTH* | AT1G05900 | HOM04D004019 | HOM03D005173 |
| | *NTH* | AT2G31450 | HOM04D004019 | HOM03D005173 |
| | *OGG1* | AT1G21710 | HOM04D006939 | HOM03D005744 |
| | *PARG1* | AT2G31870 | HOM04D003287 | HOM03D003504 |
| | *PARG2* | AT2G31865 | HOM04D003287 | HOM03D003504 |
| | *PARP1* | AT2G31320 | HOM04D001195 | HOM03D000597 |
| | *PARP2* | AT4G02390 | HOM04D001195 | HOM03D000597 |
| | *PARP3* | AT5G22470 | HOM04D001195 | HOM03D000597 |
| | *PNKP* | AT3G14890 | HOM04D005170 | HOM03D004809 |
| | *Tag* | AT1G13635 | HOM04D000784 | HOM03D001279 |

| | | | | |
|---|---|---|---|---|
| | *Tag* | AT1G15970 | HOM04D000784 | HOM03D001279 |
| | *Tag* | AT1G75090 | HOM04D000784 | HOM03D001279 |
| | *Tag* | AT1G80850 | HOM04D000784 | HOM03D001279 |
| | *Tag* | AT3G12710 | HOM04D000784 | HOM03D001279 |
| | *Tag* | AT5G44680 | HOM04D000784 | HOM03D001279 |
| | *Tag* | AT5G57970 | HOM04D000784 | HOM03D001279 |
| | *TDP1* | AT5G15170 | HOM04D005673 | HOM03D004707 |
| | *UNG* | AT3G18630 | HOM04D003441 | HOM03D003393 |
| | *XRCC1* | AT1G80420 | HOM04D006984 | HOM03D003667 |
| Nucleotide excision repair | *CCNH* | AT5G27620 | HOM04D005036 | HOM03D003364 |
| | *CSA* | AT1G19750 | HOM04D005364 | HOM03D005285 |
| | *CSA* | AT1G27840 | HOM04D005364 | HOM03D005285 |
| | *CUL4* | AT5G46210 | HOM04D000338 | HOM03D000143 |
| | *DDB1* | AT4G05420 | HOM04D003108 | HOM03D000591 |
| | *DDB1* | AT4G21100 | HOM04D003108 | HOM03D000591 |
| | *DDB2* | AT5G58760 | HOM04D007014 | HOM03D003898 |
| | *GTF2H1* | AT1G55750 | HOM04D004318 | HOM03D003099 |
| | *GTF2H1* | AT3G61420 | HOM04D004318 | HOM03D003099 |
| | *GTF2H2* | AT1G05055 | HOM04D006174 | HOM03D006192 |
| | *GTF2H3* | AT1G18340 | HOM04D006212 | HOM03D006663 |
| | *GTF2H4* | AT4G17020 | HOM04D005140 | HOM03D003940 |
| | *GTF2H5* | AT1G12400 | HOM04D007085 | HOM03D008072 |
| | *GTF2H5* | AT1G62886 | HOM04D007085 | HOM03D008072 |
| | *LIG1* | AT1G08130 | HOM04D001683 | HOM03D001412 |
| | *LIG1* | AT1G49250 | HOM04D001683 | HOM03D001412 |
| | *Mfd* | AT3G02060 | HOM04D005818 | HOM03D003238 |
| | *MMS19* | AT5G48120 | HOM04D004480 | HOM03D004191 |
| | *MNAT1* | AT4G30820 | HOM04D005360 | HOM03D004449 |
| | *RAD1/UVH1/ERCC4/XPF* | AT5G41150 | HOM04D005466 | HOM03D003505 |
| | *RAD23A* | AT1G79650 | HOM04D001203 | HOM03D001632 |
| | *RAD23B* | AT1G16190 | HOM04D001203 | HOM03D001632 |
| | *RAD23C* | AT3G02540 | HOM04D001203 | HOM03D001632 |
| | *RAD23D* | AT5G38470 | HOM04D001203 | HOM03D001632 |
| | *RBX1* | AT3G42830 | HOM04D001544 | HOM03D001542 |
| | *RBX1* | AT5G20570 | HOM04D001544 | HOM03D001542 |
| | *RFC1* | AT5G22010 | HOM04D004689 | HOM03D002834 |

| | | | | |
|---|---|---|---|---|
| | *RFC2* | AT1G21690 | HOM04D001345 | HOM03D001196 |
| | *RFC3* | AT1G77470 | HOM04D001345 | HOM03D001196 |
| | *RFC4* | AT1G63160 | HOM04D001345 | HOM03D001196 |
| | *RFC5* | AT5G27740 | HOM04D001694 | HOM03D003877 |
| | *RPA1* | AT2G06510 | HOM04D000929 | HOM03D000629 |
| | *RPA1* | AT4G19130 | HOM04D000929 | HOM03D000629 |
| | *RPA1* | AT5G08020 | HOM04D000929 | HOM03D000629 |
| | *RPA1* | AT5G45400 | HOM04D000929 | HOM03D000629 |
| | *RPA1* | AT5G61000 | HOM04D000929 | HOM03D000629 |
| | *RPA2* | AT2G24490 | HOM04D002638 | HOM03D003134 |
| | *RPA2* | AT3G02920 | HOM04D002638 | HOM03D003134 |
| | *RPA3* | AT3G52630 | HOM04D003942 | HOM03D005396 |
| | *RPA3* | AT4G18590 | HOM04D003942 | HOM03D005396 |
| | *UVR1/UVX3/XPG/ERCC5* | AT3G28030 | HOM04D005866 | HOM03D002893 |
| | *UVR7/ERCC1* | AT3G05210 | HOM04D005591 | HOM03D004203 |
| | *UvrD* | AT4G25120 | HOM04D002964 | HOM03D005360 |
| | *XAB2* | AT5G28740 | HOM04D003069 | HOM03D002694 |
| | *XPB/ERCC3* | AT5G41360 | HOM04D003675 | HOM03D002803 |
| | *XPB/ERCC3* | AT5G41370 | HOM04D003675 | HOM03D002803 |
| | *XPC* | AT5G16630 | HOM04D005966 | HOM03D004314 |
| | *XPD/UVH6/ERCC2* | AT1G03190 | HOM04D004614 | HOM03D005289 |
| Homologous recombination repair | *BRCA2* | AT4G00020 | HOM04D004670 | HOM03D008142 |
| | *BRCA2* | AT5G01630 | HOM04D004670 | HOM03D008142 |
| | *EME1* | AT2G21800 | HOM04D005249 | HOM03D007551 |
| | *EME1* | AT2G22140 | HOM04D005249 | HOM03D007551 |
| | *MIM* | AT5G61460 | HOM04D003618 | HOM03D003447 |
| | *MND1* | AT4G29170 | HOM04D005684 | HOM03D007966 |
| | *MRE11A* | AT5G54260 | HOM04D004854 | HOM03D005935 |
| | *MUS81* | AT4G30870 | HOM04D004990 | HOM03D004705 |
| | *NBS1* | AT3G02680 | HOM04D006113 | HOM03D004683 |
| | *RAD50* | AT2G31970 | HOM04D005302 | HOM03D003113 |
| | *RAD51B* | AT2G28560 | HOM04D007144 | HOM03D007435 |
| | *RAD51C* | AT2G45280 | HOM04D007012 | HOM03D007195 |
| | *RAD51D* | AT1G07745 | HOM04D006740 | HOM03D007750 |
| | *RecG* | AT2G01440 | HOM04D003779 | HOM03D003370 |
| | *SSB* | AT3G18580 | HOM04D002728 | HOM03D002499 |

| | | | | |
|---|---|---|---|---|
| | *SSB* | AT4G11060 | HOM04D002728 | HOM03D002499 |
| | *TOP3* | AT2G32000 | HOM04D002223 | HOM03D002059 |
| | *TOP3* | AT5G63920 | HOM04D002223 | HOM03D002059 |
| | *XRCC2* | AT5G64520 | HOM04D006906 | HOM03D008620 |
| Mismatch repair | *MLH1* | AT4G09140 | HOM04D005281 | HOM03D005583 |
| | *MLH3* | AT4G35520 | HOM04D003331 | HOM03D005080 |
| | *MSH1* | AT3G24320 | HOM04D004513 | HOM03D005511 |
| | *MSH5* | AT3G20475 | HOM04D005333 | HOM03D007428 |
| | *Muts_like* | AT1G65070 | HOM04D001403 | HOM03D001852 |
| | *Muts_like* | AT5G54090 | HOM04D001403 | HOM03D001852 |
| | *PMS1* | AT4G02460 | HOM04D002177 | HOM03D002554 |
| Non-homologous end-joining repair | *ATRAD21.1* | AT5G40840 | HOM04D001275 | HOM03D001079 |
| | *ATRAD21.2* | AT3G59550 | HOM04D001275 | HOM03D001079 |
| | *ATRAD21.3* | AT5G16270 | HOM04D001275 | HOM03D001079 |
| | *KU70* | AT1G16970 | HOM04D005046 | HOM03D004691 |
| | *KU80* | AT1G48050 | HOM04D005174 | HOM03D002193 |
| | *LIG4* | AT5G57160 | HOM04D005047 | HOM03D002488 |
| | *PRKDC* | AT1G50030 | HOM04D002601 | HOM03D001652 |
| | *XRCC4* | AT3G23100 | HOM04D006340 | HOM03D005209 |
| Editing and processing nuclease | *FEN1* | AT5G26680 | HOM04D003408 | HOM03D002630 |
| | *FLJ35220* | AT4G31150 | HOM04D005935 | HOM03D006237 |
| | *HEX1/EXO1* | AT1G18090 | HOM04D002577 | HOM03D005538 |
| | *HEX1/EXO1* | AT1G29630 | HOM04D002577 | HOM03D005538 |
| | *SPO11-1* | AT3G13170 | HOM04D001259 | HOM03D001513 |
| | *SPO11-2* | AT1G63990 | HOM04D001259 | HOM03D001513 |
| | *SPO11-3* | AT5G02820 | HOM04D001259 | HOM03D001513 |
| Modulation of nucleotide pool | *DUT1* | AT3G46940 | HOM04D003033 | HOM03D002613 |
| | *NUDX1* | AT1G68760 | HOM04D003418 | HOM03D005023 |
| | *RNR1* | AT2G21790 | HOM04D002376 | HOM03D001347 |
| | *RNR2a* | AT3G23580 | HOM04D002018 | HOM03D001558 |
| | *TSO2* | AT3G27060 | HOM04D002018 | HOM03D001558 |
| DNA plymerase | *NUDX1* | AT1G68760 | HOM04D003418 | HOM03D005023 |
| | *POLD2* | AT2G42120 | HOM04D005157 | HOM03D004054 |
| | *POLD3* | AT1G78650 | HOM04D002072 | HOM03D004484 |
| | *POLD4* | AT1G09815 | HOM04D004732 | HOM03D003548 |
| | *POLE* | AT1G08260 | HOM04D003276 | HOM03D002351 |

| | | | | |
|---|---|---|---|---|
| | *POLE* | AT2G27120 | HOM04D003276 | HOM03D002351 |
| | *POLE* | AT5G22110 | HOM04D004989 | HOM03D007043 |
| | *POLH* | AT5G44740 | HOM04D004091 | HOM03D007442 |
| | *Polk* | AT1G49980 | HOM04D002775 | HOM03D006067 |
| | *POLL* | AT1G10520 | HOM04D006123 | HOM03D007561 |
| | *REV1* | AT5G44750 | HOM04D004212 | HOM03D005524 |
| | *REV7* | AT1G16590 | HOM04D006848 | HOM03D005648 |
| Rad6 pathway | *MMS2* | AT1G23260 | HOM04D001492 | HOM03D001161 |
| | *MMS2* | AT1G70660 | HOM04D001492 | HOM03D001161 |
| | *MMS2* | AT2G36060 | HOM04D001492 | HOM03D001161 |
| | *MMS2* | AT3G52560 | HOM04D001492 | HOM03D001161 |
| Direct reversal of damage | *ABH3/AlkB* | AT2G22260 | HOM04D007234 | HOM03D007275 |
| | *AlkB* | AT1G11780 | HOM04D006501 | HOM03D006029 |
| | *PHR1* | AT1G12370 | HOM04D005911 | HOM03D005566 |
| DNA damage response | *AXR1* | AT1G05180 | HOM04D003724 | HOM03D003484 |
| | *BRU1* | AT3G18730 | HOM04D004030 | HOM03D008954 |
| | *CHEK2* | AT4G04720 | HOM04D000039 | HOM03D000063 |
| | *COP1* | AT2G32950 | HOM04D000650 | HOM03D000501 |
| | *DET1* | AT4G10180 | HOM04D005851 | HOM03D003960 |
| | *DRT101* | AT5G18070 | HOM04D004359 | HOM03D004660 |
| | *DRT102* | AT3G04880 | HOM04D006441 | HOM03D003323 |
| | *DRT111* | AT1G30480 | HOM04D004921 | HOM03D003999 |
| | *HUS1* | AT1G52530 | HOM04D004876 | HOM03D005957 |
| | *PR19B/PUB60-1* | AT1G04510 | HOM04D003246 | HOM03D004531 |
| | *PR19B/PUB60-2* | AT2G33340 | HOM04D003246 | HOM03D004531 |
| | *PRD1* | AT4G14180 | HOM04D006666 | HOM03D007084 |
| | *RAD1* | AT4G17760 | HOM04D006209 | HOM03D007251 |
| | *RAD17* | AT5G66130 | HOM04D005902 | HOM03D006532 |
| | *RAD9* | AT3G05480 | HOM04D005486 | HOM03D007064 |
| | *REX1* | AT5G04910 | HOM04D006322 | HOM03D006889 |
| | *SMC1* | AT3G54670 | HOM04D003489 | HOM03D003237 |
| | *SMC3* | AT2G27170 | HOM04D003467 | HOM03D002271 |
| | *SMC4* | AT5G48600 | HOM04D003434 | HOM03D002909 |
| | *SMC5* | AT5G15920 | HOM04D004387 | HOM03D001853 |
| | *SMC6* | AT5G07660 | HOM04D003618 | HOM03D003447 |
| | *SOG1* | AT1G25580 | HOM04D000656 | HOM03D000769 |

| | | | | |
|---|---|---|---|---|
| | *SSRP1* | AT3G28730 | HOM04D003180 | HOM03D002008 |
| | *WRN* | AT4G13870 | HOM04D006594 | HOM03D006683 |

**Appendix Table S2.** The results of Fisher's exact test to test enrichment or dilution of each life form in each of significantly different cluster.

| | | Number of target life form in target cluster | Number of target life form in all species | p-values | Q-values |
|---|---|---|---|---|---|
| Cluster1 | Tree | 0 | 23 | 0.284 | 0.621 |
| | Perennial herb | 1 | 15 | 1 | 1 |
| | Annual herb | 0 | 21 | 0.545 | 0.793 |
| | Alga | 2 | 2 | 0.00164 | 0.0262 |
| Cluster2 | Tree | 5 | 23 | 0.00566 | 0.0452 |
| | Perennial herb | 0 | 15 | 0.321 | 0.621 |
| | Annual herb | 0 | 21 | 0.154 | 0.437 |
| | Alga | 0 | 2 | 1 | 1 |
| Cluster3 | Tree | 8 | 23 | 0.388 | 0.621 |
| | Perennial herb | 5 | 15 | 0.741 | 0.988 |
| | Annual herb | 4 | 21 | 0.371 | 0.621 |
| | Alga | 0 | 2 | 1 | 1 |
| Cluster4 | Tree | 10 | 23 | 0.0658 | 0.263 |
| | Perennial herb | 9 | 15 | 1 | 1 |
| | Annual herb | 17 | 21 | 0.0145 | 0.0774 |
| | Alga | 0 | 2 | 0.164 | 0.437 |

**Appendix Table S3.** The results of PGLS regressions in 121 gene families.

(A) Trees versus annual herbs

| Symbol of gene family | ID of gene family in Dicots PLAZA 4.0 | ID of gene family in Gymno PLAZA 1.0 | Genes within the gene family | Coefficient | Standard error | t-value | p-value | Q-value |
|---|---|---|---|---|---|---|---|---|
| PARP | HOM04D001195 | HOM03D000597 | PARP1, PARP2, PARP3 | -0.363 | 0.090 | -4.014 | 0.000 | 0.022 |
| RNR2, TSO2 | HOM04D002018 | HOM03D001558 | RNR2a, TSO2 | -0.651 | 0.201 | -3.239 | 0.002 | 0.081 |
| BRCA2 | HOM04D004670 | HOM03D008142 | BRCA2 | 0.672 | 0.206 | 3.254 | 0.002 | 0.081 |
| DRT102 | HOM04D006441 | HOM03D003323 | DRT102 | -0.674 | 0.255 | -2.639 | 0.011 | 0.263 |
| MPG/MAG | HOM04D007180 | HOM03D007182 | MPG/MAG | 0.322 | 0.128 | 2.505 | 0.015 | 0.263 |
| PNKP | HOM04D005170 | HOM03D004809 | PNKP | 0.444 | 0.177 | 2.509 | 0.015 | 0.263 |
| PMS1 | HOM04D002177 | HOM03D002554 | PMS1 | 0.301 | 0.118 | 2.556 | 0.013 | 0.263 |
| Tag | HOM04D000784 | HOM03D001279 | Tag | -0.326 | 0.135 | -2.407 | 0.019 | 0.268 |
| BRU1 | HOM04D004030 | HOM03D008954 | BRU1 | -0.606 | 0.262 | -2.313 | 0.024 | 0.268 |
| SPO11 | HOM04D001259 | HOM03D001513 | SPO11-1, SPO11-2, SPO11-3 | 0.530 | 0.222 | 2.383 | 0.021 | 0.268 |
| MSH1 | HOM04D004513 | HOM03D005511 | MSH1 | 0.473 | 0.204 | 2.319 | 0.024 | 0.268 |
| *PARG* | HOM04D003287 | HOM03D003504 | *PARG1, PARG2* | 0.530 | 0.238 | 2.230 | 0.030 | 0.286 |
| SOG1 | HOM04D000656 | HOM03D000769 | SOG1 | 0.254 | 0.115 | 2.216 | 0.031 | 0.286 |
| RFC1 | HOM04D004689 | HOM03D002834 | RFC1 | -0.344 | 0.158 | -2.177 | 0.034 | 0.291 |
| MSH5 | HOM04D005333 | HOM03D007428 | MSH5 | 0.524 | 0.250 | 2.099 | 0.040 | 0.326 |
| RAD51D | HOM04D006740 | HOM03D007750 | RAD51D | 0.376 | 0.183 | 2.055 | 0.045 | 0.337 |
| PR19B/PUB60 | HOM04D003246 | HOM03D004531 | PR19B/PUB60-1, PR19B/PUB60-2 | 0.377 | 0.192 | 1.960 | 0.055 | 0.391 |
| RAD9 | HOM04D005486 | HOM03D007064 | RAD9 | -0.612 | 0.358 | -1.710 | 0.093 | 0.594 |
| RAD23 | HOM04D001203 | HOM03D001632 | RAD23A, RAD23B, RAD23C, RAD23D | 0.201 | 0.118 | 1.708 | 0.093 | 0.594 |
| CUL4 | HOM04D000338 | HOM03D000143 | CUL4 | -0.228 | 0.138 | -1.645 | 0.106 | 0.638 |
| FEN1 | HOM04D003408 | HOM03D002630 | FEN1 | -0.697 | 0.434 | -1.606 | 0.114 | 0.656 |
| XRCC2 | HOM04D006906 | HOM03D008620 | XRCC2 | 0.287 | 0.200 | 1.440 | 0.156 | 0.784 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RBX1 | HOM04D001544 | HOM03D001542 | RBX1 | 0.231 | 0.160 | 1.450 | 0.153 | 0.784 |
| CHEK2 | HOM04D000039 | HOM03D000063 | CHEK2 | 0.150 | 0.103 | 1.452 | 0.152 | 0.784 |
| RPA2 | HOM04D002638 | HOM03D003134 | RPA2 | -0.218 | 0.159 | -1.371 | 0.176 | 0.786 |
| GTF2H5 | HOM04D007085 | HOM03D008072 | GTF2H5 | -0.413 | 0.296 | -1.396 | 0.168 | 0.786 |
| XPB/ERCC3 | HOM04D003675 | HOM03D002803 | XPB/ERCC3 | 0.328 | 0.236 | 1.391 | 0.170 | 0.786 |
| DRT111 | HOM04D004921 | HOM03D003999 | DRT111 | -0.273 | 0.206 | -1.324 | 0.191 | 0.786 |
| OGG1 | HOM04D006939 | HOM03D005744 | OGG1 | 0.181 | 0.138 | 1.312 | 0.195 | 0.786 |
| SMC4 | HOM04D003434 | HOM03D002909 | SMC4 | 0.278 | 0.206 | 1.350 | 0.182 | 0.786 |
| APTX | HOM04D004756 | HOM03D002833 | APTX | -0.190 | 0.152 | -1.246 | 0.218 | 0.826 |
| AXR1 | HOM04D003724 | HOM03D003484 | AXR1 | 0.234 | 0.188 | 1.245 | 0.218 | 0.826 |
| MMS2 | HOM04D001492 | HOM03D001161 | MMS2 | 0.171 | 0.142 | 1.205 | 0.233 | 0.855 |
| COP1 | HOM04D000650 | HOM03D000501 | COP1 | -0.107 | 0.133 | -0.803 | 0.425 | 0.872 |
| GTF2H4 | HOM04D005140 | HOM03D003940 | GTF2H4 | -0.129 | 0.148 | -0.868 | 0.389 | 0.872 |
| POLD4 | HOM04D004732 | HOM03D003548 | POLD4 | -0.185 | 0.233 | -0.793 | 0.431 | 0.872 |
| MUTY | HOM04D005552 | HOM03D004454 | MUTY | -0.205 | 0.219 | -0.936 | 0.353 | 0.872 |
| CSA | HOM04D005364 | HOM03D005285 | CSA | -0.131 | 0.142 | -0.919 | 0.362 | 0.872 |
| CCNH | HOM04D005036 | HOM03D003364 | CCNH | -0.159 | 0.160 | -0.995 | 0.324 | 0.872 |
| KU70 | HOM04D005046 | HOM03D004691 | KU70 | 0.146 | 0.179 | 0.815 | 0.419 | 0.872 |
| Polk | HOM04D002775 | HOM03D006067 | Polk | -0.276 | 0.281 | -0.981 | 0.331 | 0.872 |
| APE2 | HOM04D004425 | HOM03D006661 | APE2 | -0.169 | 0.182 | -0.924 | 0.359 | 0.872 |
| RNR1 | HOM04D002376 | HOM03D001347 | RNR1 | 0.132 | 0.153 | 0.862 | 0.393 | 0.872 |
| MMS19 | HOM04D004480 | HOM03D004191 | MMS19 | -0.176 | 0.197 | -0.895 | 0.375 | 0.872 |
| SMC3 | HOM04D003467 | HOM03D002271 | SMC3 | -0.157 | 0.168 | -0.932 | 0.355 | 0.872 |
| SMC6, MIM | HOM04D003618 | HOM03D003447 | SMC6, MIM | -0.294 | 0.259 | -1.133 | 0.262 | 0.872 |
| XPD/UVH6/ERCC2 | HOM04D004614 | HOM03D005289 | XPD/UVH6/ERCC2 | 0.251 | 0.317 | 0.790 | 0.433 | 0.872 |
| POLD3 | HOM04D002072 | HOM03D004484 | POLD3 | 0.192 | 0.231 | 0.829 | 0.411 | 0.872 |
| GTF2H3 | HOM04D006212 | HOM03D006663 | GTF2H3 | 0.123 | 0.118 | 1.038 | 0.304 | 0.872 |
| RAD1/UVH1/ERCC4/XPF | HOM04D005466 | HOM03D003505 | RAD1/UVH1/ERCC4/XPF | -0.140 | 0.149 | -0.940 | 0.351 | 0.872 |
| MND1 | HOM04D005684 | HOM03D007966 | MND1 | 0.160 | 0.179 | 0.894 | 0.375 | 0.872 |
| KU80 | HOM04D005174 | HOM03D002193 | KU80 | -0.153 | 0.193 | -0.789 | 0.433 | 0.872 |
| PRKDC | HOM04D002601 | HOM03D001652 | PRKDC | -0.265 | 0.339 | -0.780 | 0.438 | 0.872 |
| GTF2H2 | HOM04D006174 | HOM03D006192 | GTF2H2 | 0.169 | 0.189 | 0.897 | 0.374 | 0.872 |
| Muts_like | HOM04D001403 | HOM03D001852 | Muts_like | 0.114 | 0.144 | 0.795 | 0.430 | 0.872 |
| RAD51B | HOM04D007144 | HOM03D007435 | RAD51B | 0.202 | 0.240 | 0.841 | 0.404 | 0.872 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DML | HOM04D001046 | HOM03D001428 | DML1, DML2, DML3 | 0.127 | 0.163 | 0.778 | 0.440 | 0.872 |
| UNG | HOM04D003441 | HOM03D003393 | UNG | 0.348 | 0.315 | 1.106 | 0.273 | 0.872 |
| UVR1/UVX3/XPG/ERCC5 | HOM04D005866 | HOM03D002893 | UVR1/UVX3/XPG/ERCC5 | -0.144 | 0.153 | -0.940 | 0.351 | 0.872 |
| POLH | HOM04D004091 | HOM03D007442 | POLH | 0.250 | 0.273 | 0.914 | 0.365 | 0.872 |
| PRD1 | HOM04D006666 | HOM03D007084 | PRD1 | 0.129 | 0.157 | 0.822 | 0.414 | 0.872 |
| RFC5 | HOM04D001694 | HOM03D003877 | RFC5 | -0.137 | 0.211 | -0.650 | 0.519 | 0.880 |
| DRT101 | HOM04D004359 | HOM03D004660 | DRT101 | -0.125 | 0.200 | -0.623 | 0.536 | 0.880 |
| XRCC1 | HOM04D006984 | HOM03D003667 | XRCC1 | 0.103 | 0.167 | 0.618 | 0.539 | 0.880 |
| EME1 | HOM04D005249 | HOM03D007551 | EME1 | 0.117 | 0.182 | 0.640 | 0.525 | 0.880 |
| SSB | HOM04D002728 | HOM03D002499 | SSB | 0.086 | 0.125 | 0.688 | 0.494 | 0.880 |
| POLD2 | HOM04D005157 | HOM03D004054 | POLD2 | -0.100 | 0.158 | -0.635 | 0.528 | 0.880 |
| MNAT1 | HOM04D005360 | HOM03D004449 | MNAT1 | 0.061 | 0.102 | 0.598 | 0.553 | 0.880 |
| REV7 | HOM04D006848 | HOM03D005648 | REV7 | -0.133 | 0.221 | -0.605 | 0.547 | 0.880 |
| XAB2 | HOM04D003069 | HOM03D002694 | XAB2 | -0.122 | 0.203 | -0.601 | 0.551 | 0.880 |
| MRE11A | HOM04D004854 | HOM03D005935 | MRE11A | 0.087 | 0.139 | 0.627 | 0.533 | 0.880 |
| HUS1 | HOM04D004876 | HOM03D005957 | HUS1 | -0.315 | 0.513 | -0.615 | 0.541 | 0.880 |
| RAD1 | HOM04D006209 | HOM03D007251 | RAD1 | 0.110 | 0.173 | 0.635 | 0.528 | 0.880 |
| HEX1/EXO1 | HOM04D002577 | HOM03D005538 | HEX1/EXO1 | 0.134 | 0.207 | 0.648 | 0.519 | 0.880 |
| POLL | HOM04D006123 | HOM03D007561 | POLL | 0.131 | 0.209 | 0.625 | 0.535 | 0.880 |
| POLE | HOM04D004989 | HOM03D007043 | POLE | 0.115 | 0.189 | 0.608 | 0.545 | 0.880 |
| REX1 | HOM04D006322 | HOM03D006889 | REX1 | 0.088 | 0.153 | 0.575 | 0.567 | 0.881 |
| DET1 | HOM04D005851 | HOM03D003960 | DET1 | 0.092 | 0.163 | 0.564 | 0.575 | 0.881 |
| UvrD | HOM04D002964 | HOM03D005360 | UvrD | -0.164 | 0.289 | -0.569 | 0.572 | 0.881 |
| FLJ35220 | HOM04D005935 | HOM03D006237 | FLJ35220 | 0.074 | 0.136 | 0.546 | 0.587 | 0.885 |
| MUS81 | HOM04D004990 | HOM03D004705 | MUS81 | -0.097 | 0.213 | -0.455 | 0.651 | 0.885 |
| FPG | HOM04D005473 | HOM03D004609 | FPG | -0.057 | 0.126 | -0.456 | 0.650 | 0.885 |
| UVR7/ERCC1 | HOM04D005591 | HOM03D004203 | UVR7/ERCC1 | -0.112 | 0.247 | -0.455 | 0.651 | 0.885 |
| MBD4 | HOM04D004958 | HOM03D003502 | MBD4 | 0.080 | 0.153 | 0.524 | 0.603 | 0.885 |
| LIG1 | HOM04D001683 | HOM03D001412 | LIG1 | 0.068 | 0.146 | 0.467 | 0.643 | 0.885 |
| ATRAD21 | HOM04D001275 | HOM03D001079 | ATRAD21.1, ATRAD21.2, ATRAD21.3 | -0.097 | 0.188 | -0.516 | 0.608 | 0.885 |
| POLE | HOM04D003276 | HOM03D002351 | POLE | 0.089 | 0.194 | 0.457 | 0.650 | 0.885 |
| DDB1 | HOM04D003108 | HOM03D000591 | DDB1 | -0.122 | 0.246 | -0.498 | 0.621 | 0.885 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| GTF2H1 | HOM04D004318 | HOM03D003099 | GTF2H1 | -0.095 | 0.177 | -0.536 | 0.594 | 0.885 |
| SMC1 | HOM04D003489 | HOM03D003237 | SMC1 | 0.088 | 0.202 | 0.433 | 0.667 | 0.886 |
| NBS1 | HOM04D006113 | HOM03D004683 | NBS1 | 0.146 | 0.345 | 0.424 | 0.673 | 0.886 |
| SSRP1 | HOM04D003180 | HOM03D002008 | SSRP1 | 0.064 | 0.151 | 0.425 | 0.672 | 0.886 |
| DUT1 | HOM04D003033 | HOM03D002613 | DUT1 | 0.149 | 0.371 | 0.401 | 0.690 | 0.898 |
| WRN | HOM04D006594 | HOM03D006683 | WRN | 0.064 | 0.189 | 0.338 | 0.737 | 0.949 |
| XPC | HOM04D005966 | HOM03D004314 | XPC | -0.052 | 0.159 | -0.325 | 0.746 | 0.950 |
| Mfd | HOM04D005818 | HOM03D003238 | Mfd | -0.053 | 0.184 | -0.288 | 0.774 | 0.976 |
| MLH3 | HOM04D003331 | HOM03D005080 | MLH3 | 0.149 | 0.536 | 0.278 | 0.782 | 0.976 |
| AlkB | HOM04D006501 | HOM03D006029 | AlkB | -0.012 | 0.144 | -0.084 | 0.934 | 0.992 |
| MLH1 | HOM04D005281 | HOM03D005583 | MLH1 | -0.038 | 0.156 | -0.242 | 0.810 | 0.992 |
| TDP1 | HOM04D005673 | HOM03D004707 | TDP1 | 0.015 | 0.149 | 0.098 | 0.922 | 0.992 |
| ABH3/AlkB | HOM04D007234 | HOM03D007275 | ABH3/AlkB | 0.017 | 0.176 | 0.094 | 0.926 | 0.992 |
| TOP3 | HOM04D002223 | HOM03D002059 | TOP3 | -0.005 | 0.173 | -0.031 | 0.976 | 0.992 |
| APE1L | HOM04D006817 | HOM03D005832 | APE1L | 0.017 | 0.141 | 0.122 | 0.903 | 0.992 |
| RAD50 | HOM04D005302 | HOM03D003113 | RAD50 | -0.002 | 0.149 | -0.015 | 0.988 | 0.992 |
| RFC2 | HOM04D001345 | HOM03D001196 | RFC2, RFC3, RFC4 | -0.015 | 0.091 | -0.168 | 0.867 | 0.992 |
| DDB2 | HOM04D007014 | HOM03D003898 | DDB2 | -0.001 | 0.134 | -0.010 | 0.992 | 0.992 |
| XRCC4 | HOM04D006340 | HOM03D005209 | XRCC4 | 0.024 | 0.242 | 0.097 | 0.923 | 0.992 |
| SMC5 | HOM04D004387 | HOM03D001853 | SMC5 | 0.023 | 0.230 | 0.100 | 0.921 | 0.992 |
| RAD17 | HOM04D005902 | HOM03D006532 | RAD17 | -0.057 | 0.292 | -0.195 | 0.846 | 0.992 |
| RPA3 | HOM04D003942 | HOM03D005396 | RPA3 | 0.004 | 0.212 | 0.017 | 0.986 | 0.992 |
| NUDX1 | HOM04D003418 | HOM03D005023 | NUDX1 | -0.057 | 0.332 | -0.172 | 0.864 | 0.992 |
| PHR1 | HOM04D005911 | HOM03D005566 | PHR1 | 0.005 | 0.121 | 0.038 | 0.969 | 0.992 |
| RecG | HOM04D003779 | HOM03D003370 | RecG | 0.008 | 0.175 | 0.043 | 0.966 | 0.992 |
| REV1 | HOM04D004212 | HOM03D005524 | REV1 | 0.028 | 0.274 | 0.102 | 0.919 | 0.992 |
| LIG4 | HOM04D005047 | HOM03D002488 | LIG4 | 0.009 | 0.164 | 0.057 | 0.955 | 0.992 |
| RPA1 | HOM04D000929 | HOM03D000629 | RPA1 | 0.037 | 0.188 | 0.197 | 0.845 | 0.992 |
| NTH | HOM04D004019 | HOM03D005173 | NTH | 0.030 | 0.204 | 0.148 | 0.883 | 0.992 |
| MAGLP/AlkA | HOM04D002929 | HOM03D004685 | MAGLP/AlkA | 0.034 | 0.179 | 0.189 | 0.851 | 0.992 |
| APE1 | HOM04D004383 | HOM03D004400 | APE1 | 0.042 | 0.212 | 0.199 | 0.843 | 0.992 |
| HMGB1 | HOM04D000711 | HOM03D000500 | HMGB1 | 0.008 | 0.126 | 0.066 | 0.948 | 0.992 |
| RAD51C | HOM04D007012 | HOM03D007195 | RAD51C | 0.032 | 0.174 | 0.183 | 0.855 | 0.992 |

(B) Trees versus perennial herbs

| Symbol of gene family | ID of gene family in Dicots PLAZA 4.0 | ID of gene family in Gymno PLAZA 1.0 | Genes within the gene family | Coefficient | Standard error | t-value | p-value | Q-value |
|---|---|---|---|---|---|---|---|---|
| PARP | HOM04D001195 | HOM03D000597 | PARP1, PARP2, PARP3 | -0.395 | 0.111 | -3.560 | 0.001 | 0.046 |
| RNR2, TSO2 | HOM04D002018 | HOM03D001558 | RNR2a, TSO2 | -0.554 | 0.202 | -2.743 | 0.008 | 0.194 |
| BRCA2 | HOM04D004670 | HOM03D008142 | BRCA2 | 0.262 | 0.253 | 1.034 | 0.305 | 0.689 |
| DRT102 | HOM04D006441 | HOM03D003323 | DRT102 | -0.566 | 0.272 | -2.076 | 0.042 | 0.361 |
| MPG/MAG | HOM04D007180 | HOM03D007182 | MPG/MAG | 0.039 | 0.157 | 0.250 | 0.803 | 0.936 |
| PNKP | HOM04D005170 | HOM03D004809 | PNKP | 0.030 | 0.187 | 0.158 | 0.875 | 0.968 |
| PMS1 | HOM04D002177 | HOM03D002554 | PMS1 | -0.012 | 0.144 | -0.085 | 0.933 | 0.968 |
| Tag | HOM04D000784 | HOM03D001279 | Tag | -0.646 | 0.136 | -4.749 | 0.000 | 0.002 |
| BRU1 | HOM04D004030 | HOM03D008954 | BRU1 | -0.764 | 0.295 | -2.590 | 0.012 | 0.208 |
| SPO11 | HOM04D001259 | HOM03D001513 | SPO11-1, SPO11-2, SPO11-3 | 0.511 | 0.223 | 2.288 | 0.026 | 0.280 |
| MSH1 | HOM04D004513 | HOM03D005511 | MSH1 | 0.151 | 0.250 | 0.604 | 0.548 | 0.814 |
| *PARG* | HOM04D003287 | HOM03D003504 | *PARG1, PARG2* | 0.178 | 0.246 | 0.725 | 0.472 | 0.732 |
| SOG1 | HOM04D000656 | HOM03D000769 | SOG1 | -0.065 | 0.122 | -0.531 | 0.597 | 0.839 |
| RFC1 | HOM04D004689 | HOM03D002834 | RFC1 | -0.308 | 0.194 | -1.587 | 0.118 | 0.540 |
| MSH5 | HOM04D005333 | HOM03D007428 | MSH5 | 0.615 | 0.306 | 2.007 | 0.050 | 0.376 |
| RAD51D | HOM04D006740 | HOM03D007750 | RAD51D | 0.205 | 0.225 | 0.913 | 0.365 | 0.689 |
| PR19B/PUB60 | HOM04D003246 | HOM03D004531 | PR19B/PUB60-1, PR19B/PUB60-2 | -0.242 | 0.199 | -1.219 | 0.228 | 0.662 |
| RAD9 | HOM04D005486 | HOM03D007064 | RAD9 | -0.452 | 0.360 | -1.256 | 0.214 | 0.662 |
| RAD23 | HOM04D001203 | HOM03D001632 | RAD23A, RAD23B, RAD23C, RAD23D | -0.110 | 0.119 | -0.918 | 0.363 | 0.689 |
| CUL4 | HOM04D000338 | HOM03D000143 | CUL4 | 0.005 | 0.170 | 0.027 | 0.979 | 0.968 |
| FEN1 | HOM04D003408 | HOM03D002630 | FEN1 | -0.612 | 0.435 | -1.406 | 0.165 | 0.634 |
| XRCC2 | HOM04D006906 | HOM03D008620 | XRCC2 | 0.560 | 0.245 | 2.288 | 0.026 | 0.280 |
| RBX1 | HOM04D001544 | HOM03D001542 | RBX1 | -0.134 | 0.169 | -0.793 | 0.431 | 0.715 |
| CHEK2 | HOM04D000039 | HOM03D000063 | CHEK2 | 0.010 | 0.105 | 0.097 | 0.923 | 0.968 |
| RPA2 | HOM04D002638 | HOM03D003134 | RPA2 | -0.530 | 0.160 | -3.317 | 0.002 | 0.064 |
| GTF2H5 | HOM04D007085 | HOM03D008072 | GTF2H5 | -0.413 | 0.321 | -1.287 | 0.203 | 0.662 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| XPB/ERCC3 | HOM04D003675 | HOM03D002803 | XPB/ERCC3 | 0.288 | 0.246 | 1.168 | 0.248 | 0.662 |
| DRT111 | HOM04D004921 | HOM03D003999 | DRT111 | -0.309 | 0.252 | -1.222 | 0.227 | 0.662 |
| OGG1 | HOM04D006939 | HOM03D005744 | OGG1 | 0.158 | 0.161 | 0.983 | 0.330 | 0.689 |
| SMC4 | HOM04D003434 | HOM03D002909 | SMC4 | 0.234 | 0.253 | 0.925 | 0.359 | 0.689 |
| APTX | HOM04D004756 | HOM03D002833 | APTX | -0.435 | 0.187 | -2.327 | 0.024 | 0.280 |
| AXR1 | HOM04D003724 | HOM03D003484 | AXR1 | 0.260 | 0.204 | 1.275 | 0.208 | 0.662 |
| MMS2 | HOM04D001492 | HOM03D001161 | MMS2 | -0.010 | 0.151 | -0.066 | 0.947 | 0.968 |
| COP1 | HOM04D000650 | HOM03D000501 | COP1 | -0.368 | 0.134 | -2.750 | 0.008 | 0.194 |
| GTF2H4 | HOM04D005140 | HOM03D003940 | GTF2H4 | -0.361 | 0.182 | -1.982 | 0.052 | 0.376 |
| POLD4 | HOM04D004732 | HOM03D003548 | POLD4 | -0.418 | 0.236 | -1.771 | 0.082 | 0.443 |
| MUTY | HOM04D005552 | HOM03D004454 | MUTY | -0.337 | 0.221 | -1.526 | 0.133 | 0.584 |
| CSA | HOM04D005364 | HOM03D005285 | CSA | -0.209 | 0.175 | -1.199 | 0.236 | 0.662 |
| CCNH | HOM04D005036 | HOM03D003364 | CCNH | -0.216 | 0.196 | -1.103 | 0.275 | 0.662 |
| KU70 | HOM04D005046 | HOM03D004691 | KU70 | -0.282 | 0.219 | -1.286 | 0.204 | 0.662 |
| Polk | HOM04D002775 | HOM03D006067 | Polk | -0.321 | 0.284 | -1.130 | 0.263 | 0.662 |
| APE2 | HOM04D004425 | HOM03D006661 | APE2 | -0.230 | 0.224 | -1.027 | 0.309 | 0.689 |
| RNR1 | HOM04D002376 | HOM03D001347 | RNR1 | 0.148 | 0.159 | 0.931 | 0.356 | 0.689 |
| MMS19 | HOM04D004480 | HOM03D004191 | MMS19 | -0.201 | 0.242 | -0.832 | 0.409 | 0.703 |
| SMC3 | HOM04D003467 | HOM03D002271 | SMC3 | -0.164 | 0.192 | -0.852 | 0.398 | 0.703 |
| SMC6, MIM | HOM04D003618 | HOM03D003447 | SMC6, MIM | -0.215 | 0.260 | -0.824 | 0.414 | 0.703 |
| XPD/UVH6/ERCC2 | HOM04D004614 | HOM03D005289 | XPD/UVH6/ERCC2 | 0.230 | 0.319 | 0.721 | 0.474 | 0.732 |
| POLD3 | HOM04D002072 | HOM03D004484 | POLD3 | -0.182 | 0.246 | -0.742 | 0.461 | 0.732 |
| GTF2H3 | HOM04D006212 | HOM03D006663 | GTF2H3 | 0.088 | 0.145 | 0.606 | 0.547 | 0.814 |
| RAD1/UVH1/ERCC4/XPF | HOM04D005466 | HOM03D003505 | RAD1/UVH1/ERCC4/XPF | -0.097 | 0.183 | -0.533 | 0.596 | 0.839 |
| MND1 | HOM04D005684 | HOM03D007966 | MND1 | 0.112 | 0.220 | 0.511 | 0.612 | 0.839 |
| KU80 | HOM04D005174 | HOM03D002193 | KU80 | -0.112 | 0.225 | -0.498 | 0.621 | 0.839 |
| PRKDC | HOM04D002601 | HOM03D001652 | PRKDC | 0.184 | 0.366 | 0.502 | 0.618 | 0.839 |
| GTF2H2 | HOM04D006174 | HOM03D006192 | GTF2H2 | -0.109 | 0.231 | -0.472 | 0.639 | 0.853 |
| Muts_like | HOM04D001403 | HOM03D001852 | Muts_like | 0.058 | 0.151 | 0.383 | 0.703 | 0.871 |
| RAD51B | HOM04D007144 | HOM03D007435 | RAD51B | 0.080 | 0.256 | 0.314 | 0.755 | 0.898 |
| DML | HOM04D001046 | HOM03D001428 | DML1, DML2, DML3 | -0.022 | 0.182 | -0.118 | 0.906 | 0.968 |
| UNG | HOM04D003441 | HOM03D003393 | UNG | -0.046 | 0.315 | -0.147 | 0.884 | 0.968 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| UVR1/UVX3/XPG/ERCC5 | HOM04D005866 | HOM03D002893 | UVR1/UVX3/XPG/ERCC5 | 0.008 | 0.169 | 0.049 | 0.961 | 0.968 |
| POLH | HOM04D004091 | HOM03D007442 | POLH | 0.017 | 0.335 | 0.052 | 0.959 | 0.968 |
| PRD1 | HOM04D006666 | HOM03D007084 | PRD1 | -0.030 | 0.178 | -0.167 | 0.868 | 0.968 |
| RFC5 | HOM04D001694 | HOM03D003877 | RFC5 | -0.516 | 0.215 | -2.405 | 0.020 | 0.280 |
| DRT101 | HOM04D004359 | HOM03D004660 | DRT101 | -0.425 | 0.223 | -1.905 | 0.062 | 0.376 |
| XRCC1 | HOM04D006984 | HOM03D003667 | XRCC1 | -0.248 | 0.204 | -1.212 | 0.230 | 0.662 |
| EME1 | HOM04D005249 | HOM03D007551 | EME1 | 0.212 | 0.191 | 1.108 | 0.273 | 0.662 |
| SSB | HOM04D002728 | HOM03D002499 | SSB | -0.167 | 0.154 | -1.087 | 0.282 | 0.662 |
| POLD2 | HOM04D005157 | HOM03D004054 | POLD2 | -0.213 | 0.193 | -1.100 | 0.276 | 0.662 |
| MNAT1 | HOM04D005360 | HOM03D004449 | MNAT1 | -0.116 | 0.125 | -0.932 | 0.355 | 0.689 |
| REV7 | HOM04D006848 | HOM03D005648 | REV7 | -0.209 | 0.229 | -0.916 | 0.363 | 0.689 |
| XAB2 | HOM04D003069 | HOM03D002694 | XAB2 | 0.207 | 0.249 | 0.830 | 0.410 | 0.703 |
| MRE11A | HOM04D004854 | HOM03D005935 | MRE11A | -0.145 | 0.171 | -0.850 | 0.399 | 0.703 |
| HUS1 | HOM04D004876 | HOM03D005957 | HUS1 | -0.382 | 0.516 | -0.740 | 0.463 | 0.732 |
| RAD1 | HOM04D006209 | HOM03D007251 | RAD1 | 0.126 | 0.212 | 0.595 | 0.554 | 0.814 |
| HEX1/EXO1 | HOM04D002577 | HOM03D005538 | HEX1/EXO1 | -0.068 | 0.208 | -0.327 | 0.745 | 0.898 |
| POLL | HOM04D006123 | HOM03D007561 | POLL | 0.005 | 0.256 | 0.019 | 0.985 | 0.968 |
| POLE | HOM04D004989 | HOM03D007043 | POLE | 0.033 | 0.232 | 0.142 | 0.888 | 0.968 |
| REX1 | HOM04D006322 | HOM03D006889 | REX1 | -0.267 | 0.187 | -1.428 | 0.159 | 0.630 |
| DET1 | HOM04D005851 | HOM03D003960 | DET1 | -0.209 | 0.170 | -1.233 | 0.223 | 0.662 |
| UvrD | HOM04D002964 | HOM03D005360 | UvrD | -0.155 | 0.354 | -0.438 | 0.663 | 0.857 |
| FLJ35220 | HOM04D005935 | HOM03D006237 | FLJ35220 | 0.368 | 0.167 | 2.205 | 0.032 | 0.313 |
| MUS81 | HOM04D004990 | HOM03D004705 | MUS81 | -0.504 | 0.261 | -1.930 | 0.059 | 0.376 |
| FPG | HOM04D005473 | HOM03D004609 | FPG | -0.134 | 0.154 | -0.867 | 0.389 | 0.703 |
| UVR7/ERCC1 | HOM04D005591 | HOM03D004203 | UVR7/ERCC1 | -0.220 | 0.248 | -0.887 | 0.379 | 0.703 |
| MBD4 | HOM04D004958 | HOM03D003502 | MBD4 | 0.148 | 0.187 | 0.790 | 0.433 | 0.715 |
| LIG1 | HOM04D001683 | HOM03D001412 | LIG1 | -0.131 | 0.180 | -0.731 | 0.468 | 0.732 |
| ATRAD21 | HOM04D001275 | HOM03D001079 | ATRAD21.1, ATRAD21.2, ATRAD21.3 | -0.099 | 0.189 | -0.525 | 0.602 | 0.839 |
| POLE | HOM04D003276 | HOM03D002351 | POLE | -0.094 | 0.224 | -0.419 | 0.677 | 0.862 |
| DDB1 | HOM04D003108 | HOM03D000591 | DDB1 | -0.116 | 0.290 | -0.401 | 0.690 | 0.864 |
| GTF2H1 | HOM04D004318 | HOM03D003099 | GTF2H1 | 0.073 | 0.217 | 0.335 | 0.739 | 0.898 |
| SMC1 | HOM04D003489 | HOM03D003237 | SMC1 | -0.168 | 0.248 | -0.676 | 0.502 | 0.765 |
| NBS1 | HOM04D006113 | HOM03D004683 | NBS1 | 0.051 | 0.347 | 0.148 | 0.883 | 0.968 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SSRP1 | HOM04D003180 | HOM03D002008 | SSRP1 | -0.003 | 0.165 | -0.020 | 0.984 | 0.968 |
| DUT1 | HOM04D003033 | HOM03D002613 | DUT1 | -0.171 | 0.373 | -0.460 | 0.647 | 0.855 |
| WRN | HOM04D006594 | HOM03D006683 | WRN | -0.381 | 0.232 | -1.641 | 0.106 | 0.528 |
| XPC | HOM04D005966 | HOM03D004314 | XPC | -0.099 | 0.195 | -0.509 | 0.613 | 0.839 |
| Mfd | HOM04D005818 | HOM03D003238 | Mfd | -0.404 | 0.210 | -1.923 | 0.060 | 0.376 |
| MLH3 | HOM04D003331 | HOM03D005080 | MLH3 | 1.246 | 0.658 | 1.895 | 0.063 | 0.376 |
| AlkB | HOM04D006501 | HOM03D006029 | AlkB | -0.473 | 0.177 | -2.676 | 0.010 | 0.194 |
| MLH1 | HOM04D005281 | HOM03D005583 | MLH1 | -0.406 | 0.192 | -2.118 | 0.039 | 0.353 |
| TDP1 | HOM04D005673 | HOM03D004707 | TDP1 | -0.328 | 0.183 | -1.794 | 0.078 | 0.443 |
| ABH3/AlkB | HOM04D007234 | HOM03D007275 | ABH3/AlkB | -0.376 | 0.216 | -1.737 | 0.088 | 0.455 |
| TOP3 | HOM04D002223 | HOM03D002059 | TOP3 | -0.288 | 0.180 | -1.598 | 0.116 | 0.540 |
| APE1L | HOM04D006817 | HOM03D005832 | APE1L | -0.215 | 0.150 | -1.431 | 0.158 | 0.630 |
| RAD50 | HOM04D005302 | HOM03D003113 | RAD50 | -0.264 | 0.182 | -1.450 | 0.153 | 0.630 |
| RFC2 | HOM04D001345 | HOM03D001196 | RFC2, RFC3, RFC4 | -0.121 | 0.112 | -1.082 | 0.284 | 0.662 |
| DDB2 | HOM04D007014 | HOM03D003898 | DDB2 | -0.186 | 0.164 | -1.134 | 0.262 | 0.662 |
| XRCC4 | HOM04D006340 | HOM03D005209 | XRCC4 | -0.283 | 0.243 | -1.168 | 0.248 | 0.662 |
| SMC5 | HOM04D004387 | HOM03D001853 | SMC5 | -0.278 | 0.244 | -1.139 | 0.259 | 0.662 |
| RAD17 | HOM04D005902 | HOM03D006532 | RAD17 | -0.416 | 0.358 | -1.163 | 0.250 | 0.662 |
| RPA3 | HOM04D003942 | HOM03D005396 | RPA3 | -0.203 | 0.215 | -0.942 | 0.350 | 0.689 |
| NUDX1 | HOM04D003418 | HOM03D005023 | NUDX1 | -0.333 | 0.334 | -0.995 | 0.324 | 0.689 |
| PHR1 | HOM04D005911 | HOM03D005566 | PHR1 | -0.136 | 0.149 | -0.914 | 0.365 | 0.689 |
| RecG | HOM04D003779 | HOM03D003370 | RecG | -0.095 | 0.214 | -0.443 | 0.660 | 0.857 |
| REV1 | HOM04D004212 | HOM03D005524 | REV1 | 0.116 | 0.281 | 0.412 | 0.682 | 0.862 |
| LIG4 | HOM04D005047 | HOM03D002488 | LIG4 | 0.062 | 0.193 | 0.319 | 0.751 | 0.898 |
| RPA1 | HOM04D000929 | HOM03D000629 | RPA1 | 0.056 | 0.192 | 0.291 | 0.772 | 0.909 |
| NTH | HOM04D004019 | HOM03D005173 | NTH | -0.017 | 0.228 | -0.073 | 0.942 | 0.968 |
| MAGLP/AlkA | HOM04D002929 | HOM03D004685 | MAGLP/AlkA | 0.017 | 0.189 | 0.089 | 0.929 | 0.968 |
| APE1 | HOM04D004383 | HOM03D004400 | APE1 | 0.040 | 0.224 | 0.178 | 0.859 | 0.968 |
| HMGB1 | HOM04D000711 | HOM03D000500 | HMGB1 | 0.011 | 0.135 | 0.080 | 0.936 | 0.968 |
| RAD51C | HOM04D007012 | HOM03D007195 | RAD51C | -0.005 | 0.214 | -0.025 | 0.980 | 0.968 |

(C) Estimated values of Pagel's lambda, Log likelihood and AIC

| Symbol of gene family | ID of gene family in Dicots PLAZA 4.0 | ID of gene family in Gymno PLAZA 1.0 | Genes within the gene family | Pagel's lambda | Log likelihood | AIC |
|---|---|---|---|---|---|---|
| PARP | HOM04D001195 | HOM03D000597 | PARP1, PARP2, PARP3 | 0.000 | -17.739 | 45.479 |
| RNR2, TSO2 | HOM04D002018 | HOM03D001558 | RNR2a, TSO2 | 0.898 | -47.740 | 105.480 |
| BRCA2 | HOM04D004670 | HOM03D008142 | BRCA2 | 0.000 | -66.428 | 142.856 |
| DRT102 | HOM04D006441 | HOM03D003323 | DRT102 | 0.424 | -68.013 | 146.026 |
| MPG/MAG | HOM04D007180 | HOM03D007182 | MPG/MAG | 0.000 | -38.431 | 86.861 |
| PNKP | HOM04D005170 | HOM03D004809 | PNKP | 0.478 | -45.649 | 101.299 |
| PMS1 | HOM04D002177 | HOM03D002554 | PMS1 | 0.000 | -33.323 | 76.646 |
| Tag | HOM04D000784 | HOM03D001279 | Tag | 0.982 | -21.904 | 53.808 |
| BRU1 | HOM04D004030 | HOM03D008954 | BRU1 | 0.177 | -74.024 | 158.048 |
| SPO11 | HOM04D001259 | HOM03D001513 | SPO11-1, SPO11-2, SPO11-3 | 0.960 | -52.296 | 114.592 |
| MSH1 | HOM04D004513 | HOM03D005511 | MSH1 | 0.000 | -65.718 | 141.437 |
| *PARG* | HOM04D003287 | HOM03D003504 | *PARG1, PARG2* | 0.620 | -61.158 | 132.317 |
| SOG1 | HOM04D000656 | HOM03D000769 | SOG1 | 0.433 | -20.668 | 51.335 |
| RFC1 | HOM04D004689 | HOM03D002834 | RFC1 | 0.000 | -50.759 | 111.519 |
| MSH5 | HOM04D005333 | HOM03D007428 | MSH5 | 0.000 | -77.709 | 165.417 |
| RAD51D | HOM04D006740 | HOM03D007750 | RAD51D | 0.000 | -59.367 | 128.735 |
| PR19B/PUB60 | HOM04D003246 | HOM03D004531 | PR19B/PUB60-1, PR19B/PUB60-2 | 0.631 | -48.608 | 107.216 |
| RAD9 | HOM04D005486 | HOM03D007064 | RAD9 | 0.976 | -79.638 | 169.275 |
| RAD23 | HOM04D001203 | HOM03D001632 | RAD23A, RAD23B, RAD23C, RAD23D | 0.801 | -17.612 | 45.223 |
| CUL4 | HOM04D000338 | HOM03D000143 | CUL4 | 0.000 | -42.867 | 95.734 |
| FEN1 | HOM04D003408 | HOM03D002630 | FEN1 | 1.000 | -88.081 | 186.161 |
| XRCC2 | HOM04D006906 | HOM03D008620 | XRCC2 | 0.000 | -64.489 | 138.979 |
| RBX1 | HOM04D001544 | HOM03D001542 | RBX1 | 0.477 | -39.570 | 89.139 |
| CHEK2 | HOM04D000039 | HOM03D000063 | CHEK2 | 0.779 | -10.166 | 30.331 |
| RPA2 | HOM04D002638 | HOM03D003134 | RPA2 | 0.981 | -31.506 | 73.012 |
| GTF2H5 | HOM04D007085 | HOM03D008072 | GTF2H5 | 0.344 | -78.000 | 166.000 |
| XPB/ERCC3 | HOM04D003675 | HOM03D002803 | XPB/ERCC3 | 0.556 | -61.573 | 133.146 |
| DRT111 | HOM04D004921 | HOM03D003999 | DRT111 | 0.000 | -66.281 | 142.563 |

| | | | | | | |
|---|---|---|---|---|---|---|
| OGG1 | HOM04D006939 | HOM03D005744 | OGG1 | 0.073 | -39.057 | 88.113 |
| SMC4 | HOM04D003434 | HOM03D002909 | SMC4 | 0.000 | -66.339 | 142.677 |
| APTX | HOM04D004756 | HOM03D002833 | APTX | 0.000 | -48.579 | 107.158 |
| AXR1 | HOM04D003724 | HOM03D003484 | AXR1 | 0.339 | -51.385 | 112.770 |
| MMS2 | HOM04D001492 | HOM03D001161 | MMS2 | 0.444 | -33.201 | 76.403 |
| COP1 | HOM04D000650 | HOM03D000501 | COP1 | 0.837 | -24.068 | 58.137 |
| GTF2H4 | HOM04D005140 | HOM03D003940 | GTF2H4 | 0.000 | -46.953 | 103.906 |
| POLD4 | HOM04D004732 | HOM03D003548 | POLD4 | 0.798 | -57.807 | 125.614 |
| MUTY | HOM04D005552 | HOM03D004454 | MUTY | 0.851 | -53.541 | 117.081 |
| CSA | HOM04D005364 | HOM03D005285 | CSA | 0.000 | -44.532 | 99.065 |
| CCNH | HOM04D005036 | HOM03D003364 | CCNH | 0.000 | -51.277 | 112.554 |
| KU70 | HOM04D005046 | HOM03D004691 | KU70 | 0.000 | -57.992 | 125.985 |
| Polk | HOM04D002775 | HOM03D006067 | Polk | 0.844 | -68.378 | 146.756 |
| APE2 | HOM04D004425 | HOM03D006661 | APE2 | 0.000 | -59.142 | 128.284 |
| RNR1 | HOM04D002376 | HOM03D001347 | RNR1 | 0.587 | -35.661 | 81.322 |
| MMS19 | HOM04D004480 | HOM03D004191 | MMS19 | 0.000 | -63.682 | 137.363 |
| SMC3 | HOM04D003467 | HOM03D002271 | SMC3 | 0.135 | -49.015 | 108.029 |
| SMC6, MIM | HOM04D003618 | HOM03D003447 | SMC6, MIM | 0.926 | -62.273 | 134.546 |
| XPD/UVH6/ERCC2 | HOM04D004614 | HOM03D005289 | XPD/UVH6/ERCC2 | 0.989 | -71.625 | 153.250 |
| POLD3 | HOM04D002072 | HOM03D004484 | POLD3 | 0.455 | -61.769 | 133.538 |
| GTF2H3 | HOM04D006212 | HOM03D006663 | GTF2H3 | 0.000 | -33.634 | 77.269 |
| RAD1/UVH1/ERCC4/ XPF | HOM04D005466 | HOM03D003505 | RAD1/UVH1/ERCC4 /XPF | 0.000 | -47.147 | 104.293 |
| MND1 | HOM04D005684 | HOM03D007966 | MND1 | 0.000 | -58.139 | 126.278 |
| KU80 | HOM04D005174 | HOM03D002193 | KU80 | 0.082 | -58.784 | 127.569 |
| PRKDC | HOM04D002601 | HOM03D001652 | PRKDC | 0.363 | -85.723 | 181.445 |
| GTF2H2 | HOM04D006174 | HOM03D006192 | GTF2H2 | 0.000 | -61.145 | 132.291 |
| Muts_like | HOM04D001403 | HOM03D001852 | Muts_like | 0.521 | -32.759 | 75.517 |
| RAD51B | HOM04D007144 | HOM03D007435 | RAD51B | 0.425 | -64.408 | 138.817 |
| DML | HOM04D001046 | HOM03D001428 | DML1, DML2, DML3 | 0.214 | -45.141 | 100.282 |
| UNG | HOM04D003441 | HOM03D003393 | UNG | 1.000 | -69.099 | 148.198 |
| UVR1/UVX3/XPG/ER CC5 | HOM04D005866 | HOM03D002893 | UVR1/UVX3/XPG/E RCC5 | 0.268 | -40.515 | 91.031 |
| POLH | HOM04D004091 | HOM03D007442 | POLH | 0.000 | -82.940 | 175.881 |
| PRD1 | HOM04D006666 | HOM03D007084 | PRD1 | 0.162 | -44.222 | 98.443 |
| RFC5 | HOM04D001694 | HOM03D003877 | RFC5 | 0.745 | -52.574 | 115.149 |

| | | | | | | |
|---|---|---|---|---|---|---|
| DRT101 | HOM04D004359 | HOM03D004660 | DRT101 | 0.216 | -57.200 | 124.400 |
| XRCC1 | HOM04D006984 | HOM03D003667 | XRCC1 | 0.000 | -53.811 | 117.621 |
| EME1 | HOM04D005249 | HOM03D007551 | EME1 | 0.530 | -46.715 | 103.430 |
| SSB | HOM04D002728 | HOM03D002499 | SSB | 0.000 | -37.018 | 84.036 |
| POLD2 | HOM04D005157 | HOM03D004054 | POLD2 | 0.000 | -50.579 | 111.157 |
| MNAT1 | HOM04D005360 | HOM03D004449 | MNAT1 | 0.000 | -24.630 | 59.260 |
| REV7 | HOM04D006848 | HOM03D005648 | REV7 | 0.609 | -56.922 | 123.845 |
| XAB2 | HOM04D003069 | HOM03D002694 | XAB2 | 0.000 | -65.491 | 140.983 |
| MRE11A | HOM04D004854 | HOM03D005935 | MRE11A | 0.000 | -43.258 | 96.516 |
| HUS1 | HOM04D004876 | HOM03D005957 | HUS1 | 0.977 | -100.915 | 211.831 |
| RAD1 | HOM04D006209 | HOM03D007251 | RAD1 | 0.000 | -55.904 | 121.807 |
| HEX1/EXO1 | HOM04D002577 | HOM03D005538 | HEX1/EXO1 | 0.956 | -48.350 | 106.701 |
| POLL | HOM04D006123 | HOM03D007561 | POLL | 0.000 | -67.196 | 144.391 |
| POLE | HOM04D004989 | HOM03D007043 | POLE | 0.000 | -61.228 | 132.455 |
| REX1 | HOM04D006322 | HOM03D006889 | REX1 | 0.000 | -48.619 | 107.239 |
| DET1 | HOM04D005851 | HOM03D003960 | DET1 | 0.571 | -39.535 | 89.070 |
| UvrD | HOM04D002964 | HOM03D005360 | UvrD | 0.000 | -86.277 | 182.554 |
| FLJ35220 | HOM04D005935 | HOM03D006237 | FLJ35220 | 0.000 | -41.921 | 93.843 |
| MUS81 | HOM04D004990 | HOM03D004705 | MUS81 | 0.000 | -68.329 | 146.658 |
| FPG | HOM04D005473 | HOM03D004609 | FPG | 0.000 | -37.219 | 84.438 |
| UVR7/ERCC1 | HOM04D005591 | HOM03D004203 | UVR7/ERCC1 | 0.990 | -56.662 | 123.324 |
| MBD4 | HOM04D004958 | HOM03D003502 | MBD4 | 0.000 | -48.643 | 107.287 |
| LIG1 | HOM04D001683 | HOM03D001412 | LIG1 | 0.000 | -46.208 | 102.417 |
| ATRAD21 | HOM04D001275 | HOM03D001079 | ATRAD21.1, ATRAD21.2, ATRAD21.3 | 0.922 | -43.381 | 96.762 |
| POLE | HOM04D003276 | HOM03D002351 | POLE | 0.106 | -58.205 | 126.410 |
| DDB1 | HOM04D003108 | HOM03D000591 | DDB1 | 0.056 | -73.957 | 157.913 |
| GTF2H1 | HOM04D004318 | HOM03D003099 | GTF2H1 | 0.000 | -57.265 | 124.529 |
| SMC1 | HOM04D003489 | HOM03D003237 | SMC1 | 0.000 | -65.284 | 140.567 |
| NBS1 | HOM04D006113 | HOM03D004683 | NBS1 | 0.963 | -78.210 | 166.421 |
| SSRP1 | HOM04D003180 | HOM03D002008 | SSRP1 | 0.302 | -38.935 | 87.871 |
| DUT1 | HOM04D003033 | HOM03D002613 | DUT1 | 0.939 | -83.126 | 176.251 |
| WRN | HOM04D006594 | HOM03D006683 | WRN | 0.000 | -61.360 | 132.720 |
| XPC | HOM04D005966 | HOM03D004314 | XPC | 0.000 | -50.961 | 111.922 |
| Mfd | HOM04D005818 | HOM03D003238 | Mfd | 0.134 | -54.263 | 118.526 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MLH3 | HOM04D003331 | HOM03D005080 | MLH3 | 0.000 | -122.777 | 255.554 |
| AlkB | HOM04D006501 | HOM03D006029 | AlkB | 0.000 | -45.282 | 100.564 |
| MLH1 | HOM04D005281 | HOM03D005583 | MLH1 | 0.000 | -50.089 | 110.178 |
| TDP1 | HOM04D005673 | HOM03D004707 | TDP1 | 0.000 | -47.272 | 104.543 |
| ABH3/AlkB | HOM04D007234 | HOM03D007275 | ABH3/AlkB | 0.000 | -57.186 | 124.373 |
| TOP3 | HOM04D002223 | HOM03D002059 | TOP3 | 0.553 | -43.169 | 96.338 |
| APE1L | HOM04D006817 | HOM03D005832 | APE1L | 0.422 | -32.999 | 75.998 |
| RAD50 | HOM04D005302 | HOM03D003113 | RAD50 | 0.000 | -47.088 | 104.176 |
| RFC2 | HOM04D001345 | HOM03D001196 | RFC2, RFC3, RFC4 | 0.000 | -18.103 | 46.206 |
| DDB2 | HOM04D007014 | HOM03D003898 | DDB2 | 0.000 | -40.874 | 91.748 |
| XRCC4 | HOM04D006340 | HOM03D005209 | XRCC4 | 0.922 | -58.184 | 126.367 |
| SMC5 | HOM04D004387 | HOM03D001853 | SMC5 | 0.447 | -61.472 | 132.945 |
| RAD17 | HOM04D005902 | HOM03D006532 | RAD17 | 0.000 | -86.841 | 183.682 |
| RPA3 | HOM04D003942 | HOM03D005396 | RPA3 | 0.781 | -52.585 | 115.170 |
| NUDX1 | HOM04D003418 | HOM03D005023 | NUDX1 | 0.917 | -77.134 | 164.268 |
| PHR1 | HOM04D005911 | HOM03D005566 | PHR1 | 0.000 | -35.055 | 80.111 |
| RecG | HOM04D003779 | HOM03D003370 | RecG | 0.000 | -56.650 | 123.300 |
| REV1 | HOM04D004212 | HOM03D005524 | REV1 | 0.668 | -68.927 | 147.854 |
| LIG4 | HOM04D005047 | HOM03D002488 | LIG4 | 0.066 | -49.841 | 109.683 |
| RPA1 | HOM04D000929 | HOM03D000629 | RPA1 | 0.739 | -46.005 | 102.009 |
| NTH | HOM04D004019 | HOM03D005173 | NTH | 0.197 | -58.668 | 127.335 |
| MAGLP/AlkA | HOM04D002929 | HOM03D004685 | MAGLP/AlkA | 0.496 | -46.105 | 102.210 |
| APE1 | HOM04D004383 | HOM03D004400 | APE1 | 0.482 | -56.276 | 122.551 |
| HMGB1 | HOM04D000711 | HOM03D000500 | HMGB1 | 0.380 | -26.929 | 63.858 |
| RAD51C | HOM04D007012 | HOM03D007195 | RAD51C | 0.000 | -56.528 | 123.055 |

**Appendix Table S4.** The list of 11 tree species for analysis of the relationship between the copy number ratio of *PARP* and the growth rate.

| Species | Group | Reference |
|---|---|---|
| *Eucalyptus grandis* | Angiosperm | Burns and Honkala (1990b) |
| *Malus domestica* | Angiosperm | Liebhard et al. (2003) |
| *Populus trichocarpa* | Angiosperm | Burns and Honkala (1990b) |
| *Prunus persica* | Angiosperm | Burns and Honkala (1990b) |
| *Picea abies* | Gymnosperm | Kostler (1956) |
| *Picea glauca* | Gymnosperm | Burns and Honkala (1990a) |
| *Picea sitchensis* | Gymnosperm | Burns and Honkala (1990a) |
| *Pinus pinaster* | Gymnosperm | Bravo-Oviedo, Rio and Montero (2004) |
| *Pinus sylvestris* | Gymnosperm | Burns and Honkala (1990a) |
| *Pinus taeda* | Gymnosperm | Burns and Honkala (1990a) |
| *Pseudotsuga menziesii* | Gymnosperm | Burns and Honkala (1990a) |

**Appendix Table S5.** The species list for the analyses. 23 tree species, four shrub species, 15 perennial herb species, 21 annual herb species and 2 alga species were included. Four shrub species were eliminated from the analyses.

| | Species name | Reference |
|---|---|---|
| Tree: 23 species | | |
| Angiosperm | *Amborella trichopoda* | Angiosperm Phylogeny Website |
| | *Carica papaya* | PLANTS database |
| | *Citrus clementina* | Plants For A Future |
| | *Coffea canephora* | Plants of the World online |
| | *Eucalyptus grandis* | PLANTS database |
| | *Hevea brasiliensis* | Plants of the World online |
| | *Malus domestica* | PLANTS database |
| | *Populus trichocarpa* | PLANTS database |
| | *Prunus persica* | PLANTS database |
| | *Pyrus bretschneideri* | Plants For A Future |
| | *Theobroma cacao* | PLANTS database |
| | *Ziziphus jujuba* | The University and Jepson Herbaria |
| Gymnosperm | *Cycas micholitzii* | The Gymnosperm Database |
| | *Ginkgo biloba* | The Gymnosperm Database |
| | *Gnetum Montanum* | The Gymnosperm Database |
| | *Picea abies* | The Gymnosperm Database |
| | *Picea glauca* | The Gymnosperm Database |
| | *Picea sitchensis* | The Gymnosperm Database |
| | *Pinus pinaster* | The Gymnosperm Database |
| | *Pinus sylvestris* | The Gymnosperm Database |
| | *Pinus taeda* | The Gymnosperm Database |
| | *Pseudotsuga menziesii* | The Gymnosperm Database |
| | *Taxus baccata* | The Gymnosperm Database |
| Shrub: 4 species | | |
| | *Actinidia chinensis* | PLANTS database |
| | *Gossypium raimondii* | Gotmare V, Singh P, Tule BN (2000) |
| | *Manihot esculenta* | PLANTS database |
| | *Vitis vinifera* | PLANTS database |

| | | |
|---|---|---|
| Perennial herb: 15 species | | |
| | *Arabidopsis lyrata* | PLANTS database |
| | *Brassica oleracea* | PLANTS database |
| | *Cajanus cajan* | PLANTS database |
| | *Capsicum annuum* | PLANTS database |
| | *Erythranthe guttata* | The University and Jepson Herbaria |
| | *Fragaria vesca* | PLANTS database |
| | *Marchantia polymorpha* | University of Massachusetts Weed Herbarium |
| | *Nelumbo nucifera* | PLANTS database |
| | *Oryza sativa* ssp. *japonica* | Takasaki et al. (1994) |
| | *Ricinus communis* | PLANTS database |
| | *Selaginella moellendorffii* | Zhang, Hans, Kato (2013) |
| | *Solanum lycopersicum* | PLANTS database |
| | *Solanum tuberosum* | PLANTS database |
| | *Trifolium pratense* | PLANTS database |
| | *Utricularia gibba* | PLANTS database |
| Annual herb: 21 species | | |
| | *Amaranthus hypochondriacus* | PLANTS database |
| | *Arabidopsis thaliana* | PLANTS database |
| | *Arachis ipaensis* | Plants of the World online |
| | *Beta vulgaris* | PLANTS database |
| | *Brassica rapa* | PLANTS database |
| | *Capsella rubella* | PLANTS database |
| | *Chenopodium quinoa* | Plants For A Future |
| | *Cicer arietinum* | PLANTS database |
| | *Citrullus lanatus* | PLANTS database |
| | *Corchorus olitorius* | PLANTS database |
| | *Cucumis melo* | PLANTS database |
| | *Cucumis sativus L.* | PLANTS database |
| | *Daucus carota* | PLANTS database |
| | *Glycine max* | PLANTS database |
| | *Medicago truncatula* | Tivoli et al. 2006 |
| | *Petunia axillaris* | PLANTS database |
| | *Physcomitrella patens* | D. Cove 2005 |
| | *Schrenkiella parvula* | Inan, G., Q. Zhang, et al. (2004) |
| | *Tarenaya hassleriana* | PLANTS database |

| | | |
|---|---|---|
| | *Vigna radiata var. radiata* | PLANTS database |
| | *Zea mays* | PLANTS database |
| Alga: 2 species | | |
| | *Chlamydomonas reinhardtii* | Merchant SS et al. 2007 |
| | *Micromonas commoda* | Baren et al. 2016 |

**Appendix Table S6.** The list of species used for the analyses considering the phylogenetic relationships. 23 tree species, 15 perennial herb species and 21 annual herb species were used. Two alga species (*Chlamydomonas reinhardtii* and *Micromonas commoda*) were removed from the analyses because the no sequence data of two alga species was available.

| | Species name |
|---|---|
| Tree: 23 species | |
| Angiosperm | *Amborella trichopoda* |
| | *Carica papaya* |
| | *Citrus clementina* |
| | *Coffea canephora* |
| | *Eucalyptus grandis* |
| | *Hevea brasiliensis* |
| | *Malus domestica* |
| | *Populus trichocarpa* |
| | *Prunus persica* |
| | *Pyrus bretschneideri* |
| | *Theobroma cacao* |
| | *Ziziphus jujuba* |
| Gymnosperm | *Cycas micholitzii* |
| | *Ginkgo biloba* |
| | *Gnetum Montanum* |
| | *Picea abies* |
| | *Picea glauca* |
| | *Picea sitchensis* |
| | *Pinus pinaster* |
| | *Pinus sylvestris* |
| | *Pinus taeda* |
| | *Pseudotsuga menziesii* |
| | *Taxus baccata* |
| Perennial herb: 15 species | |
| | *Arabidopsis lyrata* |

| | |
|---|---|
| | *Brassica oleracea* |
| | *Cajanus cajan* |
| | *Capsicum annuum* |
| | *Erythranthe guttata* |
| | *Fragaria vesca* |
| | *Marchantia polymorpha* |
| | *Nelumbo nucifera* |
| | *Oryza sativa* ssp. *japonica* |
| | *Ricinus communis* |
| | *Selaginella moellendorffii* |
| | *Solanum lycopersicum* |
| | *Solanum tuberosum* |
| | *Trifolium pratense* |
| | *Utricularia gibba* |
| Annual herb: 21 species | |
| | *Amaranthus hypochondriacus* |
| | *Arabidopsis thaliana* |
| | *Arachis ipaensis* |
| | *Beta vulgaris* |
| | *Brassica rapa* |
| | *Capsella rubella* |
| | *Chenopodium quinoa* |
| | *Cicer arietinum* |
| | *Citrullus lanatus* |
| | *Corchorus olitorius* |
| | *Cucumis melo* |
| | *Cucumis sativus L.* |
| | *Daucus carota* |
| | *Glycine max* |
| | *Medicago truncatula* |
| | *Petunia axillaris* |
| | *Physcomitrella patens* |
| | *Schrenkiella parvula* |
| | *Tarenaya hassleriana* |
| | *Vigna radiata var. radiata* |
| | *Zea mays* |

**Appendix Table S7.** The list of 189 *PARP* genes used for the construction of the phylogenetic tree to compare the domain structures of *PARP* genes.

| Gene ID | Species |
|---|---|
| Achn065121 | *Actinidia Chinensis* |
| Achn068031 | *Actinidia Chinensis* |
| Achn200491 | *Actinidia Chinensis* |
| Achn295181 | *Actinidia Chinensis* |
| Achn352311 | *Actinidia Chinensis* |
| Achn359611 | *Actinidia Chinensis* |
| AH002646 | *Amaranthus hypochondriacus* |
| AH013261 | *Amaranthus hypochondriacus* |
| AH022095 | *Amaranthus hypochondriacus* |
| ATR0680G113 | *Amborella trichopoda* |
| ATR0680G401 | *Amborella trichopoda* |
| ATR0706G118 | *Amborella trichopoda* |
| ATR0807G166 | *Amborella trichopoda* |
| AL4G26550 | *Arabidopsis lyrata* |
| AL6G33490 | *Arabidopsis lyrata* |
| AL6G50730 | *Arabidopsis lyrata* |
| AT2G31320 | *Arabidopsis thaliana* |
| AT4G02390 | *Arabidopsis thaliana* |
| AT5G22470 | *Arabidopsis thaliana* |
| Araip.5M8X8 | *Arachis ipaensis* |
| Araip.JYP5G | *Arachis ipaensis* |
| Araip.SKT5W | *Arachis ipaensis* |
| Araip.ZRL1S | *Arachis ipaensis* |
| Bv5_120830_cunf | *Beta vulgaris* |
| Bv7_163730_kdcj | *Beta vulgaris* |
| Bo2g100450 | *Brassica oleracea* |
| Bo2g100460 | *Brassica oleracea* |
| Bo3g052580 | *Brassica oleracea* |
| Bo4g052260 | *Brassica oleracea* |
| Bo9g148430 | *Brassica oleracea* |

| | |
|---|---|
| Brara.B02605 | *Brassica rapa* |
| Brara.C02811 | *Brassica rapa* |
| Brara.E01231 | *Brassica rapa* |
| Brara.J01467 | *Brassica rapa* |
| C.cajan_06726.g | *Cajanus cajan* |
| C.cajan_09672.g | *Cajanus cajan* |
| C.cajan_21742.g | *Cajanus cajan* |
| Carubv10000452m.g | *Capsella rubella* |
| Carubv10002547m.g | *Capsella rubella* |
| Carubv10022570m.g | *Capsella rubella* |
| CAN.G1214.7 | *Capsicum annuum* |
| CAN.G386.7 | *Capsicum annuum* |
| CAN.G461.11 | *Capsicum annuum* |
| CAN.G942.10 | *Capsicum annuum* |
| Cpa.g.sc32.96 | *Carica papaya* |
| Cpa.g.sc50.44 | *Carica papaya* |
| Cpa.g.sc9.254 | *Carica papaya* |
| AUR62008678 | *Chenopodium quinoa* |
| AUR62009776 | *Chenopodium quinoa* |
| AUR62011902 | *Chenopodium quinoa* |
| AUR62024743 | *Chenopodium quinoa* |
| AUR62025568 | *Chenopodium quinoa* |
| AUR62039221 | *Chenopodium quinoa* |
| Ca_03469.g | *Cicer arietinum* |
| Ca_12212.g | *Cicer arietinum* |
| Ca_16481.g | *Cicer arietinum* |
| Cla005994.g | *Citrullus lanatus* |
| Cla005995.g | *Citrullus lanatus* |
| Cla008646.g | *Citrullus lanatus* |
| Cla015093.g | *Citrullus lanatus* |
| Ciclev10018683m.g | *Citrus clementina* |
| Ciclev10019312m.g | *Citrus clementina* |
| Ciclev10027891m.g | *Citrus clementina* |
| Cc01_g09360 | *Coffea canephora* |
| Cc01_g18530 | *Coffea canephora* |
| Cc01_g20930 | *Coffea canephora* |

| | |
|---|---|
| COL.COLO4_05598 | *Corchorus olitorius* |
| COL.COLO4_05599 | *Corchorus olitorius* |
| COL.COLO4_19902 | *Corchorus olitorius* |
| MELO3C015996 | *Cucumis melo* |
| MELO3C021418 | *Cucumis melo* |
| MELO3C024039 | *Cucumis melo* |
| Cucsa.053430 | *Cucumis sativus* |
| Cucsa.205510 | *Cucumis sativus* |
| Cucsa.385080 | *Cucumis sativus* |
| DCAR_012388 | *Daucas carota* |
| DCAR_018467 | *Daucas carota* |
| Migut.D00147 | *Erythranthe guttata* |
| Migut.D00407 | *Erythranthe guttata* |
| Migut.D02355 | *Erythranthe guttata* |
| Eucgr.H01106 | *Eucalyptus grandis* |
| Eucgr.J00484 | *Eucalyptus grandis* |
| Eucgr.K03285 | *Eucalyptus grandis* |
| FVE08249 | *Fragaria vesca* |
| FVE10614 | *Fragaria vesca* |
| FVE22043 | *Fragaria vesca* |
| Glyma.02G017200 | *Glycine max* |
| Glyma.03G161300 | *Glycine max* |
| Glyma.10G017700 | *Glycine max* |
| Glyma.11G184100 | *Glycine max* |
| Glyma.12G088300 | *Glycine max* |
| Glyma.19G162800 | *Glycine max* |
| Gorai.007G127600 | *Gossypium raimondii* |
| Gorai.007G144300 | *Gossypium raimondii* |
| Gorai.009G086300 | *Gossypium raimondii* |
| HBR0402G047 | *Hevea brasiliensis* |
| HBR0402G050 | *Hevea brasiliensis* |
| HBR2393G008 | *Hevea brasiliensis* |
| MDO.mRNA.g.2470.6 | *Malus domestica* |
| MDO.mRNA.g.2470.7 | *Malus domestica* |
| MDO.mRNA.g.2809.8 | *Malus domestica* |
| MDO.mRNA.g.3996.2 | *Malus domestica* |

| | |
|---|---|
| MDO.mRNA.g.4017.1 | *Malus domestica* |
| MDO.mRNA.g.6120.22 | *Malus domestica* |
| MDO.mRNA.g.6120.24 | *Malus domestica* |
| Manes.01G220000 | *Manihot esculenta* |
| Manes.01G220100 | *Manihot esculenta* |
| Manes.05G087700 | *Manihot esculenta* |
| Manes.11G160900 | *Manihot esculenta* |
| Mapoly0074s0022 | *Marchantia polymorpha* |
| Mapoly0154s0015 | *Marchantia polymorpha* |
| Medtr1g088375 | *Medicago truncatula* |
| Medtr1g088400 | *Medicago truncatula* |
| Medtr4g053530 | *Medicago truncatula* |
| Medtr7g096520 | *Medicago truncatula* |
| NNU_03475 | *Nelumbo nucifera* |
| NNU_14032 | *Nelumbo nucifera* |
| NNU_19038 | *Nelumbo nucifera* |
| LOC_Os01g24940 | *Oryza sativa japonica* |
| LOC_Os02g32860 | *Oryza sativa japonica* |
| LOC_Os07g23110 | *Oryza sativa japonica* |
| Peaxi162Scf00134g00123 | *Petunia axillaris* |
| Peaxi162Scf00445g00511 | *Petunia axillaris* |
| Peaxi162Scf00751g00223 | *Petunia axillaris* |
| Peaxi162Scf01281g00019 | *Petunia axillaris* |
| Pp3c1_22640 | *Physcomitrella patens* |
| Pp3c22_13240 | *Physcomitrella patens* |
| Pp3c8_13220 | *Physcomitrella patens* |
| Pp3c8_17220 | *Physcomitrella patens* |
| PAB00011220 | *Picea abies* |
| PAB00016058 | *Picea abies* |
| PAB00021042 | *Picea abies* |
| PAB00059084 | *Picea abies* |
| Potri.002G041300 | *Populus trichocarpa* |
| Potri.004G184100 | *Populus trichocarpa* |
| Potri.009G143932 | *Populus trichocarpa* |
| Potri.014G128000 | *Populus trichocarpa* |
| Potri.014G128200 | *Populus trichocarpa* |

| | |
|---|---|
| Prupe.6G127600 | *Prunus persica* |
| Prupe.8G227600 | *Prunus persica* |
| Prupe.8G262600 | *Prunus persica* |
| Pbr003510.1.g | *Pyrus bretschneideri* |
| Pbr009023.1.g | *Pyrus bretschneideri* |
| Pbr009024.1.g | *Pyrus bretschneideri* |
| Pbr025332.1.g | *Pyrus bretschneideri* |
| Pbr026324.1.g | *Pyrus bretschneideri* |
| Pbr026355.1.g | *Pyrus bretschneideri* |
| RCO.g.29883.000089 | *Ricinus communis* |
| RCO.g.30055.000011 | *Ricinus communis* |
| Tp4g13800 | *Schrenkiella parvula* |
| Tp6g02270 | *Schrenkiella parvula* |
| Tp6g22780 | *Schrenkiella parvula* |
| SMO118G0342 | *Selaginella moellendorffii* |
| SMO353G0427 | *Selaginella moellendorffii* |
| SMO364G0756 | *Selaginella moellendorffii* |
| SMO367G0269 | *Selaginella moellendorffii* |
| Solyc01g009470.1 | *Solanum lycopersicum* |
| Solyc03g117970.2 | *Solanum lycopersicum* |
| Solyc08g074730.1 | *Solanum lycopersicum* |
| Solyc08g074740.2 | *Solanum lycopersicum* |
| Solyc11g067250.1 | *Solanum lycopersicum* |
| PGSC0003DMG400007402 | *Solanum tuberosum* |
| PGSC0003DMG401030070 | *Solanum tuberosum* |
| PGSC0003DMG402030070 | *Solanum tuberosum* |
| THA.LOC104799546 | *Tarenaya hassleriana* |
| THA.LOC104800882 | *Tarenaya hassleriana* |
| THA.LOC104801277 | *Tarenaya hassleriana* |
| TCA.TCM_004107 | *Theobroma cacao* |
| TCA.TCM_004119 | *Theobroma cacao* |
| TCA.TCM_004671 | *Theobroma cacao* |
| TCA.TCM_041443 | *Theobroma cacao* |
| TPR.G17213 | *Trifolium pratense* |
| TPR.G18318 | *Trifolium pratense* |
| TPR.G34005 | *Trifolium pratense* |

| | |
|---|---|
| UGI.Scf00161.10239 | *Utricularia gibba* |
| UGI.Scf01208.20459 | *Utricularia gibba* |
| Vradi02g06900 | *Vigna radiata* |
| Vradi03g01470 | *Vigna radiata* |
| GSVIVG01028029001 | *Vitis vinifera* |
| GSVIVG01028296001 | *Vitis vinifera* |
| GSVIVG01036149001 | *Vitis vinifera* |
| Zm00001d005168 | *Zea mays* |
| Zm00001d009231 | *Zea mays* |
| Zm00001d016694 | *Zea mays* |
| ZJU.LOC107405971 | *Ziziphus jujuba* |
| ZJU.LOC107406331 | *Ziziphus jujuba* |
| ZJU.LOC107409492 | *Ziziphus jujuba* |
| ZJU.LOC107425942 | *Ziziphus jujuba* |
| ZJU.LOC107426250 | *Ziziphus jujuba* |

**Appendix Table S8.** The list of 332 *PARP* genes. (a) 131 *PARP* genes used for the construction of the phylogenetic tree. (b) 201 *PARP* genes removed from the construction of the phylogenetic tree by increasing gap-free site using MaxAlign.

(a)

| Gene ID | Species |
|---|---|
| AH022095 | *Amaranthus hypochondriacus* |
| ATR0680G401 | *Amborella trichopoda* |
| ATR0706G118 | *Amborella trichopoda* |
| AL4G26550 | *Arabidopsis lyrata* |
| AL6G33490 | *Arabidopsis lyrata* |
| AL6G50730 | *Arabidopsis lyrata* |
| AT2G31320 | *Arabidopsis thaliana* |
| AT4G02390 | *Arabidopsis thaliana* |
| AT5G22470 | *Arabidopsis thaliana* |
| Araip.SKT5W | *Arachis ipaensis* |
| Bv5_120830_cunf | *Beta vulgaris* |
| Bv7_163730_kdcj | *Beta vulgaris* |
| Bo3g052580 | *Brassica oleracea* |
| Bo4g052260 | *Brassica oleracea* |
| Bo9g148430 | *Brassica oleracea* |
| Brara.B02605 | *Brassica rapa* |
| Brara.C02811 | *Brassica rapa* |
| Brara.E01231 | *Brassica rapa* |
| Brara.J01467 | *Brassica rapa* |
| C.cajan_06726.g | *Cajanus cajan* |
| C.cajan_09672.g | *Cajanus cajan* |
| C.cajan_21742.g | *Cajanus cajan* |
| Carubv10000452m.g | *Capsella rubella* |
| Carubv10022570m.g | *Capsella rubella* |
| CAN.G461.11 | *Capsicum annuum* |
| CAN.G942.10 | *Capsicum annuum* |
| Cpa.g.sc32.96 | *Carica papaya* |
| Cpa.g.sc9.254 | *Carica papaya* |

(b)

| Gene ID | Species |
|---|---|
| AH002646 | *Amaranthus hypochondriacus* |
| AH013261 | *Amaranthus hypochondriacus* |
| ATR0081G030 | *Amborella trichopoda* |
| ATR0081G068 | *Amborella trichopoda* |
| ATR0680G113 | *Amborella trichopoda* |
| ATR0807G166 | *Amborella trichopoda* |
| AL1G59810 | *Arabidopsis lyrata* |
| Araip.2HK6U | *Arachis ipaensis* |
| Araip.49RC6 | *Arachis ipaensis* |
| Araip.4IM8E | *Arachis ipaensis* |
| Araip.5M8X8 | *Arachis ipaensis* |
| Araip.JYP5G | *Arachis ipaensis* |
| Araip.L3J2U | *Arachis ipaensis* |
| Araip.ZRL1S | *Arachis ipaensis* |
| Bo2g100450 | *Brassica oleracea* |
| Bo2g100460 | *Brassica oleracea* |
| Carubv10002547m.g | *Capsella rubella* |
| CAN.G1214.7 | *Capsicum annuum* |
| CAN.G1214.9 | *Capsicum annuum* |
| CAN.G386.6 | *Capsicum annuum* |
| CAN.G386.7 | *Capsicum annuum* |
| Cpa.g.sc50.44 | *Carica papaya* |
| Cpa.g.sc50.45 | *Carica papaya* |
| AUR62008678 | *Chenopodium quinoa* |
| AUR62011902 | *Chenopodium quinoa* |
| AUR62025568 | *Chenopodium quinoa* |
| Cla005995.g | *Citrullus lanatus* |
| Cla008646.g | *Citrullus lanatus* |

| | |
|---|---|
| AUR62009776 | *Chenopodium quinoa* |
| AUR62024743 | *Chenopodium quinoa* |
| AUR62039221 | *Chenopodium quinoa* |
| Ca_03469.g | *Cicer arietinum* |
| Ca_12212.g | *Cicer arietinum* |
| Ca_16481.g | *Cicer arietinum* |
| Cla005994.g | *Citrullus lanatus* |
| Cla015093.g | *Citrullus lanatus* |
| Ciclev10018683m.g | *Citrus clementina* |
| Ciclev10019312m.g | *Citrus clementina* |
| Cc00_g22450 | *Coffea canephora* |
| Cc01_g18530 | *Coffea canephora* |
| Cc01_g20930 | *Coffea canephora* |
| COL.COLO4_19902 | *Corchorus olitorius* |
| MELO3C024039 | *Cucumis melo* |
| Cucsa.053430 | *Cucumis sativus* |
| Cucsa.205510 | *Cucumis sativus* |
| Cucsa.385080 | *Cucumis sativus* |
| DCAR_012388 | *Daucas carota* |
| DCAR_018467 | *Daucas carota* |
| Migut.D00147 | *Erythranthe guttata* |
| Migut.D00407 | *Erythranthe guttata* |
| Migut.D02355 | *Erythranthe guttata* |
| Eucgr.H01106 | *Eucalyptus grandis* |
| Eucgr.J00484 | *Eucalyptus grandis* |
| Eucgr.K03285 | *Eucalyptus grandis* |
| FVE08249 | *Fragaria vesca* |
| FVE22043 | *Fragaria vesca* |
| Glyma.03G161300 | *Glycine max* |
| Glyma.10G017700 | *Glycine max* |
| Glyma.11G184100 | *Glycine max* |
| Glyma.12G088300 | *Glycine max* |
| Glyma.19G162800 | *Glycine max* |
| GMO00017089 | *Gnetum montanum* |
| GMO00017354 | *Gnetum montanum* |
| HBR0402G050 | *Hevea brasiliensis* |
| Ciclev10023303m.g | *Citrus clementina* |
| Ciclev10027891m.g | *Citrus clementina* |
| Cc01_g09350 | *Coffea canephora* |
| Cc01_g09360 | *Coffea canephora* |
| COL.COLO4_05598 | *Corchorus olitorius* |
| COL.COLO4_05599 | *Corchorus olitorius* |
| COL.COLO4_23334 | *Corchorus olitorius* |
| COL.COLO4_23335 | *Corchorus olitorius* |
| MELO3C015996 | *Cucumis melo* |
| MELO3C021418 | *Cucumis melo* |
| MELO3C021419 | *Cucumis melo* |
| MELO3C021420 | *Cucumis melo* |
| CMI00004336 | *Cycas micholitzii* |
| CMI00005428 | *Cycas micholitzii* |
| CMI00018239 | *Cycas micholitzii* |
| CMI00021647 | *Cycas micholitzii* |
| DCAR_012185 | *Daucas carota* |
| DCAR_012186 | *Daucas carota* |
| FVE10614 | *Fragaria vesca* |
| GBI00004299 | *Ginkgo biloba* |
| GBI00008714 | *Ginkgo biloba* |
| GBI00009097 | *Ginkgo biloba* |
| GBI00023514 | *Ginkgo biloba* |
| Glyma.02G017200 | *Glycine max* |
| Glyma.10G124100 | *Glycine max* |
| HBR0402G047 | *Hevea brasiliensis* |
| HBR1831G016 | *Hevea brasiliensis* |
| HBR3468G023 | *Hevea brasiliensis* |
| MDO.mRNA.g.2470.6 | *Malus domestica* |
| MDO.mRNA.g.2470.7 | *Malus domestica* |
| MDO.mRNA.g.2809.7 | *Malus domestica* |
| MDO.mRNA.g.2809.8 | *Malus domestica* |
| MDO.mRNA.g.357.5 | *Malus domestica* |
| MDO.mRNA.g.357.6 | *Malus domestica* |
| MDO.mRNA.g.357.7 | *Malus domestica* |
| MDO.mRNA.g.3996.2 | *Malus domestica* |

| | |
|---|---|
| HBR2393G008 | *Hevea brasiliensis* |
| Mapoly0074s0022 | *Marchantia polymorpha* |
| Mapoly0154s0015 | *Marchantia polymorpha* |
| Medtr1g088375 | *Medicago truncatula* |
| Medtr4g053530 | *Medicago truncatula* |
| Medtr7g096520 | *Medicago truncatula* |
| NNU_14032 | *Nelumbo nucifera* |
| NNU_19038 | *Nelumbo nucifera* |
| LOC_Os01g24940 | *Oryza sativa japonica* |
| LOC_Os02g32860 | *Oryza sativa japonica* |
| LOC_Os07g23110 | *Oryza sativa japonica* |
| Peaxi162Scf00445g00511 | *Petunia axillaris* |
| Peaxi162Scf00757g00223 | *Petunia axillaris* |
| Peaxi162Scf01281g00019 | *Petunia axillaris* |
| Pp3c1_22640 | *Physcomitrella patens* |
| Pp3c22_13240 | *Physcomitrella patens* |
| Pp3c8_17220 | *Physcomitrella patens* |
| PAB00021042 | *Picea abies* |
| PPI00058999 | *Pinus pinaster* |
| PPI00073846 | *Pinus pinaster* |
| PSY00007693 | *Pinus sylvestris* |
| PSY00015729 | *Pinus sylvestris* |
| PTA00003970 | *Pinus taeda* |
| PTA00019626 | *Pinus taeda* |
| Potri.002G041300 | *Populus trichocarpa* |
| Potri.014G128200 | *Populus trichocarpa* |
| Prupe.6G127600 | *Prunus persica* |
| Prupe.8G227600 | *Prunus persica* |
| Prupe.8G262600 | *Prunus persica* |
| PME00007555 | *Pseudotsuga menziesii* |
| PME00051383 | *Pseudotsuga menziesii* |
| PME00094295 | *Pseudotsuga menziesii* |
| Pbr003510.1.g | *Pyrus bretschneideri* |
| Pbr009023.1.g | *Pyrus bretschneideri* |
| Pbr025332.1.g | *Pyrus bretschneideri* |
| Pbr026324.1.g | *Pyrus bretschneideri* |
| MDO.mRNA.g.4017.1 | *Malus domestica* |
| MDO.mRNA.g.4963.10 | *Malus domestica* |
| MDO.mRNA.g.4963.9 | *Malus domestica* |
| MDO.mRNA.g.6120.21 | *Malus domestica* |
| MDO.mRNA.g.6120.22 | *Malus domestica* |
| MDO.mRNA.g.6120.24 | *Malus domestica* |
| Mapoly0030s0138 | *Marchantia polymorpha* |
| Medtr1g088400 | *Medicago truncatula* |
| NNU_03475 | *Nelumbo nucifera* |
| NNU_07935 | *Nelumbo nucifera* |
| LOC_Os01g24920 | *Oryza sativa japonica* |
| Peaxi162Scf00134g00123 | *Petunia axillaris* |
| Peaxi162Scf00751g00217 | *Petunia axillaris* |
| Peaxi162Scf00751g00223 | *Petunia axillaris* |
| Peaxi162Scf00751g00224 | *Petunia axillaris* |
| Pp3c8_13220 | *Physcomitrella patens* |
| PAB00001919 | *Picea abies* |
| PAB00002850 | *Picea abies* |
| PAB00002955 | *Picea abies* |
| PAB00011220 | *Picea abies* |
| PAB00016058 | *Picea abies* |
| PAB00043164 | *Picea abies* |
| PAB00044039 | *Picea abies* |
| PAB00046641 | *Picea abies* |
| PAB00059084 | *Picea abies* |
| PGL00009845 | *Picea glauca* |
| PGL00011348 | *Picea glauca* |
| PSI00003629 | *Picea sitchensis* |
| PPI00000081 | *Pinus pinaster* |
| PPI00003071 | *Pinus pinaster* |
| PPI00037856 | *Pinus pinaster* |
| PPI00038529 | *Pinus pinaster* |
| PPI00042280 | *Pinus pinaster* |
| PPI00050647 | *Pinus pinaster* |
| PPI00052742 | *Pinus pinaster* |
| PPI00053106 | *Pinus pinaster* |

| | |
|---|---|
| Pbr026355.1.g | *Pyrus bretschneideri* |
| RCO.g.29883.000089 | *Ricinus communis* |
| RCO.g.30055.000011 | *Ricinus communis* |
| Tp4g13800 | *Schrenkiella parvula* |
| Tp6g02270 | *Schrenkiella parvula* |
| Tp6g22780 | *Schrenkiella parvula* |
| SMO118G0342 | *Selaginella moellendorffii* |
| SMO353G0427 | *Selaginella moellendorffii* |
| SMO364G0756 | *Selaginella moellendorffii* |
| SMO367G0269 | *Selaginella moellendorffii* |
| Solyc01g009470.1 | *Solanum lycopersicum* |
| Solyc03g117970.2 | *Solanum lycopersicum* |
| Solyc08g074730.1 | *Solanum lycopersicum* |
| Solyc08g074740.2 | *Solanum lycopersicum* |
| Solyc11g067250.1 | *Solanum lycopersicum* |
| PGSC0003DMG400007402 | *Solanum tuberosum* |
| PGSC0003DMG401030070 | *Solanum tuberosum* |
| PGSC0003DMG402030070 | *Solanum tuberosum* |
| THA.LOC104799546 | *Tarenaya hassleriana* |
| THA.LOC104800882 | *Tarenaya hassleriana* |
| THA.LOC104801277 | *Tarenaya hassleriana* |
| TCA.TCM_004107 | *Theobroma cacao* |
| TCA.TCM_004671 | *Theobroma cacao* |
| TCA.TCM_041443 | *Theobroma cacao* |
| TPR.G17213 | *Trifolium pratense* |
| TPR.G34005 | *Trifolium pratense* |
| UGI.Scf00161.10239 | *Utricularia gibba* |
| Vradi02g06900 | *Vigna radiata* |
| Zm00001d016694 | *Zea mays* |
| ZJU.LOC107405971 | *Ziziphus jujuba* |
| ZJU.LOC107425942 | *Ziziphus jujuba* |

| | |
|---|---|
| PPI00066288 | *Pinus pinaster* |
| PPI00071432 | *Pinus pinaster* |
| PPI00075862 | *Pinus pinaster* |
| PPI00076222 | *Pinus pinaster* |
| PSY00000933 | *Pinus sylvestris* |
| PSY00000934 | *Pinus sylvestris* |
| PSY00002174 | *Pinus sylvestris* |
| PSY00003099 | *Pinus sylvestris* |
| PSY00011283 | *Pinus sylvestris* |
| PSY00011284 | *Pinus sylvestris* |
| PSY00017560 | *Pinus sylvestris* |
| PSY00027634 | *Pinus sylvestris* |
| PTA00011977 | *Pinus taeda* |
| PTA00012649 | *Pinus taeda* |
| PTA00029900 | *Pinus taeda* |
| PTA00044382 | *Pinus taeda* |
| PTA00044383 | *Pinus taeda* |
| PTA00044384 | *Pinus taeda* |
| PTA00048519 | *Pinus taeda* |
| PTA00076307 | *Pinus taeda* |
| Potri.004G184100 | *Populus trichocarpa* |
| Potri.009G136500 | *Populus trichocarpa* |
| Potri.009G143866 | *Populus trichocarpa* |
| Potri.009G143932 | *Populus trichocarpa* |
| Potri.014G128000 | *Populus trichocarpa* |
| Potri.014G128100 | *Populus trichocarpa* |
| Prupe.3G262400 | *Prunus persica* |
| Prupe.3G262700 | *Prunus persica* |
| Prupe.5G191000 | *Prunus persica* |
| PME00008631 | *Pseudotsuga menziesii* |
| PME00008632 | *Pseudotsuga menziesii* |
| PME00019315 | *Pseudotsuga menziesii* |
| PME00038040 | *Pseudotsuga menziesii* |
| PME00051377 | *Pseudotsuga menziesii* |
| PME00051378 | *Pseudotsuga menziesii* |
| PME00051379 | *Pseudotsuga menziesii* |

| | |
|---|---|
| PME00051380 | *Pseudotsuga menziesii* |
| PME00051381 | *Pseudotsuga menziesii* |
| PME00051382 | *Pseudotsuga menziesii* |
| PME00051384 | *Pseudotsuga menziesii* |
| PME00051385 | *Pseudotsuga menziesii* |
| PME00051386 | *Pseudotsuga menziesii* |
| PME00051387 | *Pseudotsuga menziesii* |
| PME00051388 | *Pseudotsuga menziesii* |
| PME00051389 | *Pseudotsuga menziesii* |
| PME00051390 | *Pseudotsuga menziesii* |
| PME00068074 | *Pseudotsuga menziesii* |
| PME00068076 | *Pseudotsuga menziesii* |
| PME00068077 | *Pseudotsuga menziesii* |
| PME00068078 | *Pseudotsuga menziesii* |
| PME00068079 | *Pseudotsuga menziesii* |
| PME00068080 | *Pseudotsuga menziesii* |
| PME00068082 | *Pseudotsuga menziesii* |
| PME00068083 | *Pseudotsuga menziesii* |
| PME00068084 | *Pseudotsuga menziesii* |
| PME00068085 | *Pseudotsuga menziesii* |
| PME00068086 | *Pseudotsuga menziesii* |
| PME00068088 | *Pseudotsuga menziesii* |
| PME00068089 | *Pseudotsuga menziesii* |
| PME00068090 | *Pseudotsuga menziesii* |
| PME00068092 | *Pseudotsuga menziesii* |
| PME00068093 | *Pseudotsuga menziesii* |
| PME00068094 | *Pseudotsuga menziesii* |
| PME00068096 | *Pseudotsuga menziesii* |
| PME00068097 | *Pseudotsuga menziesii* |
| PME00068099 | *Pseudotsuga menziesii* |
| PME00068100 | *Pseudotsuga menziesii* |
| PME00068102 | *Pseudotsuga menziesii* |
| PME00068103 | *Pseudotsuga menziesii* |
| PME00068104 | *Pseudotsuga menziesii* |
| PME00068105 | *Pseudotsuga menziesii* |
| PME00068107 | *Pseudotsuga menziesii* |

| | |
|---|---|
| PME00068108 | *Pseudotsuga menziesii* |
| PME00099600 | *Pseudotsuga menziesii* |
| PME00099602 | *Pseudotsuga menziesii* |
| PME00131356 | *Pseudotsuga menziesii* |
| PME00142152 | *Pseudotsuga menziesii* |
| Pbr003252.1.g | *Pyrus bretschneideri* |
| Pbr009024.1.g | *Pyrus bretschneideri* |
| Pbr021722.1.g | *Pyrus bretschneideri* |
| Pbr035027.1.g | *Pyrus bretschneideri* |
| RCO.g.29986.000043 | *Ricinus communis* |
| RCO.g.29986.000044 | *Ricinus communis* |
| SMO364G0880 | *Selaginella moellendorffii* |
| TBA00002240 | *Taxus baccata* |
| TBA00007115 | *Taxus baccata* |
| TBA00007116 | *Taxus baccata* |
| TBA00027172 | *Taxus baccata* |
| TCA.TCM_004119 | *Theobroma cacao* |
| TCA.TCM_004120 | *Theobroma cacao* |
| TPR.G16288 | *Trifolium pratense* |
| TPR.G18318 | *Trifolium pratense* |
| UGI.Scf01208.20459 | *Utricularia gibba* |
| Vradi03g01470 | *Vigna radiata* |
| Zm00001d005168 | *Zea mays* |
| Zm00001d009230 | *Zea mays* |
| Zm00001d009231 | *Zea mays* |
| ZJU.LOC107406331 | *Ziziphus jujuba* |
| ZJU.LOC107409492 | *Ziziphus jujuba* |
| ZJU.LOC107426250 | *Ziziphus jujuba* |
| ZJU.LOC107426308 | *Ziziphus jujuba* |

**Appendix Table S9.** The list of *PARP* genes within gymnosperm species. (a) 24 *PARP* genes used for the construction of the phylogenetic tree. (b) 88 *PARP* genes removed from the construction of the phylogenetic tree by increasing gap-free site using MaxAlign.

(a)

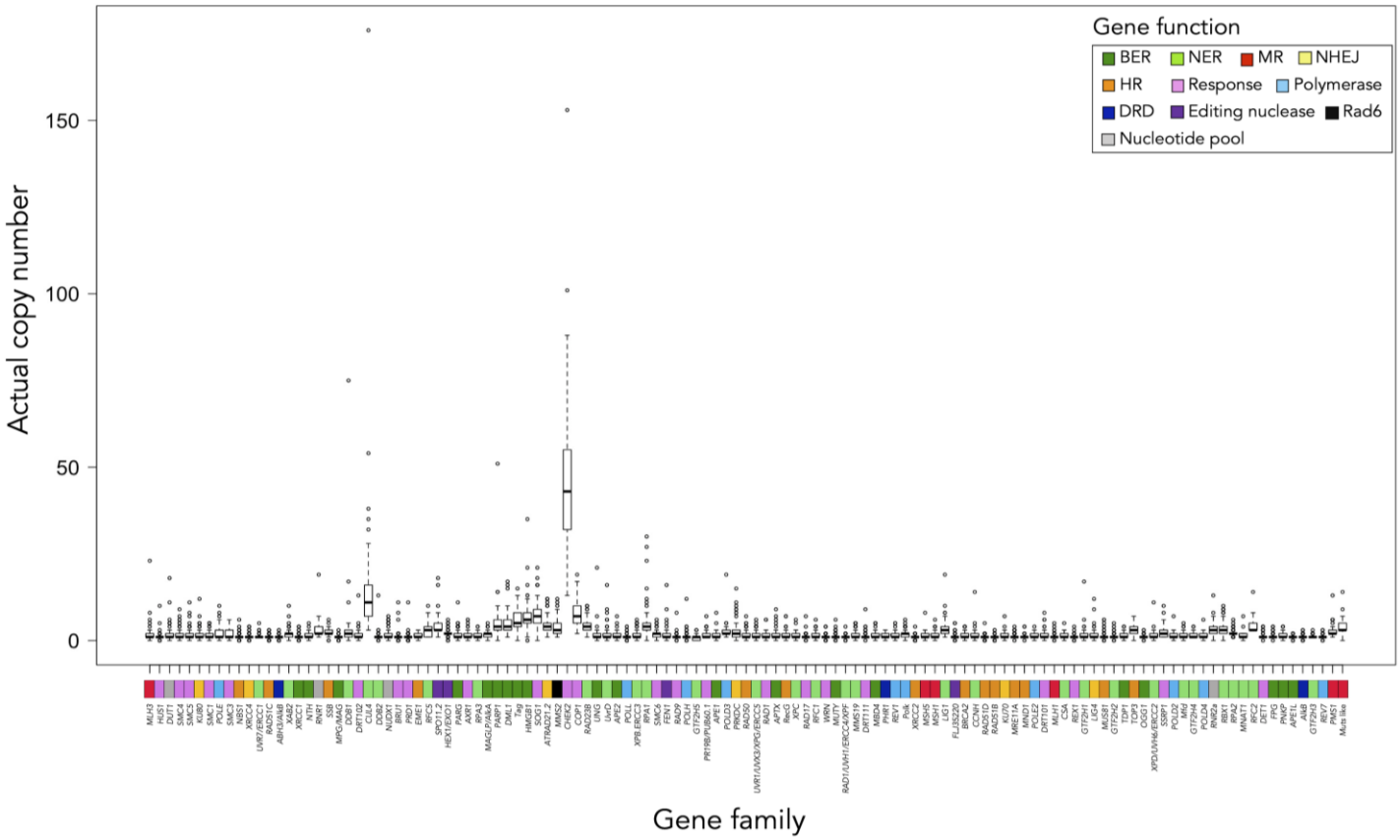
| Gene ID | Species |
|---|---|
| GMO00017089 | *Gnetum montanum* |
| GMO00017354 | *Gnetum montanum* |
| PAB00021042 | *Picea abies* |
| PPI00058999 | *Pinus pinaster* |
| PPI00073846 | *Pinus pinaster* |
| PSY00003099 | *Pinus sylvestris* |
| PSY00007693 | *Pinus sylvestris* |
| PSY00015729 | *Pinus sylvestris* |
| PTA00003970 | *Pinus taeda* |
| PTA00019626 | *Pinus taeda* |
| PME00007555 | *Pseudotsuga menziesii* |
| PME00008631 | *Pseudotsuga menziesii* |
| PME00051377 | *Pseudotsuga menziesii* |
| PME00051379 | *Pseudotsuga menziesii* |
| PME00051381 | *Pseudotsuga menziesii* |
| PME00051383 | *Pseudotsuga menziesii* |
| PME00051384 | *Pseudotsuga menziesii* |
| PME00051386 | *Pseudotsuga menziesii* |
| PME00051387 | *Pseudotsuga menziesii* |
| PME00051389 | *Pseudotsuga menziesii* |
| PME00068085 | *Pseudotsuga menziesii* |
| PME00068096 | *Pseudotsuga menziesii* |
| PME00068099 | *Pseudotsuga menziesii* |
| PME00094295 | *Pseudotsuga menziesii* |

(b)

| Gene ID | *Species* |
|---|---|
| CMI00004336 | *Cycas micholitzii* |
| CMI00005428 | *Cycas micholitzii* |
| CMI00018239 | *Cycas micholitzii* |
| CMI00021647 | *Cycas micholitzii* |
| GBI00004299 | *Ginkgo biloba* |
| GBI00008714 | *Ginkgo biloba* |
| GBI00009097 | *Ginkgo biloba* |
| GBI00023514 | *Ginkgo biloba* |
| PAB00001919 | *Picea abies* |
| PAB00002850 | *Picea abies* |
| PAB00002955 | *Picea abies* |
| PAB00011220 | *Picea abies* |
| PAB00016058 | *Picea abies* |
| PAB00043164 | *Picea abies* |
| PAB00044039 | *Picea abies* |
| PAB00046641 | *Picea abies* |
| PAB00059084 | *Picea abies* |
| PGL00009845 | *Picea glauca* |
| PGL00011348 | *Picea glauca* |
| PSI00003629 | *Picea sitchensis* |
| PPI00000081 | *Pinus pinaster* |
| PPI00003071 | *Pinus pinaster* |
| PPI00037856 | *Pinus pinaster* |
| PPI00038529 | *Pinus pinaster* |
| PPI00042280 | *Pinus pinaster* |
| PPI00050647 | *Pinus pinaster* |

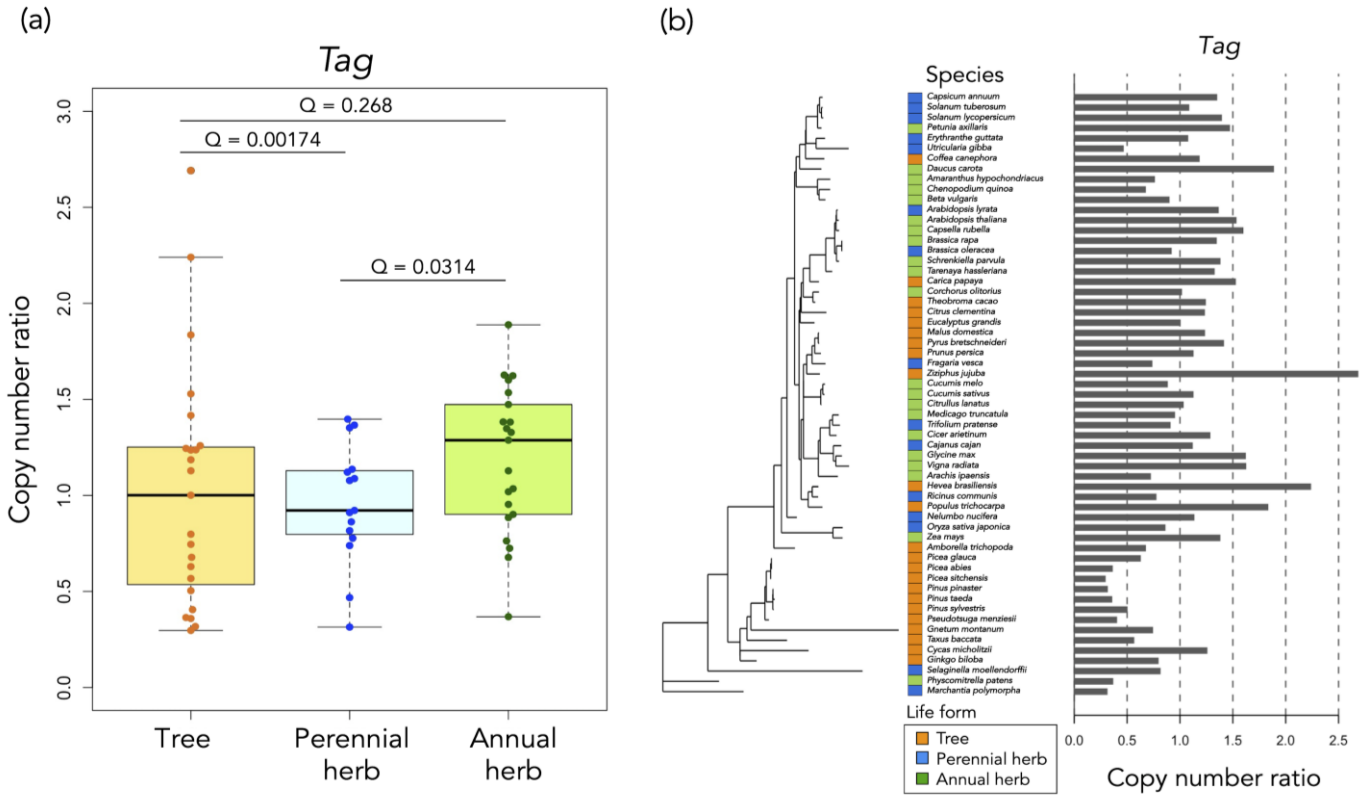
| | |
|---|---|
| PPI00052742 | *Pinus pinaster* |
| PPI00053106 | *Pinus pinaster* |
| PPI00066288 | *Pinus pinaster* |
| PPI00071432 | *Pinus pinaster* |
| PPI00075862 | *Pinus pinaster* |
| PPI00076222 | *Pinus pinaster* |
| PSY00000933 | *Pinus sylvestris* |
| PSY00000934 | *Pinus sylvestris* |
| PSY00002174 | *Pinus sylvestris* |
| PSY00011283 | *Pinus sylvestris* |
| PSY00011284 | *Pinus sylvestris* |
| PSY00017560 | *Pinus sylvestris* |
| PSY00027634 | *Pinus sylvestris* |
| PTA00011977 | *Pinus taeda* |
| PTA00012649 | *Pinus taeda* |
| PTA00029900 | *Pinus taeda* |
| PTA00044382 | *Pinus taeda* |
| PTA00044383 | *Pinus taeda* |
| PTA00044384 | *Pinus taeda* |
| PTA00048519 | *Pinus taeda* |
| PTA00076307 | *Pinus taeda* |
| PME00008632 | *Pseudotsuga menziesii* |
| PME00019315 | *Pseudotsuga menziesii* |
| PME00038040 | *Pseudotsuga menziesii* |
| PME00051378 | *Pseudotsuga menziesii* |
| PME00051380 | *Pseudotsuga menziesii* |
| PME00051382 | *Pseudotsuga menziesii* |
| PME00051385 | *Pseudotsuga menziesii* |
| PME00051388 | *Pseudotsuga menziesii* |
| PME00051390 | *Pseudotsuga menziesii* |
| PME00068074 | *Pseudotsuga menziesii* |
| PME00068076 | *Pseudotsuga menziesii* |
| PME00068077 | *Pseudotsuga menziesii* |
| PME00068078 | *Pseudotsuga menziesii* |
| PME00068079 | *Pseudotsuga menziesii* |
| PME00068080 | *Pseudotsuga menziesii* |

| | |
|---|---|
| PME00068082 | *Pseudotsuga menziesii* |
| PME00068083 | *Pseudotsuga menziesii* |
| PME00068084 | *Pseudotsuga menziesii* |
| PME00068086 | *Pseudotsuga menziesii* |
| PME00068088 | *Pseudotsuga menziesii* |
| PME00068089 | *Pseudotsuga menziesii* |
| PME00068090 | *Pseudotsuga menziesii* |
| PME00068092 | *Pseudotsuga menziesii* |
| PME00068093 | *Pseudotsuga menziesii* |
| PME00068094 | *Pseudotsuga menziesii* |
| PME00068097 | *Pseudotsuga menziesii* |
| PME00068100 | *Pseudotsuga menziesii* |
| PME00068102 | *Pseudotsuga menziesii* |
| PME00068103 | *Pseudotsuga menziesii* |
| PME00068104 | *Pseudotsuga menziesii* |
| PME00068105 | *Pseudotsuga menziesii* |
| PME00068107 | *Pseudotsuga menziesii* |
| PME00068108 | *Pseudotsuga menziesii* |
| PME00099600 | *Pseudotsuga menziesii* |
| PME00099602 | *Pseudotsuga menziesii* |
| PME00131356 | *Pseudotsuga menziesii* |
| PME00142152 | *Pseudotsuga menziesii* |
| TBA00002240 | *Taxus baccata* |
| TBA00007115 | *Taxus baccata* |
| TBA00007116 | *Taxus baccata* |
| TBA00027172 | *Taxus baccata* |

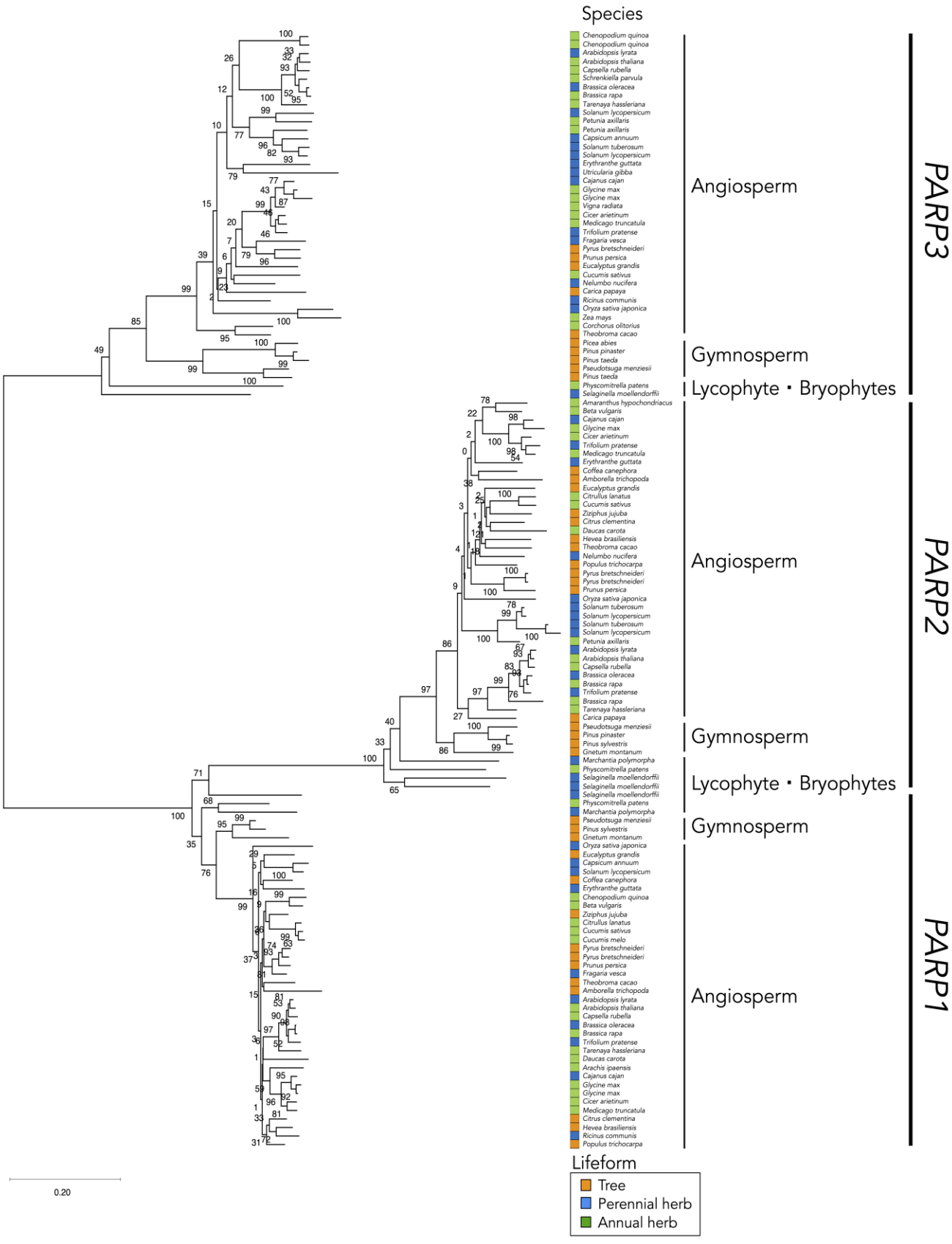



**Appendix Figure S1.** The actual copy number of 121 gene families associated with DNA repair, related to Figure 1. The symbols of the genes within each gene family are shown on the horizontal axis. The horizontal line inside the box showed the median and the length of box showed the interquartile range (range between the 25$^{th}$ to 75$^{th}$ percentiles). The whiskers indicated points within 1.5 times the interquartile rage. The points beyond the whisker range indicated the outliers. The gene families were ordered according to the result of hierarchical clustering. The order of gene families corresponded to the order of gene families in main figure 1a. Each gene family was categorized into one of 11 groups: BER, base excision repair; NER, nucleotide excision repair; MR, mismatch repair; NHEJ, nonhomologous end-joining repair; HR, homologous recombination repair; Response, DNA damage response; Polymerase,

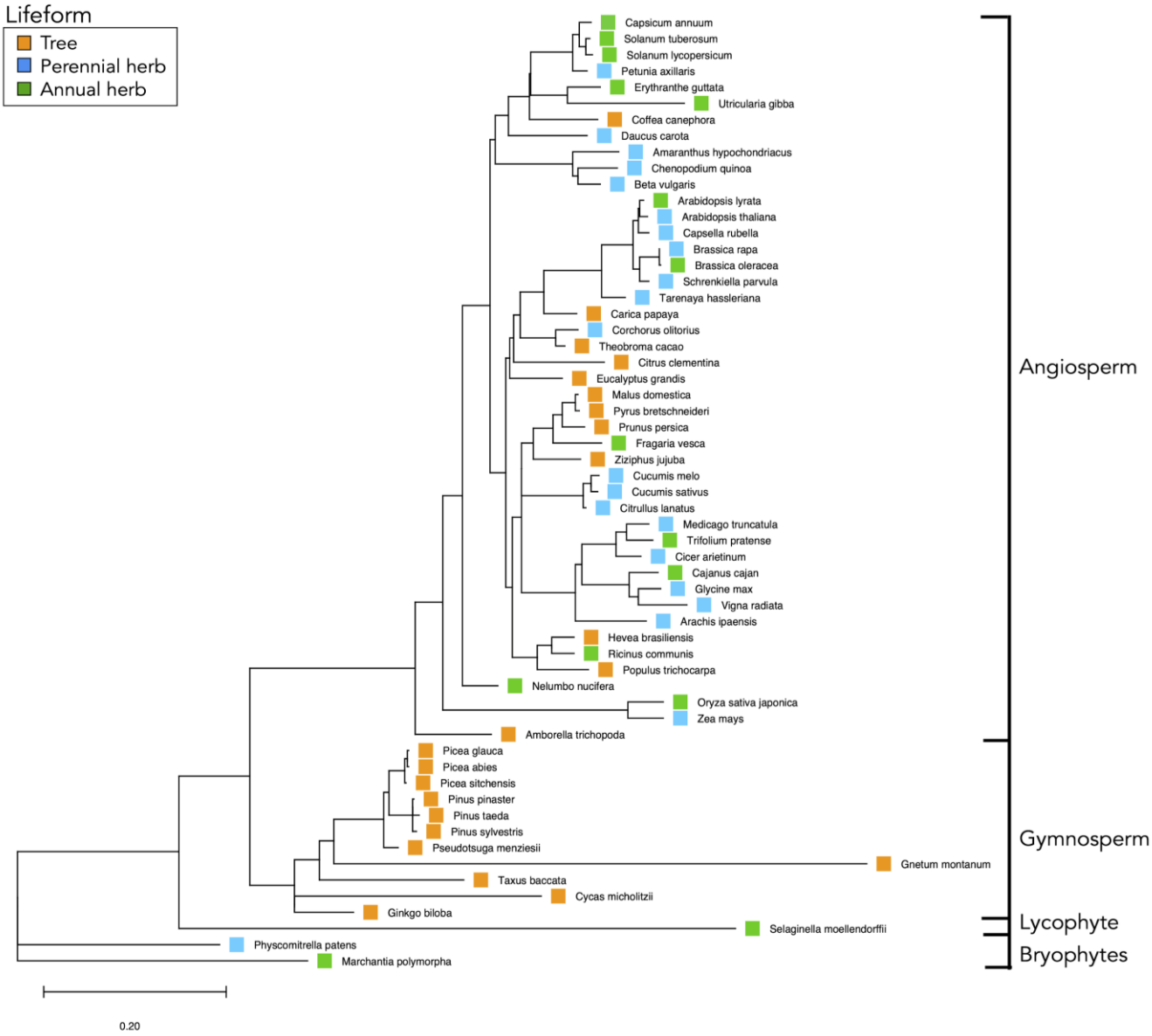
DNA polymerase; DRD, direct reversal of damage; Editing nuclease, editing and processing nuclease; Rad6, Rad6 pathway; Nucleotide pool, modulation of nucleotide pool.



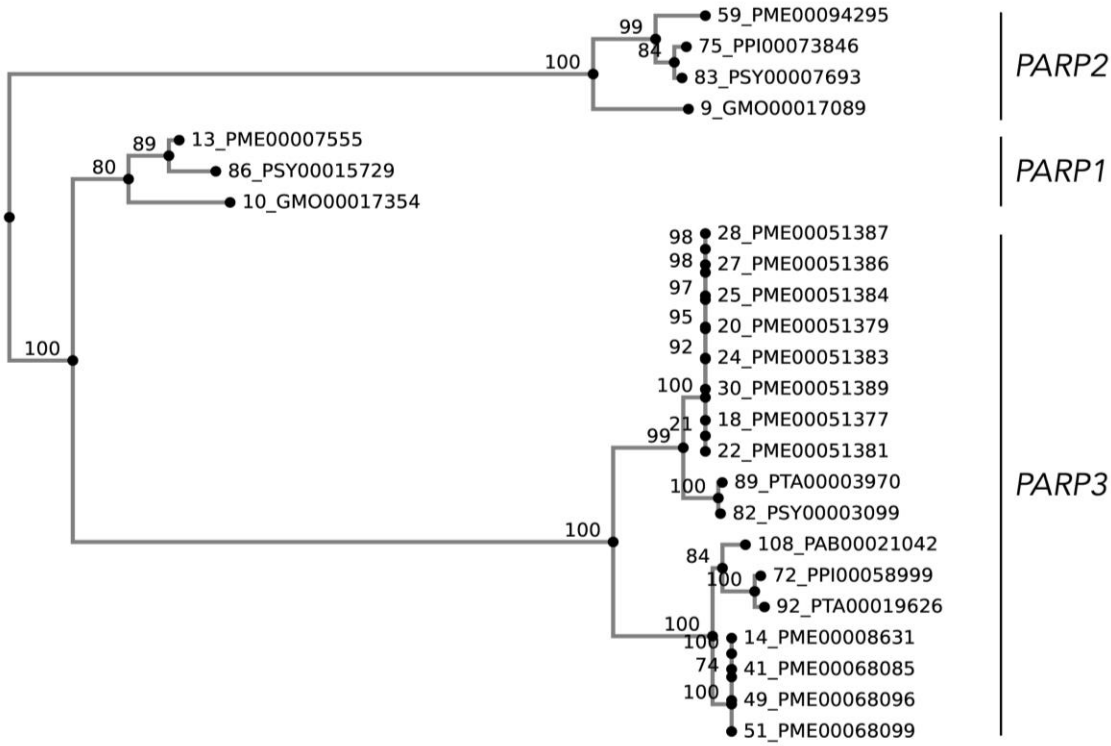

**Appendix Figure S2.** Comparison analysis of the copy number ratio of *Tag* gene families among life forms, related to Figure 2. (a) Box plot of the copy number ratios in different life forms. Tree species had significantly higher copy number ratios than perennial herb species (coefficient = −0.646, standard error = 0.136, *t*-value = −4.75, P-value = $1.46\times10^{-5}$, Q-value = 0.00174). There was no significant difference between tree species and annual herb species (coefficient = −0.326, standard error = 0.135, *t*-value = −2.41, P-value = 0.0194, Q-value = 0.268). The horizontal line inside the box showed the median and the length of box showed the interquartile range (range between the 25$^{th}$ to 75$^{th}$ percentiles). The whiskers indicated points within 1.5 times the interquartile rage. The points beyond the whisker range indicated the outliers. (b) The phylogenetic relationships of the copy number ratios of the *Tag* gene family. The estimated Pagel's lambda was 0.982.


Species
Chenopodium quinoa
Chenopodium quinoa
Arabidopsis lyrata
Arabidopsis thaliana
Capsella rubella
Schrenkiella parvula
Brassica oleracea
Brassica rapa
Tarenaya hassleriana
Solanum lycopersicum
Petunia axillaris
Petunia axillaris
Capsicum annuum
Solanum tuberosum
Solanum lycopersicum
Erythranthe guttata
Utricularia gibba
Cajanus cajan
Glycine max
Glycine max
Vigna radiata
Cicer arietinum
Medicago truncatula
Trifolium pratense
Fragaria vesca
Pyrus bretschneideri
Prunus persica
Eucalyptus grandis
Cucumis sativus
Nelumbo nucifera
Carica papaya
Ricinus communis
Oryza sativa japonica
Zea mays
Corchorus olitorius
Theobroma cacao
Picea abies
Pinus pinaster
Pinus taeda
Pseudotsuga menziesii
Pinus taeda
Physcomitrella patens
Selaginella moellendorffii
Amaranthus hypochondriacus
Beta vulgaris
Cajanus cajan
Glycine max
Cicer arietinum
Trifolium pratense
Medicago truncatula
Erythranthe guttata
Coffea canephora
Amborella trichopoda
Eucalyptus grandis
Citrullus lanatus
Cucumis sativus
Ziziphus jujuba
Citrus clementina
Daucas carota
Hevea brasiliensis
Theobroma cacao
Nelumbo nucifera
Populus trichocarpa
Pyrus bretschneideri
Pyrus bretschneideri
Prunus persica
Oryza sativa japonica
Solanum tuberosum
Solanum lycopersicum
Solanum tuberosum
Solanum lycopersicum
Petunia axillaris
Arabidopsis lyrata
Arabidopsis thaliana
Capsella rubella
Brassica oleracea
Brassica rapa
Trifolium pratense
Brassica rapa
Tarenaya hassleriana
Carica papaya
Pseudotsuga menziesii
Pinus pinaster
Pinus sylvestris
Gnetum montanum
Marchantia polymorpha
Physcomitrella patens
Selaginella moellendorffii
Selaginella moellendorffii
Selaginella moellendorffii
Physcomitrella patens
Marchantia polymorpha
Pseudotsuga menziesii
Pinus sylvestris
Gnetum montanum
Oryza sativa japonica
Eucalyptus grandis
Capsicum annuum
Solanum lycopersicum
Coffea canephora
Erythranthe guttata
Chenopodium quinoa
Beta vulgaris
Ziziphus jujuba
Citrullus lanatus
Cucumis sativus
Cucumis melo
Pyrus bretschneideri
Pyrus bretschneideri
Prunus persica
Fragaria vesca
Theobroma cacao
Amborella trichopoda
Arabidopsis lyrata
Arabidopsis thaliana
Capsella rubella
Brassica oleracea
Brassica rapa
Trifolium pratense
Tarenaya hassleriana
Daucas carota
Arachis paensis
Cajanus cajan
Glycine max
Glycine max
Cicer arietinum
Medicago truncatula
Citrus clementina
Hevea brasiliensis
Ricinus communis
Populus trichocarpa
Angiosperm
Gymnosperm
Lycophyte・Bryophytes
Angiosperm
Gymnosperm
Lycophyte・Bryophytes
Gymnosperm
Angiosperm
PARP3
PARP2
PARP1
Lifeform
Tree
Perennial herb
Annual herb
0.20

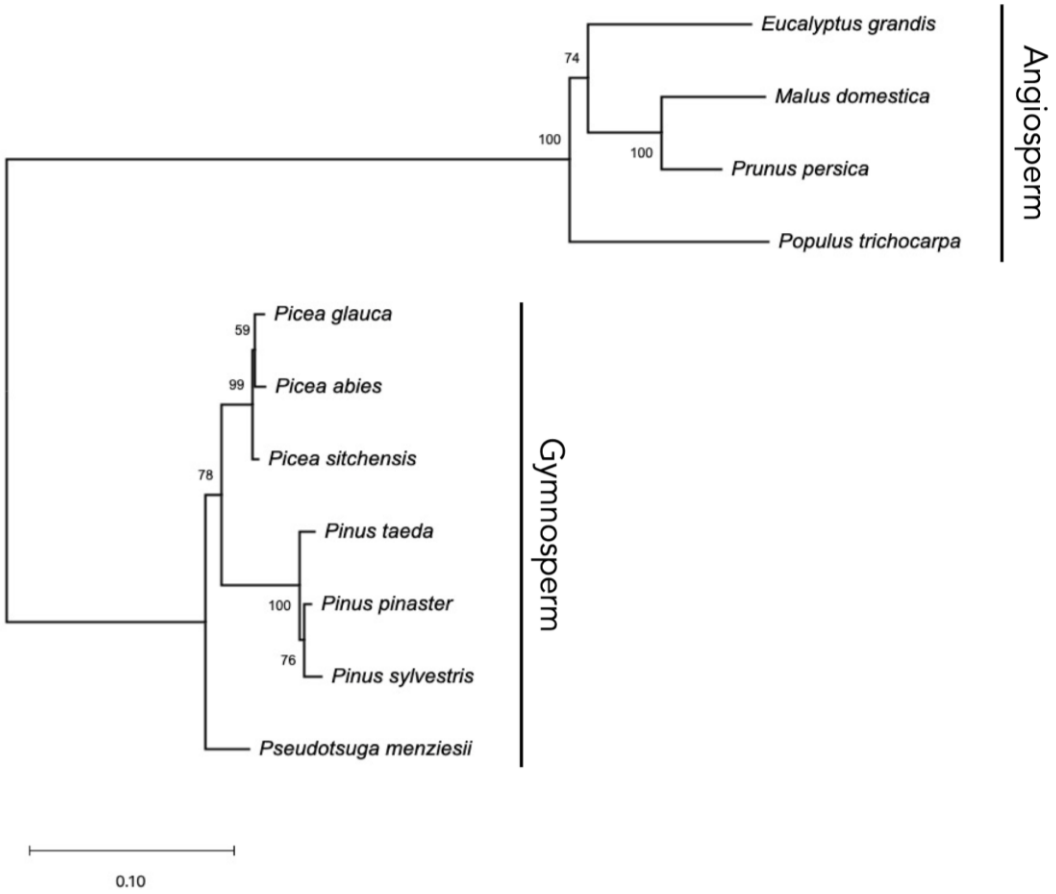
**Appendix Figure S3.** The phylogenetic tree of *PARP* gene family of species in the dataset, related to Figure 3 and Table 2. 131 genes in the species including angiosperms, gymnosperms, lycophyte, and bryophytes. The numbers given on each branch were bootstrap values.




**Appendix Figure S4.** The phylogenetic tree of species for analyses, related to STAR Methods. 23 tree species (orange), 15 perennial herb species (blue), and 21 annual herb species (green) were included.




**Appendix Figure S5.** The phylogenetic tree of 24 *PARP* genes within gymnosperm species, related to STAR Methods. 24 *PARP* genes within gymnosperm species were divided into three distinct clades (*PARP1*, *PARP2*, and *PARP3*). The numbers given on each branch were bootstrap values.




**Appendix Figure S6.** The phylogenetic tree of 11 tree species for analyses of the relationship between the growth rate and the copy number ratio of *PARP*, related to STAR Methods, Figure 5 and Table 3. Four angiosperm and seven gymnosperm species were included. The numbers given on each branch were bootstrap values.

# Chapter 2: Analyses of gene copy number variation in diverse epigenetic regulatory gene families across plants: Increased copy numbers of *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) and *SILENCING DEFECTIVE 3* (*SDE3*) in long-lived trees

The study in this chapter, done in collaboration with Professor Akiko Satake, is under peer review.

**ABSTRACT**

Long-lived organisms experience high risk of damage due to the various types of stresses over their lifespans. Epigenetic regulation is involved in gene regulation, genome integrity, and inhibition of exogenous genetic elements, which are functions important for long-term survival. In the present study, to identify the epigenetic regulatory genes with increased copy number in long-lived tree species than in short-lived annual and perennial herb species, we conducted systematic comparisons of copy number variation in 121 gene families involved in various epigenetic regulatory pathways across 85 plant species with different lifespans using a genome database. Among these 121 gene families, the gene family encoding *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) and that encoding *SILENCING DEFECTIVE 3* (*SDE3*) were found to exhibit significantly higher copy number of genes in tree species than in both perennial and annual herb species. BRU1/TSK/MGO3 is involved in chromatin modifications and plays an important role in the maintenance of meristems, genome integrity, and the inheritance of chromatin states. SDE3 is involved in RNA silencing and has an important role in antiviral defense through posttranscriptional gene silencing. Increasing copy numbers of *BRU1/TSK/MGO3* and *SDE3* genes are likely to be favored in the maintenance of meristems, genome integrity, the inheritance of chromatin states, and antiviral defense in long-lived trees, and these factors could contribute to survival over a long lifespan.

**INTRODUCTION**

Organisms are exposed to many endogenous and exogenous stresses on a daily basis. Such stresses lead to damage at various levels (i.e., DNA, epigenetic state, protein, and cell). The accumulation of damage causes genomic and epigenomic instability, alteration

of gene expression, and cellular dysfunctions, resulting in disease and aging. Therefore, suppressing damage from stresses and maintaining homeostasis are required for long-lived organisms, such as trees that live for hundreds or thousands of years. Recently, a growing number of studies have shown that epigenetic regulation is involved in vital biological processes, such as the regulation of gene expression (Busslinger, 1983; Grunstein, 1997), DNA replication (Zhang et al., 2000), DNA repair (Shim et al., 2005), and the inhibition of exogenous genetic elements (Al-Kaff et al., 1998), which are important for maintaining homeostasis.

Multiple epigenetic regulatory pathways have evolved, such as those involving DNA modification, histone modification, chromatin formation and remodeling, and RNA-mediated gene silencing. DNA methylation regulates gene expression by recruiting proteins involved in gene repression or by inhibiting the binding of transcription factors to DNA (Moore et al., 2013). Loss of DNA methylation leads to activation of silenced DNA sequences, resulting in the activation of transposable elements and abnormal expression of genes (Pal & Tyler, 2016). Histone modifications are involved in the regulation of chromatin structure, activating or suppressing gene expression (Grunstein, 1997; Nakayama et al., 2001). Chromatin formation and remodeling are required for not only transcription processes but also other DNA processes, such as DNA repair (Shim et al., 2005; Chai et al., 2005), replication (Collins et al., 2002) and recombination (Fritsch et al., 2004), which are important biological processes. RNA silencing is involved in posttranscriptional gene silencing (PTGS) and transcriptional gene silencing (TGS), regulating the transcription level. Moreover, RNA silencing plays an important role in defense against viruses, microbial pathogens and transgenes (Al-Kaff et al., 1998; Ruiz-Ferrer & Voinnet, 2009). These major functions and pathways of epigenetic regulation

are highly conserved in eukaryotes (Almeida & Allshire, 2005; Fuchs et al., 2006; Lee et al., 2010; Marinov & Lynch, 2016), suggesting the universal importance of epigenetic regulation for survival of organisms.

Previously, studies on longevity have mainly focused on relationships between DNA repair and longevity (Hart & Setlow, 1974; Bürkle et al., 1994; Tian et al., 2019) because DNA repair plays an essential role in suppressing mutations due to DNA damage and maintaining genome integrity for long periods. Recently, a growing number of studies have focused on the relationships between epigenetic regulation and longevity because of the importance of epigenetic regulation in long-term genomic and epigenomic integrity (Pal & Tyler, 2016). Previous studies have investigated the effects of epigenetic regulation on longevity and identified genes related to longevity in model organisms. An example is the association of sirtuins, $NAD^+$-dependent histone deacetylases, with longevity. Sirtuins are involved in the regulation of many metabolic functions, including DNA repair, genome stability, inflammatory responses, apoptosis, the cell cycle, and mitochondrial functions (Wątroba & Szukiewicz, 2016). Overexpression or activation of Sir2 homologs extends the lifespan of worms (*Caenorhabditis elegans*) (Tissenbaum & Guarente, 2001) and fruit flies (*Drosophila melanogaster*) (Rogina & Helfand, 2004). Another example is the role in longevity and responses to environmental stresses of Dicer, which is involved in the regulation of RNA-mediated gene silencing. Dicer is an RNase III endoribonuclease and is required for the generation of microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Jinek & Doudna, 2009). Dicer is an important enzyme in the miRNA processing pathway, and its downregulation can result in the downregulation of many miRNAs, including miRNAs, which affect stress resistance and survival (Mori et al., 2012). In *C. elegans*, loss-of-function mutation of *Dicer* reduces

lifespan and stress resistance, while intestinal overexpression of *Dicer* confers stress resistance (Mori et al., 2012).

A growing number of studies have explored the functions and factors of epigenetic regulation in longevity; however, most subjects in these studies are model organisms with short lifespans (e.g., budding yeast, worms, fruit flies and mice). In particular, systematic comparisons of epigenetic regulation across species with different lifespans are not sufficiently represented. To identify the key factors and genes related to longevity and elucidate the relationship between epigenetic regulation and longevity, a comprehensive comparison is necessary across species of varying lifespan including long-lived species. Therefore, in the present study, we focus on plants, which include diverse species with a wide range of lifespans, from annual herbs with short lifespans less than one year to perennial herbs and trees with long lifespans.

To search for epigenetic regulatory genes related to tree longevity, we focused on copy number variation among species in epigenetic regulatory genes. Copy numbers of genes have changed due to gene duplication and loss. Increases in copy number via gene duplications can provide the opportunity for the evolution of phenotypic novelty and contribute to adaptive evolution (Flagel & Wendel, 2009). We have previously performed comprehensive comparative analyses of copy number variation in DNA repair gene families in plants and identified the *PARP* gene family as a unique gene family with higher copy numbers in long-lived tree species than in short-lived annual and perennial herb species, and this gene family plays important roles in DNA repair, transcription regulation, and antipathogen defense in plants as well as animals (Aoyagi Blue et al., 2021). Thus, for epigenetic regulatory genes, investigating gene families with increased copy numbers in trees through comprehensive comparison analyses of copy number

variation is effective in identifying candidate gene families that may play important roles in tree longevity.

For plant species, a growing number of studies in the model plant species *Arabidopsis thaliana* have elucidated the major epigenetic regulatory pathways and identified the genes involved in epigenetic regulation (Pikaard & Scheid, 2014). In addition, recent progress in sequencing provides genome sequence data of diverse non-model plant species, including annual and perennial herbs and trees in a wide range of taxa. In the present study, to identify the epigenetic regulatory genes with increased in tree species relative to annual and perennial herb species, we systematically compared the copy number variation of genes within 121 gene families involved in epigenetic regulation across 85 plant species, including trees, perennial herbs, annual herbs, and algae, using a genome database.

## MATERIALS AND METHODS

### Data collection and target species

We collected data on copy numbers of genes encoding proteins involved in epigenetic regulation in plant species from the Dicots PLAZA 5.0 database (Van Bel et al., 2022) (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/), which is a database of genomes of diverse plant species. This database contains information on 100 plant species, including bryophytes (*Anthoceros agrestis*, *Marchantia polymorpha* and *Physcomitrella patens*) and algae (*Chara braunii*, *Chlamydomonas reinhardtii*, *Micromonas commoda* and *Prasinoderma coloniale*), as an outgroup (Supplementary Table S1). Using the same method described in a previous study (Aoyagi Blue et al., 2021), we categorized each species included in the database into five groups according to

life form (algae, annual herbs, perennial herbs, shrubs, and trees) based on information from databases and the literature (Aoyagi Blue et al., 2021). Newly added species compared to the previous study were categorized based on other databases (eFloras [http://www.efloras.org/] and Solanaceae Source [https://solanaceaesource.myspecies.info/]) and the literature (Bisang, 2003; Kato et al., 2008; Yang et al., 2013; Borah & Ghosh, 2018; Mérai et al., 2019; Li et al., 2020; Dong et al., 2021) as well as the databases used in the previous study. We eliminated shrub species from the analyses because they have intermediate life forms, being tree-like but small (< 5 m), and have relatively shorter lifespans than trees. Thus, 85 species, including 21 tree species, 23 perennial herb species, 37 annual herb species, and four algal species, were used for our analyses (Table 1).

**Genes associated with epigenetic regulation for comparative analyses**

We selected 221 genes associated with epigenetic regulation within *Arabidopsis thaliana* based on the literature (Pikaard & Scheid, 2014; Kim, 2019) and categorized these genes into five functional groups (DNA modification, histone modification, chromatin formation or chromatin remodeling, Polycomb-group proteins and interacting components, RNA silencing) depending on the pathways described in the literature (Pikaard & Scheid, 2014; Kim, 2019) (Supplementary Table S2). Dicots PLAZA 5.0 clustered the genes into gene families by applying Tribe-MCL (Enright et al., 2002), and we used the gene families provided from the PLAZA database. The 221 epigenetic regulatory genes of *A. thaliana*, which we had selected for analyses, were grouped into 121 gene families in Dicots PLAZA 5.0. Then, we collected the data regarding copy numbers within each gene family for the species from the Dicots PLAZA 5.0 database.

**The normalized index of the copy number of genes for analysis**

Some plant species and lineages have experienced gene duplication events, including whole genome duplication (Bowers et al., 2003; Qiao et al., 2019). Species with high total numbers of genes would have high copy numbers of epigenetic regulatory genes due to gene duplication. Therefore, for the comparative analyses, we used the normalized ratio of the copy number of genes within a gene family in the focal species to the total number of genes in the species, named the "copy number ratio", instead of the actual copy numbers of genes. We calculated the copy number ratio of each gene family for a species in the same way as in a previous study (Aoyagi Blue et al., 2021).

**Construction of a phylogenetic species tree for analyses considering phylogenetic relationships**

Copy number ratios might not be statistically independent among species due to phylogenetic relationships. Thus, we need to consider phylogenetic relationships in the analysis. To adopt statistical methods that account for phylogenetic relationships of copy number ratios, we constructed a phylogenetic tree of species in the present study in the same way as in a previous study (Aoyagi Blue et al., 2021). The dataset consisted of 85 species and included four algal species, *C. braunii*, *C. reinhardtii*, *M. commoda* and *P. coloniale*, but we eliminated these algal species and one annual herbal species, *Sapria himalayana*, from the analyses considering the phylogenetic relationships because the sequence data of *rbcL* and/or *matK* to calculate branch lengths were not available for these species. Thus, we used the remaining 80 species for the construction of the

phylogenetic tree and the analyses accounting for phylogenetic relationships (Supplementary Fig. S1).

**Similarities in copy number ratio of 121 gene families among species and among gene families**

To assess similarities in copy number ratio of 121 gene families associated with epigenetic regulation among species and identify the species that generally have high copy number ratios for epigenetic regulatory gene families, we performed hierarchical clustering based on the Euclidian distance of the copy number ratio of each species using the Ward method. To test the enrichment or dilution of each life form in each of the significantly different clusters, Fisher exact tests (two-sided) were performed. Then, we controlled for the false discovery rate using the method of Storey's Q-value (Storey, 2002) and estimated the Q-value of each test using the qvalue package (ver. 2.16.0; Storey et al., 2015) in R. After the clustering analysis, we tested whether the species in each cluster had a higher or lower copy number ratio than the mean for all species. The mean copy number ratio of 121 gene families within each species was calculated. Then, we tested whether the average of the mean copy number ratio of 121 gene families within the species included in each cluster was significantly higher or lower than one (that is, the mean copy number ratio for all species) by the Wilcoxon signed rank test. After the Wilcoxon signed rank tests, we controlled for the false discovery rate and estimated the Q-value using the method explained above. Gene families were also clustered by hierarchical clustering based on the Euclidian distance of the copy number ratio of each gene family using the Ward method, and the enrichment or dilution of each gene

functional group in each of the significantly different clusters was tested using the method explained above.

To investigate whether the copy number of genes in a species was correlated with the total number of genes in a species, we evaluated the Spearman's correlation coefficient of copy number of genes and total number of genes for each gene family and tested the correlation coefficient is significantly different from zero. After the test, we controlled for the false discovery rate and estimated the Q-value using the method described above.

**Identifying the gene families with increased copy number ratios in trees**

To identify the gene families with increased copy number ratios in tree species, we compared the copy number ratio among life forms in each gene family by phylogenetic generalized least squares (PGLS) regression (Grafen, 1989). In each gene family, we performed PGLS regression with different phylogenetic models: a Brownian-motion model (Felsenstein, 1985), a Brownian-motion model with a trend, Pagel's lambda model (Pagel, 1999), Pagel's kappa model (Pagel, 1999), Pagel's delta model (Pagel, 1999), the Ornstein-Uhlenbech model (Hansen, 1997; Martins & Hansen, 1997), and the early burst model (Harmon et al., 2010). We examined the model fit across phylogenetic models based on Akaike's information criterion (AIC) value (Akaike, 1973) and selected the model with the lowest AIC value. We performed PGLS regressions and estimated values of the phylogenetic correlation parameter and the variance rate in the phylogenetic model using the phylolm package (ver. 2.6, Tung Ho & Ané, 2014) in R. After the PGLS analyses, we controlled for the false discovery rate and estimated the Q-value using the method described above.

**The phylogeny and domain structures of the *BRU1/TSK/MGO3* and *SDE3* gene families in plants**

Our analyses identified *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) and *SILENCING DEFECTIVE 3* (*SDE3*) gene families as unique gene families with increased copy number ratios in tree species (see the Results section). To investigate the evolutionary histories of these gene families in plant species, we constructed phylogenetic trees of genes within both gene families for the species included in the dataset. In addition, to assess the diversity in protein functions within the gene families among species, we compared domain structures across species in both gene families. To assess the phylogeny of genes and compare domain structures across species, we constructed phylogenetic trees with the protein domain structures in the *BRU1/TSK/MGO3* gene family and *SDE3* gene family using the tree explorer tool in Dicots PLAZA 5.0 (*BRU1/TSK/MGO3*, https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D005030; *SDE3*, https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D002863). There were 148 *BRU1/TSK/MGO3* genes in 96 species included in Dicots PLAZA 5.0, including shrubs (Supplementary Table S3). Of these 148 genes, 35 genes were removed from the construction of the phylogenetic tree by multiple sequence alignment because of low sequence similarity. Thus, the phylogenetic tree of *BRU1/TSK/MGO3* genes was constructed using 113 *BRU1/TSK/MGO3* genes from 94 species (Supplementary Table S3). There were 242 *SDE3* genes in 97 species included in Dicots PLAZA 5.0, including shrubs (Supplementary Table S4). Of these 242 genes, 32 genes were removed from the construction of the phylogenetic tree by multiple sequence

alignment. Thus, the phylogenetic tree of *SDE3* genes was constructed using 210 *SDE3* genes from 95 species (Supplementary Table S4).

To perform all statistical analyses, we used R ver. 3.4.1 (the R project, http://www.r-project.org/).

## RESULTS

### Interspecies comparison of copy number ratios of 121 epigenetic regulatory gene families

We performed hierarchical clustering based on the similarities in copy number ratio among species. Hierarchical clustering based on the Euclidian distance of the copy number ratio of each species using the Ward method showed that 85 species were divided into three clusters (Fig. 1A). Species cluster 1 consisted of two algal species, *Micromonas commoda* and *Prasinoderma coloniale,* revealing significant enrichment of algal species (Fisher exact test, Q-value = 0.0202) (Supplementary Table S5). Species cluster 2 consisted of 11 Brassicales species (10 species were members of Brassicaceae), including nine annual herb species and two perennial herb species, revealing significant enrichment of annual herb species (Fisher exact test, Q-value = 0.0338) (Supplementary Table S5). Species cluster 3 exhibited the greatest number of species, including 21 tree species, 21 perennial herb species, 28 annual herb species and two algal species, revealing no significant enrichment or dilution of a certain type of life form (Supplementary Table S5). The results of the clustering suggest that similarity in copy number ratios of 121 epigenetic regulatory gene families depends on phylogenetic relationships.

*M. commoda* and *P. coloniale,* species included in cluster 1, showed clear contrast between high and low copy number ratios among gene family clusters. The copy

number ratios of most gene families included in gene family cluster I were high, whereas the copy number ratios of most gene families in gene family cluster II were low or zero in both species. Such a contrast of high and low copy number ratios among gene family clusters was also observed in an alga, *Chlamydomonas reinhardtii*, in species cluster 3. In species cluster 1, the average of the mean copy number ratio of 121 gene families was not significantly different from the mean for all species (Wilcoxon signed-rank test, Q-value = 0.500) (Fig. 1B). The species in species cluster 2 exhibited very high copy number ratios in one gene family, which encodes SWI-SNF-related chromatin-binding proteins. The average of the mean copy number ratios of 121 gene families was not significantly different from the mean of all species (Wilcoxon signed-rank test, Q-value = 0.325) (Fig. 1B). Species cluster 3 included species with high copy number ratios for most gene families (e.g., *Trochodendron aralioides* [tree species], *Ceratophyllum demersum* [perennial herb species] and *Cardamine hirsute* [annual herb species]) and species with low copy number ratios for most gene families (e.g., *Eucalyptus grandis* [tree species], *Salvia bowleyana* [perennial herb species] and *Sapria himalayana* [annual herb species]). Therefore, the average of the mean copy number ratios of 121 gene families in species cluster 3 varied from low to high. The average of the mean copy number ratios of 121 gene families was not significantly different from the mean for all species (Wilcoxon signed-rank test, Q-value = 0.325) (Fig. 1B).

We also performed hierarchical clustering of 121 gene families to assess the similarities in copy number ratio among gene families. As the result of hierarchical clustering based on the Euclidian distance of the copy number ratio of each gene family using the Ward method, a total of 121 gene families were divided into two major clusters and one independent gene family (Fig. 1A). There were 43 gene families in gene family

cluster I, including five DNA modification gene families, 13 histone modification gene families, 13 chromatin formation gene families, four Polycomb-group protein gene families and eight RNA silencing gene families. There were 77 gene families in gene family cluster II, including ten DNA modification gene families, 15 histone modification gene families, 20 chromatin formation gene families, nine Polycomb-group protein gene families and 23 RNA silencing gene families. Only one gene family, encoding SWI-SNF-related chromatin-binding proteins, was outside of the clusters. Only 15 of the species have genes in this gene family. Twelve of the 15 species were annual and perennial herb species in Brassicales. The others were three tree species, *Theobroma cacao, Durio zibethinus* and *Quercus lobata.* In addition, the actual copy numbers in this gene family were greater within species in Brassicales than in others. These results suggest that this gene family encoding SWI-SNF-related chromatin-binding proteins has expanded in Brassicales due to gene duplications. Fisher exact tests showed no significant differences in enrichment or dilution of any type of gene function among clusters (Supplemental Table S6).

The mean actual copy number of each gene family across species was less than five, and variance among species was low in most of the gene families (Supplementary Fig. S2). However, in several gene families, the mean and variance of actual copy numbers was extremely large. For example, in the gene family encoding NAC domain-containing proteins and the gene family encoding ubiquitin-conjugating enzyme (UBC, E2) proteins, the means of the actual copy numbers were 91.28 and 43.87, and the standard deviations of the actual copy numbers were 57.62 and 26.01, respectively (Supplemental Fig. S2). There were intermediate and very low phylogenetic signals in actual copy number within these gene families (Pagel's lambda was 0.354 for the NAC

domain-containing protein gene family and $6.55\times10^{-8}$ for the UBC gene family). In addition, there was no significant relationship between actual copy number and life form. This suggests that copy number variation in these gene families is independent of phylogeny and life form. Conversely, these gene families showed strong positive correlations between actual copy number and total number of genes in a species (Spearman's rank correlation coefficients were 0.588 and 0.638, and Q-values for correlation tests were $1.86\times10^{-8}$ and $1.21\times10^{-9}$ in the NAC domain-containing protein and UBC gene families, respectively). This suggests that these gene families with large mean copy numbers and high variances increased with gene expansion due to gene duplication.

**Identifying the gene families with increased copy number ratios in tree species**

Next, to identify the gene families with increased copy number ratios in tree species, we compared copy number ratios among tree species, perennial herb species, and annual herb species using phylogenetic generalized least squares (PGLS) regressions. Among the 121 gene families, two gene families showed copy number ratios significantly higher in tree species than in both perennial and annual herb species: the gene family encoding *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) (Fig. 2A) and the gene family encoding *SILENCING DEFECTIVE 3* (*SDE3*) (Fig. 3A) (Table 2). BRU1/TSK/MGO3 is associated with chromatin formation and remodeling and is involved in DNA damage repair, the maintenance of chromatin state and the regulation of meristem development (Suzuki et al., 2004; Takeda et al., 2004; Suzuki et al., 2005; Ohno et al., 2011). Three tree species, *Trochodendron aralioides* (wheel tree), *Sequoiadendron giganteum* (giant sequoia) and *Carya illinoinensis* (pecan), exhibited the highest copy number ratios for the

*BRU1/TSK/MGO3* gene family (Fig. 2B). *T. aralioides* and *S. giganteum* also featured the largest actual copy numbers for the *BRU1/TSK/MGO3* gene family (Fig. 2C). SDE3 is an RNA helicase and is involved in posttranscriptional gene silencing and defense against viruses (Dalmay et al., 2001). Two tree species, *Citrus clementina* (orange) and *Quercus lobata* (valley oak), and one perennial herb species, *Lonicera japonica* (Japanese honeysuckle), exhibited the highest copy number ratio as well as the greatest actual copy number in the *SDE3* gene family (Fig. 3B and 3C).

In PGLS analyses, we examined model fit across phylogenetic models and selected a model for each gene family. Among 121 gene families, the Ornstein-Uhlenbech (OU) model was selected for 67 gene families, Pagel's lambda model was selected for 45 gene families, Pagel's kappa model was selected for four gene families, Pagel's delta model was selected for four gene families, and the early burst model was selected for one gene family (Supplementary Table S7). OU models were selected for both the *BRU1/TSK/MGO3* and *SDE3* gene families. In the *BRU1/TSK/MGO3* gene family, the estimated value of the phylogenetic correlation parameter $\alpha$ in the OU model was 46.00, and the estimated value of the variance rate $\sigma^2$ in the OU model was 36.35 (Table 2). In the *SDE3* gene family, the estimated value of the phylogenetic correlation parameter $\alpha$ in the OU model was 7.18, and the estimated value of the variance rate $\sigma^2$ in the OU model was 9.74 (Table 2).

One gymnosperm species, *Sequoiadendron giganteum* (giant sequoia), included in the dataset, had extraordinary long maximum lifespan and a large genome size compared to other angiosperm tree species and showed the highest copy number ratio and the actual copy number of *BRU1/TSK/MGO3* genes (Fig. 2B and C). To assess whether *S. giganteum* strongly affected the result, we performed PGLS analysis on dataset with

only angiosperm species, removing *S. giganteum*, *Selaginella moellendorffii* (a lycophyte species), and three bryophyte species (*Marchantia polymorpha*, *Physcomitrium patens* and *Anthoceros agrestis*). As the result, copy number ratios of *BRU1/TSK/MGO3* gene family and that of *SDE3* gene family were significantly higher in tree species than both in annual and perennial herb species (Supplementary Table S8 and Supplementary Figure S3). This result strongly suggests that copy number ratio of *BRU1/TSK/MGO3* gene family and that of *SDE3* gene family were significantly high in tree species.

**The evolutionary histories and diversity of the *BRU1/TSK/MGO3* and *SDE3* gene families in plants**

We identified *BRU1/TSK/MGO3* and *SDE3* gene families as unique gene families with increased copy number ratios in trees. To investigate the evolutionary histories of these gene families in plant species, we constructed a phylogenetic tree of genes with protein domain structures for each gene family using the tree explore tool in Dicots PLAZA 5.0 (*BRU1/TSK/MGO3*, https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D005030; *SDE3*, https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D002863). In the *BRU1/TSK/MGO3* gene family, there were 148 genes in 96 species, and the phylogenetic tree of genes was constructed using 113 genes from 94 species, including shrub species (Fig. 4). Most land plant species, including angiosperms as well as a gymnosperm species (*S. giganteum*), a lycophyte species (*Selaginella moellendorffii*), and several bryophyte species (*Anthoceros agrestis*, *Marchantia polymorpha* and *Physcomitrium patens*), had *BRU1/TSK/MGO3* gene(s).

Among algal species, *Chara braunii* had three *BRU1/TSK/MGO3* genes, but other algal species, *Chlamydomonas reinhardtii*, *Micromonas commoda* and *Prasinoderma coloniale*, had no *BRU1/TSK/MGO3* genes. BRU1/TSK/MGO3 domain structures were similar across species. The domains of BRU1/TSK/MGO3 mainly consisted of two domains: a tetratricopeptide repeat (TPR) domain at the N-terminal part and a leucine-rich repeat (LRR) domain at the C-terminal part. *Arabidopsis* BRU1/TSK/MGO3 possesses leucine-glycine-asparagine (LGN) repeat domains, which are classified as a subfamily of the TPR motif, and LRR domains (Suzuki et al., 2004). The TPR domain is involved in protein–protein interactions (Blatch & Lässel, 1999). The LRR domain is also involved in protein–protein and protein–ligand interactions (Matsushima & Miyashita, 2012). Although the numbers of *BRU1/TSK/MGO3* genes varied among species, the domain structure of BRU1/TSK/MGO3 was highly conserved among species (Fig. 4). This suggests that the function of BRU1/TSK/MGO3 is similar among species. However, in species with large numbers of *BRU1/TSK/MGO3* genes, such as *T. aralioides* and *S. giganteum*, some *BRU1/TSK/MGO3* genes shared low sequence similarity with other *BRU1/TSK/MGO3* genes, which were removed from construction of the phylogenetic tree. This suggests that such *BRU1/TSK/MGO3* genes would have different functions than other *BRU1/TSK/MGO3* genes.

In the *SDE3* gene family, there were 242 genes in 97 species, and the phylogenetic tree of genes was constructed using 210 genes from 95 species, including shrub species (Fig. 5 and Supplementary Fig. S4). Genes within the *SDE3* gene family were divided into two major clades based on the sequence similarities across genes (Fig. 5 and Supplementary Fig. S4). Clade 1 included 76 genes within 35 species, most of which were tree and shrub species. SDE3s in clade 1 consisted of two main domains: a

DNA2/NAM7-like helicase domain at the N-terminal part and the P-loop containing a nucleoside triphosphate hydrolase (NTPase) domain at the C-terminal part. Clade 2 included 122 genes within 84 species. SDE3s in clade 2 consisted of two main types of domains: a P-loop containing NTPase at the N-terminal part and a DNA2/NAM7 helicase domain and a DNA2/NAM7 helicase-like domain at the C-terminal part. The DNA2/NAM7 helicase domain and DNA2/NAM7 helicase-like domain are found in DNA2 and NAM7 proteins, which are involved in ATP-dependent RNA helicase activity (Kang et al., 2000; Plank & Wilkinson, 2018). The P-loop containing the NTPase domain is involved in catalyzing the hydrolysis of the β-γ phosphate bond of a bound nucleoside triphosphate (Leipe et al., 2004). This suggests that proteins in different clades would have basically common functions in terms of the domain organization, although these proteins are divided into different clades based on sequence similarity. Genes outside of clusters included 12 genes within seven species, one gymnosperm species (*S. giganteum*), one lycophyte species (*S. moellendorffii*), two bryophyte species (*M. polymorpha* and *P. patens*) and three algal species (*C. braunii*, *C. reinhardtii* and *P. coloniale*), which were the species in the earliest plant lineages (Fig. 5 and Supplementary Fig. S4). Although sequences of genes outside of clusters differed slightly from those of other genes within angiosperms, the domains consisted proteins of genes outside of clusters were conserved.

## DISCUSSION

To identify the epigenetic regulatory genes with increased copy number in tree species compared to annual and perennial herb species, we conducted systematic comparative analyses of copy number variation in 121 gene families involved in epigenetic regulation among 85 plant species with a broad range of lifespans from annual herbs with short

lifespans to perennial herbs and trees with long lifespans. Among the 121 gene families studied here, two gene families, *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) gene family and *SILENCING DEFECTIVE 3* (*SDE3*) gene family, were found to exhibit significant expansion of copy number in tree species compared to both perennial herb species and annual herb species. BRU1/TSK/MGO3 plays important roles in the maintenance of meristems and normal morphogenesis, genome integrity, and the inheritance of chromatin states. SDE3 has an important role in antiviral defense through posttranscriptional gene silencing. Our results suggest that BRU1/TSK/MGO3 and SDE3 would play important roles in tree longevity through these processes.

**Increased copy number of *BRU1/TSK/MGO3* genes in the maintenance of meristems, long-term genome integrity, and the inheritance of chromatin states**

BRU1/TSK/MGO3 is required for the maintenance of meristems and normal morphogenesis in plants. In *A thaliana,* structural and functional disorganization of meristems, including the shoot apical meristem (SAM) and the root apical meristem (RAM), and alterations in morphogenesis are observed in the *mgo3* and *tsk* mutants (Guyomarc'h et al., 2004; Suzuki et al., 2004). The sequence of tetratricopeptide repeat (TPR) domains in BRU1/TSK/MGO3 is similar to the leucin-glycine-asparagine (LGN) repeat motif in animal proteins (Guyomarc'h et al., 2004; Suzuki et al., 2004). The LGN-related protein in *Drosophila melanogaster*, Partner of Inscuteable (Pins), is involved in asymmetric cell division (Yu et al., 2000), and the Pins homolog in humans also plays a key role in asymmetric cell division (Parmentier et al., 2000). These results suggest that BRU1/TSK/MGO3 is important in the control of meristematic cell division and morphogenesis and the maintenance of meristem activity (Guyomarc'h et al., 2004;

Suzuki et al., 2004). In addition, BRU1/TSK/MGO3 plays an important role in genome maintenance. In *Arabidopsis thaliana*, *bru1* mutants are highly sensitive to genotoxic stress (Takeda et al., 2004). BRU1/TSK/MGO3 proteins are localized in the nucleus (Suzuki et al., 2004; Takeda et al., 2004), and the *BRU1/TSK/MGO3* gene is expressed in S-phase of the cell cycle (Suzuki et al., 2005). Therefore, BRU1/TSK/MGO3 is involved in an S-phase DNA damage checkpoint and postreplicative DNA repair in plants (Takeda et al., 2004). Animals have homologs of plant BRU1/TSK/MGO3, TONSOK-like (TONSL) (Ray et al., 1995; O'Donnell et al., 2010). TONSL interacts with methyl methanesulfonate-sensitivity protein 22-like (MMS22-L) and is required for the repair of DNA double-strand breaks by homologous recombination repair in human cells (Duro et al., 2010; O'Donnell et al., 2010; Piwko et al., 2011). Thus, BRU1/TSK/MGO3 and its homologs have important roles in DNA repair and long-term genome integrity. Another important function of BRU1/TSK/MGO3 is the inheritance of chromatin states and gene regulation. BRU1/TSK/MGO3 is involved not only in the inheritance of euchromatin states (Ohno et al., 2011; Ohno et al., 2014) but also in the inheritance of heterochromatin states (Takeda et al., 2004) and is required for the regulation of genes, such as *FLOWERING LOCUS C* (*FLC*), a key regulator of flowering (Guyomarc'h 2006), and genes associated with heat shock memory (Brzezinka et al., 2018). Therefore, BRU1/TSK/MGO3 plays an important role in the maintenance of meristems and morphogenesis, genome integrity, and the inheritance of chromatin states.

The maintenance of meristems and morphogenesis, genome integrity, and the inheritance of chromatin states are required for longevity. This is because stem cells in meristems provide persistent growth and development, DNA repair suppresses

mutations due to DNA damage and maintains genome integrity, and the inheritance of chromatin states is required not only for gene expression but also for DNA repair (Shim et al., 2005; Chai et al., 2005), DNA replication (Collins et al., 2002) and recombination (Fritsch et al., 2004). Our results showed that copy number ratios of the *BRU1/TSK/MGO3* gene family were high in tree species, especially in long-lived tree species (Fig. 2B). *Sequoiadendron giganteum* (giant sequoia) can live for more than 3000 years (Harvey, 1986), and *Carya illinoinensis* (pecan tree) can live for over 300 years (Smith, 1950; Brison, 1974). Increases copy number of genes via gene duplications can provide the opportunity for the evolution of phenotypic novelty and contribute to adaptive evolution (Flagel & Wendel, 2009; Weng et al., 2012; Huang et al., 2021). Our results suggests that an increased copy number of *BRU1/TSK/MGO3* genes in long-lived tree species play an important role in the maintenance of meristems and normal morphogenesis and long-term genome and epigenome integrity and are likely to favor tree longevity.

**Increased copy number of *SDE3* genes and antiviral defense**

Another gene family showing a significantly higher copy number ratio in tree species than in perennial herb species and annual herb species was the *SDE3* gene family (Fig. 3). *SDE3*s are members of the RNA helicase superfamily SF1 (Linder & Owttrim, 2009) and play a key role in antiviral defense (Dalmay et al., 2001; Garcia et al., 2012) through RNA-mediated posttranscriptional gene silencing (PTGS). In plant PTGS, SDE3 is likely required to enhance the production of double-stranded RNA from limiting amounts of transgenic or viral RNA templates by RNA-dependent RNA polymerase (RDR6) (Garcia et al., 2012). Moreover, SDE3 proteins are predicted to be

localized in the cytoplasm (Linder & Owttrim, 2009), whereas most other RNA helicases are predicted to be localized in the nucleus, and SDE3 is required for short- and long-distance cell-to-cell movement of PTGS in plants (Himber et al., 2003). *SDE3* homologs are also found in animals. *Armitage* (*Armi*), the *Drosophila SDE3* homolog, is required for RNA interference (RNAi) in *Drosophila melanogaster* (Cook et al., 2004; Tomari et al., 2004). Moloney leukemia virus protein 10 (MOV10), the *SDE3* homolog in mammals, is involved in the inhibition of the movement of transposable elements (Arjan-Odedra et al., 2012; Li et al., 2013) and the replication of retroviruses (Burdick et al., 2010; Wang et al., 2010). Therefore, SDE3s play an important role in antiviral defense in plants and animals.

Because viruses commonly infect wild plants (MacClement & Ricenterds, 1956; Raybould et al., 1999; Tugume et al., 2008) and long-lived tree species are likely to be more exposed to the risk of viral infections than are short-lived herb species, resistance to viruses is important for tree species to survive for a long time. The present study showed that the copy number ratio of the *SDE3* gene family was significantly higher in tree species than in annual and perennial herb species (Fig. 3B). The species with the highest copy number ratio of *SDE3*, *Quercus lobata* (valley oak) can live for over 350 years at its maximum lifespan (Jepson, 1910; Elias, 1980), and *Citrus clementina* (orange) lives for more than 50 years on average (LEAF Network Linking Edible Arizona Forests; https://leafnetworkaz.org/). This suggests that an increased copy number of *SDE3* genes in tree species would favor antiviral defense and longevity.

Ecological studies report that slow-growing trees tend to live longer than rapid-growing trees (Johnson & Abrams, 2009; Black et al., 2008). One of the reasons for this phenomenon is that slow-growing trees invest more energy and resources for defense

against herbivory and pathogens than for growth processes, such as photosynthesis, resulting in slow growth and long lifespans (Loehle, 1988). Indeed, a negative correlation between defense and growth is generally found in plants, and molecular factors and pathways related to the trade-off between defense and growth have been reported (Campos et al., 2016; Cui et al., 2020). The increased copy number of SDE3 in tree species would contribute to improving antiviral defense through PTGS but would indirectly affect slow growth via the trade-off between defense and growth.

**Limitations of the study and future directions**

The Present study showed that the copy numbers of *BRU1* and *SDE3* genes were significantly expanded in tree species compared with annual and perennial herb species. The present study is still limited to showing the correlational relationship between the copy number ratio and the life form of plant species. Further studies are necessary to investigate the causal role of BRU1/TSK/MGO3 and SDE3 in tree longevity. Because detailed functions of BRU1/TSK/MGO3s and SDE3s in plants, particularly perennial plant species, remain poorly understood, additional studies on non-model perennial plants as well as model plants are required. Experimental studies, such as phenotype analysis and physiological analysis on mutants and transcriptome analysis among plant species with different lifespans, will be able to shed light on the role of BRU1/TSK/MGO3 and SDE3 on plant longevity. There was only one gymnosperm species, *S. giganteum* (giant sequoia), in the dataset studied. Some gymnosperm species are known to have extraordinarily long maximum lifespans, e.g., *Pseudotsuga menziesii* (Douglas fir) and *Pinus sylvestris* (Scots pine), which are known to be able to live over 1000 years (Franklin & Dyrness, 1973). To elucidate the relationship between epigenetic regulation and tree

longevity, comparative analyses across species including gymnosperms as well as angiosperms is necessary. Thanks to the advancement of sequencing and gene annotation, comparative analyses of a large number of species, including gymnosperm and angiosperm species, will be able to be performed to identify new gene families that have important functions in long-lived tree species. Comprehensive genome data analyses will be able to reveal the pivotal processes and systems of epigenetic regulation in plant longevity.

**Conclusions**

Overall, systematic comparative analyses of copy number variation in gene families associated with various epigenetic regulatory pathways across diverse plant species revealed significantly increased copy numbers of genes of *BRU1/TSK/MGO3* and *SDE3* gene families in tree species. BRU1/TSK/MGO3 has an important role in the maintenance of meristems and normal morphogenesis, genome integrity, and the inheritance of chromatin states. SDE3 plays an important role in antiviral defense through posttranscriptional gene silencing. Our results suggest that the maintenance of meristems, genome integrity, inheritance of chromatin states, and antiviral defense would contribute to survival for a long time under the risks of damage due to stresses in plants. The present study can stimulate research to elucidate the functions and roles of BRU1/TSK/MGO3 and SDE3 in tree longevity, leading to an understanding of the relationships between epigenetic regulation and longevity.

**AKNOWLEDGMENTS**

We would like to thank Dr. Eriko Sasaki and Dr. Junko Kusumi for helpful discussions and comments on the present study. This study was funded by JSPS KAKENHI (JP26251042; JP17H06478) to A.S.

**REFERENCES**

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory (eds. B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest. (Reproduced in Breakthroughs in Statistics, Vol.1, Foundations and Basic Theory (eds. S. Kotz and N.L. Johnson), Springer-Verlag, New York, (1992).)

Al-Kaff, N.S., Covey, S.N., Kreike, M.M., Page, A.M., Pinder, R., Dale, P.J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science*, 279, 2113 2115.

Almeida, R., & Allshire, R. C. (2005). RNA silencing and genome regulation. *Trends in Cell Biology*, 15(5), 251–258. https://doi.org/10.1016/j.tcb.2005.03.006

Aoyagi Blue, Y., Kusumi, J., & Satake, A. (2021). Copy number analyses of DNA repair genes reveal the role of poly(ADP-ribose) polymerase (PARP) in tree longevity. *iScience*, 24(7), 102779. https://doi.org/https://doi.org/10.1016/j.isci.2021.102779

Arjan-Odedra, S., Swanson, C. M., Sherer, N. M., Wolinsky, S. M., & Malim, M. H. (2012). Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology*, 9, 1–13. https://doi.org/10.1186/1742-4690-9-53

Bisang, I. (2003). Population development, demographic structure, and life cycle aspects of two hornworts in Switzerland. *Lindbergia*, 28(3), 105–112. https://doi.org/10.2307/20150137

Black, B. A., Colbert, J. J., & Pederson, N. (2008). Relationships between radial growth rates and lifespan within North American tree species. *Écoscience*, 15(3), 349–357. https://doi.org/10.2980/15-3-3149

Blatch, G. L., & Lässle, M. (1999). The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays*, 21(11), 932–939. https://doi.org/10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N

Borah, D., & Ghosh, D. (2018). Sapria Himalayana. *Resonance*, 23(4), 479–489. https://doi.org/10.1007/s12045-018-0637-8

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422(6930), 433–438. https://doi.org/10.1038/nature01521

Brison, F.R. (1974). Pecan culture. Capital Printing, Austin, Tex.Cabeza de Vaca, A.N. 1983. Cabeza de Vaca's adventures in the unknowninterior of America. C. Covey (ed.). Univ. of New Mexico Press, Albuquerque.

Brzezinka, K., Altmann, S., & Bäurle, I. (2018). BRUSHY1/TONSOKU/MGOUN3 is required for heat stress memory. *Plant Cell and Environment*, 42(3), 771–781. https://doi.org/10.1111/pce.13365

Bürkle, A., Müller, M., Wolf, I., & Küpper, J. H. (1994). Poly(ADP-ribose) polymerase activity in intact or permeabilized leukocytes from mammalian species of different longevity. *Molecular and Cellular Biochemistry*, 138(1–2), 85–90. https://doi.org/10.1007/BF00928447

Burdick, R., Smith, J. L., Chaipan, C., Friew, Y., Chen, J., Venkatachari, N. J., et al. (2010). P Body-Associated Protein Mov10 Inhibits HIV-1 Replication at Multiple Stages. *Journal of Virology*, 84(19), 10241–10253. https://doi.org/10.1128/jvi.00585-10

Busslinger, M., Hurst, J., & Flavell, R. A. (1983). DNA methylation and the regulation of globin gene expression. *Cell*, 34(1), 197–206. https://doi.org/https://doi.org/10.1016/0092-8674(83)90150-2

Campos, M. L., Yoshida, Y., Major, I. T., De Oliveira Ferreira, D., Weraduwage, S. M., Froehlich, J. E., et al. (2016). Rewiring of jasmonate and phytochrome B signalling

uncouples plant growth-defense tradeoffs. *Nature Communications*, 7. https://doi.org/10.1038/ncomms12570

Chai, B., Huang, J., Cairns, B. R., & Laurent, B. C. (2005). Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes and Development*, 19(14), 1656–1661. https://doi.org/10.1101/gad.1273105

Collins, N., Poot, R. A., Kukimoto, I., García-Jiménez, C., Dellaire, G., & Varga-Weisz, P. D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nature Genetics*, 32(4), 627–632. https://doi.org/10.1038/ng1046

Cook, H. A., Koppetsch, B. S., Wu, J., & Theurkauf, W. E. (2004). The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell*, 116(6), 817–829. https://doi.org/10.1016/S0092-8674(04)00250-8

Cui, C., Wang, J. J., Zhao, J. H., Fang, Y. Y., He, X. F., Guo, H. S., et al. (2020). A Brassica miRNA Regulates Plant Growth and Immunity through Distinct Modes of Action. *Molecular Plant*, 13(2), 231–245. https://doi.org/10.1016/j.molp.2019.11.010

Dalmay, T., Horsefield, R., Braunstein, T. H., & Baulcombe, D. C. (2001). SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. *EMBO Journal*, 20(8), 2069–2077. https://doi.org/10.1093/emboj/20.8.2069

Dong, S., Liu, M., Liu, Y., Chen, F., Yang, T., Chen, L., et al. (2021). The genome of Magnolia biondii Pamp. provides insights into the evolution of Magnoliales and biosynthesis of terpenoids. *Horticulture Research*, 8(1). https://doi.org/10.1038/s41438-021-00471-9

Duro, E., Lundin, C., Ask, K., Sanchez-Pulido, L., MacArtney, T. J., Toth, R., et al. (2010). Identification of the MMS22L-TONSL Complex that promotes homologous recombination. *Molecular Cell*, 40(4), 632–644. https://doi.org/10.1016/j.molcel.2010.10.023

Elias, Thomas S. (1980). The complete trees of North America, field guide and natural history. New York. Van Nostrand Reinhold Company.

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584

Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist*, 125, 1–15.

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, 183(3), 557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Franklin, J. F. and Dyrness, C. T. (1973) Natural vegetation of Oregon and Washington. USDA For. Serv, USDA Forest Service - General Technical Report GTR-PNW-8.

Fritsch, O., Benvenuto, G., Bowler, C., Molinier, J., & Hohn, B. (2004). The INO80 protein controls homologous recombination in Arabidopsis thaliana. *Molecular Cell*, 16(3), 479–485. https://doi.org/10.1016/j.molcel.2004.09.034

Fuchs, J., Demidov, D., Houben, A., & Schubert, I. (2006). Chromosomal histone modification patterns - from conservation to diversity. *Trends in Plant Science*, 11(4), 199–208. https://doi.org/10.1016/j.tplants.2006.02.008

Garcia, D., Garcia, S., Pontier, D., Marchais, A., Renou, J. P., Lagrange, T., et al. (2012). Ago Hook and RNA Helicase Motifs Underpin Dual Roles for SDE3 in Antiviral Defense and Silencing of Nonconserved Intergenic Regions. *Molecular Cell*, 48(1), 109–120. https://doi.org/10.1016/j.molcel.2012.07.028

Grafen, A. (1989) The phylogenetic regression., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: 10.1098/rstb.1989.0106.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649), 349–352. https://doi.org/10.1038/38664

Guyomarc'h, S., Vernoux, T., Traas, J., Zhou, D. X., & Delarue, M. (2004). MGOUN3, an Arabidopsis gene with TetratricoPeptide-Repeat-related motifs, regulates meristem

cellular organization. *Journal of Experimental Botany*, 55(397), 673–684. https://doi.org/10.1093/jxb/erh069

Guyomarc'h, S., Benhamed, M., Lemonnier, G., Renou, J. P., Zhou, D. X., & Delarue, M. (2006). MGOUN3: Evidence for chromatin-mediated regulation of FLC expression. *Journal of Experimental Botany*, 57(9), 2111–2119. https://doi.org/10.1093/jxb/erj169

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5), 1341–1351. https://doi.org/10.1111/j.1558-5646.1997.tb01457.x

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, 64(8), 2385–2396. https://doi.org/10.1111/j.1558-5646.2010.01025.x

Hart, R. W., & Setlow, R. B. (1974). Correlation between deoxyribonucleic acid excision repair and life span in a number of mammalian species. *Proceedings of the National Academy of Sciences of the United States of America*, 71(6), 2169–2173. https://doi.org/10.1073/pnas.71.6.2169

Harvey, H. T. (1986). Evolution and History of Giant Sequoia. 1–3. https://www.fs.usda.gov/treesearch/pubs/27502

Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., & Voinnet, O. (2003). movement of RNA silencing. *The EMBO Journal*, 22(17), 4523–4533.

Huang, Y., Chen, J., Dong, C., Sosa, D., Xia, S., Ouyang, Y., et al. (2021). Species-specific partial gene duplication in Arabidopsis thaliana evolved novel phenotypic effects on morphological traits under strong positive selection. *The Plant Cell*, koab291. https://doi.org/10.1093/plcell/koab291

Jepson, Willis L. (1910). The silva of California. University of California Memoirs 2:1-480.

Jinek, M., & Doudna, J. A. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature*, 457(7228), 405–412. https://doi.org/10.1038/nature07755

Johnson, S. E., & Abrams, M. D. (2009). Age class, longevity and growth rate relationships: Protracted growth increases in old trees in the eastern United States. *Tree Physiology*, 29(11), 1317–1328. https://doi.org/10.1093/treephys/tpp068

Kang, H. Y., Choi, E., Bae, S. H., Lee, K. H., Gim, B. S., Kim, H. D., et al. (2000). Genetic analyses of schizosaccharomyces pombe dna2+ reveal that DNA2 plays an essential role in Okazaki fragment metabolism. *Genetics*, 155(3), 1055–1067. https://doi.org/10.1093/genetics/155.3.1055

Kato, S., Sakayama, H., Sano, S., Kasai, F., Watanabe, M. M., Tanaka, J., et al. (2008). Morphological variation and intraspecific phylogeny of the ubiquitous species Chara braunii (Charales, Charophyceae) in Japan. *Phycologia*, 47(2), 191–202. https://doi.org/10.2216/07-27.1

Kim, J. H. (2019) Chromatin remodeling and epigenetic regulation in plant DNA damage repair, *International Journal of Molecular Sciences*, 20(17). doi: 10.3390/ijms20174093.

Lee, T. F., Zhai, J., & Meyers, B. C. (2010). Conservation and divergence in eukaryotic DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9027–9028. https://doi.org/10.1073/pnas.1005440107

Leipe, D. D., Koonin, E. V., & Aravind, L. (2004). STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *Journal of Molecular Biology*, 343(1), 1–28. https://doi.org/10.1016/j.jmb.2004.08.023

Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., et al. (2013). The MOV10 helicase inhibits LINE-1 mobility. *Journal of Biological Chemistry*, 288(29), 21148–21160. https://doi.org/10.1074/jbc.M113.465856

Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., et al. (2020). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants.

*Nature Ecology and Evolution*, 4(9), 1220–1231. https://doi.org/10.1038/s41559-020-1221-7

Linder, P., & Owttrim, G. W. (2009). Plant RNA helicases: linking aberrant and silencing RNA. *Trends in Plant Science*, 14(6), 344–352. https://doi.org/10.1016/j.tplants.2009.03.007

Loehle, C. (1988) Tree life history strategies: the role of defenses. *Canadian Journal of Forest Research*. 18(2): 209-222. https://doi.org/10.1139/x88-032

MacClement, W. D., & Richards, M. G. (1956). VIRUS IN WILD PLANTS. *Canadian Journal of Botany*, 34(5), 793–799. https://doi.org/10.1139/b56-060

Marinov, G. K., & Lynch, M. (2016). Conservation and divergence of the histone code in nucleomorphs. *Biology Direct*, 11(1), 1–11. https://doi.org/10.1186/s13062-016-0119-4

Martins, E. P., & Hansen, T. F. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, 149(4), 646–667. https://doi.org/10.1086/286013

Matsushima, N., & Miyashita, H. (2012). Leucine-rich repeat (LRR) domains containing intervening motifs in plants. *Biomolecules*, 2(2), 288–311. https://doi.org/10.3390/biom2020288

Mérai, Z., Graeber, K., Wilhelmsson, P., Ullrich, K. K., Arshad, W., Grosche, C., et al. (2019). Aethionema arabicum: a novel model plant to study the light control of seed germination. *Journal of Experimental Botany*, 70(12), 3313–3328. https://doi.org/10.1093/jxb/erz146

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1), 23–38. https://doi.org/10.1038/npp.2012.112

Mori, M. A., Raghavan, P., Thomou, T., Boucher, J., Robida-Stubbs, S., MacOtela, Y., et al. (2012). Role of microRNA processing in adipose tissue in stress defense and longevity. *Cell Metabolism*, 16(3), 336–347. https://doi.org/10.1016/j.cmet.2012.07.017

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., & Grewal, S. I. S. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, 292(5514), 110–113. https://doi.org/10.1126/science.1060118

O'Donnell, L., Panier, S., Wildenhain, J., Tkach, J. M., Al-Hakim, A., Landry, M. C., et al. (2010). The MMS22L-TONSL complex mediates recovery from replication stress and homologous recombination. *Molecular Cell*, 40(4), 619–631. https://doi.org/10.1016/j.molcel.2010.10.024

Ohno, Y., Narangajavana, J., Yamamoto, A., Hattori, T., Kagaya, Y., Paszkowski, J., et al. (2011). Ectopic gene expression and organogenesis in Arabidopsis mutants missing

BRU1 required for genome maintenance. *Genetics*, 189(1), 83–95. https://doi.org/10.1534/genetics.111.130062

Ohno, Y., Nishimura, T., Hattori, T., & Takeda, S. (2014). BRU1 Maintains Configuration of the Euchromatic Subchromosomal Domain in the Nucleus of Arabidopsis, *Plant Molecular Biology Reporter*, 32(1), 19–27. https://doi.org/10.1007/s11105-013-0596-x

Pagel, M. (1999). Inferring historical patterns of biological evolution. *Nature*, 401(October 1999), 877–884.

Pal, S., & Tyler, J. K. (2016). Epigenetics and aging. *Science advances*, 2(7), e1600584. https://doi.org/10.1126/sciadv.1600584

Parmentier, M. L., Woods, D., Greig, S., Phan, P. G., Radovic, A., Bryant, P., et al. (2000). Rapsynoid/partner of inscuteable controls asymmetric division of larval neuroblasts in Drosophila. *The Journal of Neuroscience* : The Official Journal of the Society for Neuroscience, 20(14), 1–5. https://doi.org/10.1523/jneurosci.20-14-j0003.2000

Pikaard, C. S. & Scheid, O. M. (2014) Epigenetic regulation in plants, *Cold Spring Harbor Perspectives in Biology*, 6(12), pp. 1–32. doi: 10.1101/cshperspect.a019315.

Piwko, W., Buser, R., & Peter, M. (2011). Rescuing stalled replication forks: MMS22L-TONSL, a novel complex for DNA replication fork repair in human cells. *Cell Cycle*, 10(11), 1703–1705. https://doi.org/10.4161/cc.10.11.15557

M. Plank, T. D., & Wilkinson, M. F. (2018). RNA Decay Factor UPF1 Promotes Protein Decay: A Hidden Talent. *BioEssays*, 40(1), 1–6. https://doi.org/10.1002/bies.201700170

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology*, 20(1), 1–23. https://doi.org/10.1186/s13059-019-1650-2

Ray, P., Zhang, D. H., Elias, J. A., & Ray, A. (1995). Cloning of a differentially expressed IκB-related protein. *Journal of Biological Chemistry*, 270(18), 10680–10685. https://doi.org/10.1074/jbc.270.18.10680

Raybould, A. F., Maskell, L. C., Edwards, M. L., Cooper, J. I., & Gray, A. J. (1999). The prevalence and spatial distribution of viruses in natural populations of Brassica oleracea. *New Phytologist*, 141(2), 265–275. https://doi.org/10.1046/j.1469-8137.1999.00339.x

Rogina, B., & Helfand, S. L. (2004). Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), 15998–16003. https://doi.org/10.1073/pnas.0404184101

Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology*, 60, 485–510. https://doi.org/10.1146/annurev.arplant.043008.092111

Shim, E. Y., Ma, J.-L., Oum, J.-H., Yanez, Y., & Lee, S. E. (2005). The Yeast Chromatin Remodeler RSC Complex Facilitates End Joining Repair of DNA Double-Strand Breaks. *Molecular and Cellular Biology*, 25(10), 3934–3944. https://doi.org/10.1128/mcb.25.10.3934-3944.2005

Smith, J.R. 1950. Tree crops. Devin-Adair, New York.

Storey, J. D. (2002) A direct approach to false discovery rates, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(3), pp. 479–498. doi: 10.1111/1467-9868.00346.

Storey, J.D., Bass, A.J., Dabney, A., Robinson, D. (2015) qvalue: Q-value estimation for false discovery rate control. R Package. http://github.com/jdstorey/qvalue

Suzuki, T., Inagaki, S., Nakajima, S., Akashi, T., Ohto, M. A., Kobayashi, M., et al. (2004). A novel Arabidopsis gene Tonsoku is required for proper cell arrangement in root and shoot apical meristems, *Plant Journal*, 38(4), 673–684. https://doi.org/10.1111/j.1365-313X.2004.02074.x

Suzuki, T., Nakajima, S., Inagaki, S., Hirano-Nakakita, M., Matsuoka, K., Demura, T., et al. (2005). TONSOKU is expressed in S phase of the cell cycle and its defect delays cell cycle progression in Arabidopsis, *Plant and Cell Physiology*, 46(5), 736–742. https://doi.org/10.1093/pcp/pci082

Takeda, S., Tadele, Z., Hofmann, I., Probst, A. V., Angelis, K. J., Kaya, H., et al. (2004). BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis, *Genes and Development*, 18(7), 782–793. https://doi.org/10.1101/gad.295404

Tian, X., Firsanov, D., Zhang, Z., Gladyshev, V. N., Seluanov, A., Gorbunova, V., et al. (2019). SIRT6 Is Responsible for More Efficient DNA Double- Strand Break Repair in Long-Lived Species Article SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. *Cell*, 177(3), 622-638.e22. https://doi.org/10.1016/j.cell.2019.03.043

Tissenbaum, H. A., & Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. *Nature*, 410(6825), 227–230. https://doi.org/10.1038/35065638

Tomari, Y., Du, T., Haley, B., Schwarz, D. S., Bennett, R., Cook, H. A., et al. (2004). RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell*, 116(6), 831–841. https://doi.org/10.1016/S0092-8674(04)00218-1

Tugume, A. K., Mukasa, S. B., & Valkonen, J. P. T. (2008). Natural wild hosts of Sweet potato feathery mottle virus show spatial differences in virus incidence and virus-like diseases in Uganda. *Phytopathology*, 98(6), 640–652. https://doi.org/10.1094/PHYTO-98-6-0640

Tung Ho, L. S. & Ané, C. (2014) A linear-time algorithm for gaussian and non-gaussian trait evolution models, *Systematic Biology*, 63(3), pp. 397–408. doi: 10.1093/sysbio/syu005.

Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, L., Botzki, A., Coppens, F., et al. (2022). PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Research*, 50(D1), D1468–D1474. https://doi.org/10.1093/nar/gkab1024

Wang, X., Han, Y., Dang, Y., Fu, W., Zhou, T., Ptak, R. G., et al. (2010). Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *Journal of Biological Chemistry*, 285(19), 14346–14355. https://doi.org/10.1074/jbc.M110.109314

Wątroba, M., & Szukiewicz, D. (2016). The role of sirtuins in aging and age-related diseases. *Advances in Medical Sciences*, 61(1), 52–62. https://doi.org/10.1016/j.advms.2015.09.003

Weng JK, Li Y, Mo H, & Chapple C. (2012). Assembly of an Evolutionarily New Pathway for α-Pyrone Biosynthesis in Arabidopsis. *Science*, 337(6097), 960–964. https://doi.org/10.1126/science.1221614

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, et al. (2013). The reference genome of the halophytic plant Eutrema salsugineum. *Frontiers in Plant Science*, 4(MAR), 1–14. https://doi.org/10.3389/fpls.2013.00046

Yu, F., Morin, X., Cai, Y., Yang, X., & Chia, W. (2000). Analysis of partner of inscuteable, a novel player of Drosophila asymmetric divisions, reveals two distinct steps in inscuteable apical localization. *Cell*, 100(4), 399–409. https://doi.org/10.1016/S0092-8674(00)80676-5

Zhang, Z., Shibahara, K. I., & Stillman, B. (2000). PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature*, 408(6809), 221–225. https://doi.org/10.1038/35041601

## TABLES

**Table 1**. Species list for the analyses. There were 85 plant species including 21 tree species (A), 23 perennial herb species (B), 37 annual herb species (C), and four algal species (D) in the dataset. Four algal species and one annual herb species, *Sapria himalayana*, were eliminated from PGLS analyses.

| (A) Tree | (B) Perennial herb | (C) Annual herb | (D) Alga |
|---|---|---|---|
| *Acer truncatum* | *Aquilegia oxysepala* | *Aethionema arabicum* | *Chara braunii* |
| *Amborella trichopoda* | *Arabidopsis lyrata* | *Amaranthus hybridus* | *Chlamydomonas reinhardtii* |
| *Avicennia marina* | *Brassica oleracea* | *Anthoceros agrestis* | *Micromonas commoda* |
| *Carica papaya* | *Capsicum annuum* | *Arabidopsis thaliana* | *Prasinoderma coloniale* |
| *Carpinus fangiana* | *Ceratophyllum demersum* | *Arachis hypogaea* | |
| *Carya illinoinensis* | *Erythranthe guttata* | *Beta vulgaris* | |
| *Citrus clementina* | *Fragaria vesca* | *Brassica carinata* | |
| *Coffea canephora* | *Fragaria* x *ananassa* | *Brassica napus* | |
| *Davidia involucrata* | *Lonicera japonica* | *Brassica rapa* | |
| *Durio zibethinus* | *Lotus japonicus* | *Cannabis sativa* | |
| *Eucalyptus grandis* | *Marchantia polymorpha* | *Capsella rubella* | |
| *Magnolia biondii* | *Nelumbo nucifera* | *Cardamine hirsuta* | |
| *Malus domestica* | *Nicotiana tabacum* | *Chenopodium quinoa* | |
| *Olea europaea* | *Oryza sativa* ssp. *japonica* | *Cicer arietinum* L. | |
| *Populus trichocarpa* | *Salvia bowleyana* | *Citrullus lanatus* | |
| *Prunus persica* | *Sechium edule* | *Corchorus olitorius* | |
| *Punica granatum* | *Selaginella moellendorffii* | *Cucumis melo* | |
| *Quercus lobata* | *Solanum lycopersicum* | *Cucumis sativus* L. | |
| *Sequoiadendron giganteum* | *Solanum pennellii* | *Daucus carota* | |
| *Theobroma cacao* | *Solanum tuberosum* | *Erigeron canadensis* | |
| *Trochodendron aralioides* | *Trifolium pratense* | *Eutrema salsugineum* | |
| | *Utricularia gibba* | *Glycine max* | |
| | *Vanilla planifolia* | *Helianthus annuus* | |
| | | *Lactuca sativa* | |
| | | *Lupinus albus* | |
| | | *Medicago truncatula* | |
| | | *Papaver somniferum* | |

| | |
|---|---|
| | *Petunia axillaris* |
| | *Phaseolus vulgaris* |
| | *Physcomitrium patens* |
| | *Pisum sativum* |
| | *Sapria himalayana* |
| | *Schrenkiella parvula* |
| | *Striga asiatica* |
| | *Tarenaya hassleriana* |
| | *Vigna mungo* |
| | *Zea mays* |

**Table 2.** The result of phylogenetic generalized least squares (PGLS) regressions to compare the copy number ratios among life forms. The Ornstein-Uhlenbech (OU) models were selected for the *BRU1/TSK/MGO3* and *SDE3* gene families based on AIC values. a: The estimated value of the phylogenetic correlation parameter α in the OU model. b: The estimated value of the variance rate $\sigma^2$ in the OU model.

| | Trees vs. Annual herbs | | | | Trees vs. Perennial herbs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Symbol of gene family | Coefficient | Standard Error | *t*-value | Q-value | Coefficient | Standard Error | *t*-value | Q-value | Parameter | |
| *BRU1/TSK/MGO3* | –0.636 | 0.175 | –3.63 | 0.0394 | –0.690 | 0.188 | –3.67 | 0.0430 | 46.00 a | 36.35 b |
| *SDE3* | –0.821 | 0.232 | –3.54 | 0.0394 | –0.818 | 0.234 | –3.49 | 0.0430 | 7.18 a | 9.74 b |

## FIGURES

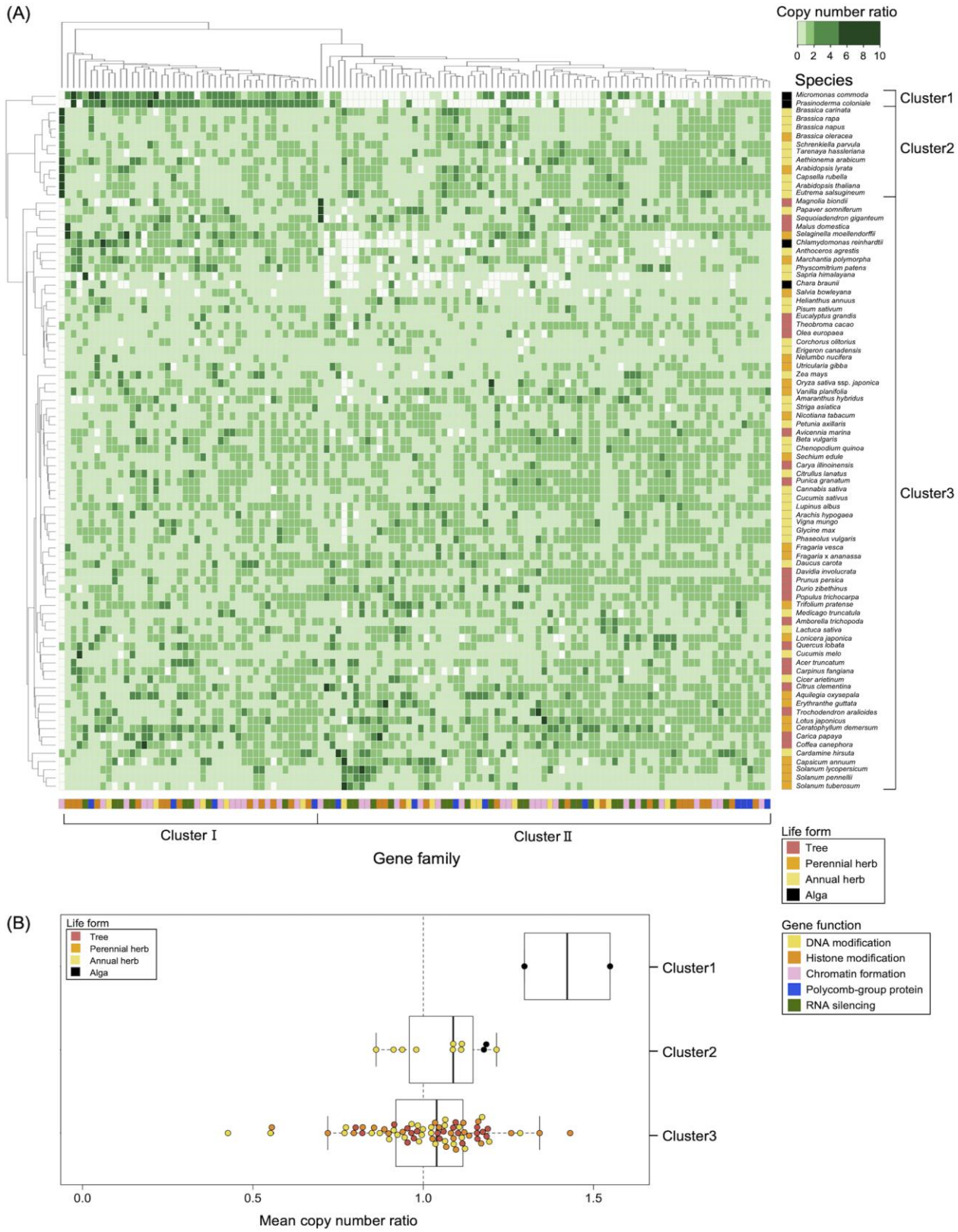

(A)
Copy number ratio
0 2 4 6 8 10
Species
Micromonas commoda
Prasinoderma coloniale
Brassica carinata
Brassica rapa
Brassica napus
Brassica oleracea
Schrenkiella parvula
Tarenaya hassleriana
Aethionema arabicum
Arabidopsis lyrata
Capsella rubella
Arabidopsis thaliana
Eutrema salsugineum
Magnolia biondii
Papaver somniferum
Sequoiadendron giganteum
Malus domestica
Selaginella moellendorffii
Chlamydomonas reinhardtii
Anthoceros agrestis
Marchantia polymorpha
Physcomitrium patens
Sapria himalayana
Chara braunii
Salvia bowleyana
Helianthus annuus
Pisum sativum
Eucalyptus grandis
Theobroma cacao
Olea europaea
Corchorus olitorius
Erigeron canadensis
Nelumbo nucifera
Utricularia gibba
Zea mays
Oryza sativa ssp. japonica
Vanilla planifolia
Amaranthus hybridus
Striga asiatica
Nicotiana tabacum
Petunia axillaris
Avicennia marina
Beta vulgaris
Chenopodium quinoa
Sechium edule
Carya illinoinensis
Citrullus lanatus
Punica granatum
Cannabis sativa
Cucumis sativus
Lupinus albus
Arachis hypogaea
Vigna mungo
Glycine max
Phaseolus vulgaris
Fragaria vesca
Fragaria x ananassa
Daucus carota
Davidia involucrata
Prunus persica
Durio zibethinus
Populus trichocarpa
Trifolium pratense
Medicago truncatula
Amborella trichopoda
Lactuca sativa
Lonicera japonica
Quercus ilex
Cucumis melo
Acer truncatum
Carpinus fangiana
Cicer arietinum
Citrus clementina
Aquilegia oxysepala
Erythranthe guttata
Trochodendron aralioides
Lotus japonicus
Ceratophyllum demersum
Carica papaya
Coffea canephora
Cardamine hirsuta
Capsicum annuum
Solanum lycopersicum
Solanum pennellii
Solanum tuberosum
Cluster1
Cluster2
Cluster3
Cluster I
Cluster II
Gene family
Life form
Tree
Perennial herb
Annual herb
Alga
(B)
Life form
Tree
Perennial herb
Annual herb
Alga
Cluster1
Cluster2
Cluster3
0.0
0.5
1.0
1.5
Mean copy number ratio
Gene function
DNA modification
Histone modification
Chromatin formation
Polycomb-group protein
RNA silencing

**Figure 1**. Interspecies comparisons of the copy number ratios of 121 epigenetic regulatory gene families. **(A)** Clustered heatmap of the copy number ratios of 121 epigenetic regulatory gene families. Hierarchical clustering was performed based on the Euclidean distance of the copy number ratio using the Ward's method. There were 85 plant species, including 21 tree species, 23 perennial herb species, 37 annual herb species, and four algal species. Each gene family was categorized into one of five functional groups: DNA modification, Histone modification, Chromatin formation, Polycomb-group proteins; or RNA silencing. **(B)** Mean copy number ratios of 121 epigenetic regulatory gene families for species in each cluster. The color of each point corresponds to the life form of the species. The horizontal line inside each box shows the median, and the length of the box shows the interquartile range (range between the 25th and 75th percentiles). The whiskers indicate points within 1.5 times the interquartile range. The points beyond the whisker range indicate the outliers.

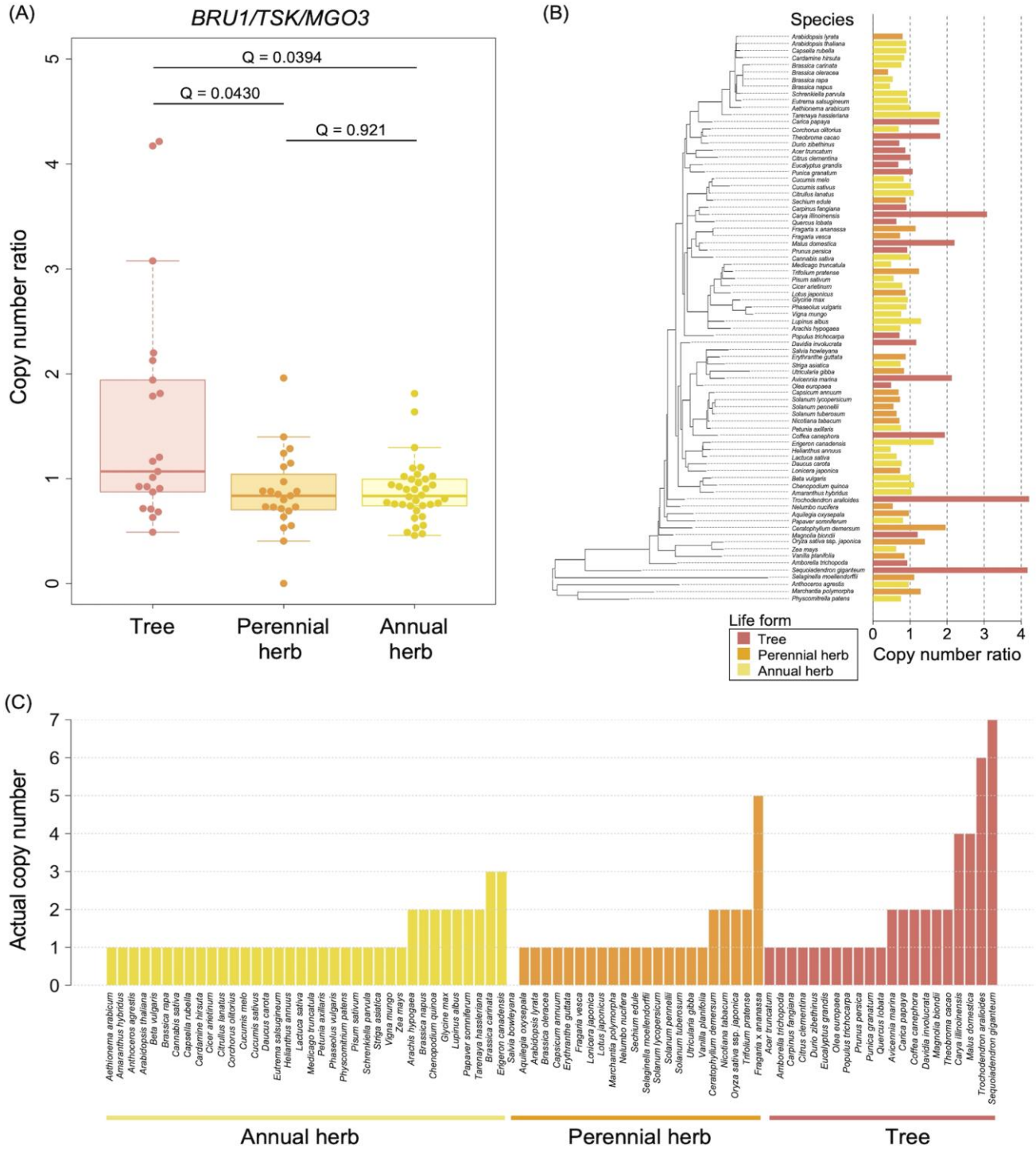



**Figure 2**. Results of the phylogenetic generalized least squares (PGLS) analysis with the Ornstein-Uhlenbech (OU) models for the *BRU1/TSK/MGO3* gene family. **(A)**

The copy number ratio of the *BRU1/TSK/MGO3* gene family in different life forms. The *BRU1/TSK/MGO3* gene family showed a significantly higher copy number ratio in tree species than in both perennial and annual herb species. The horizontal line inside each box shows the median, and the length of the box shows the interquartile range (range between the 25th and 75th percentiles). The whiskers indicate points within 1.5 times the interquartile range. The points beyond the whisker range indicate the outliers. **(B)** Phylogenetic relationships of copy number ratio of the *BRU1/TSK/MGO3* gene family. The color of each bar indicates the life form of the species. **(C)** The actual copy number of *BRU1/TSK/MGO3* genes within the gene family for a species.

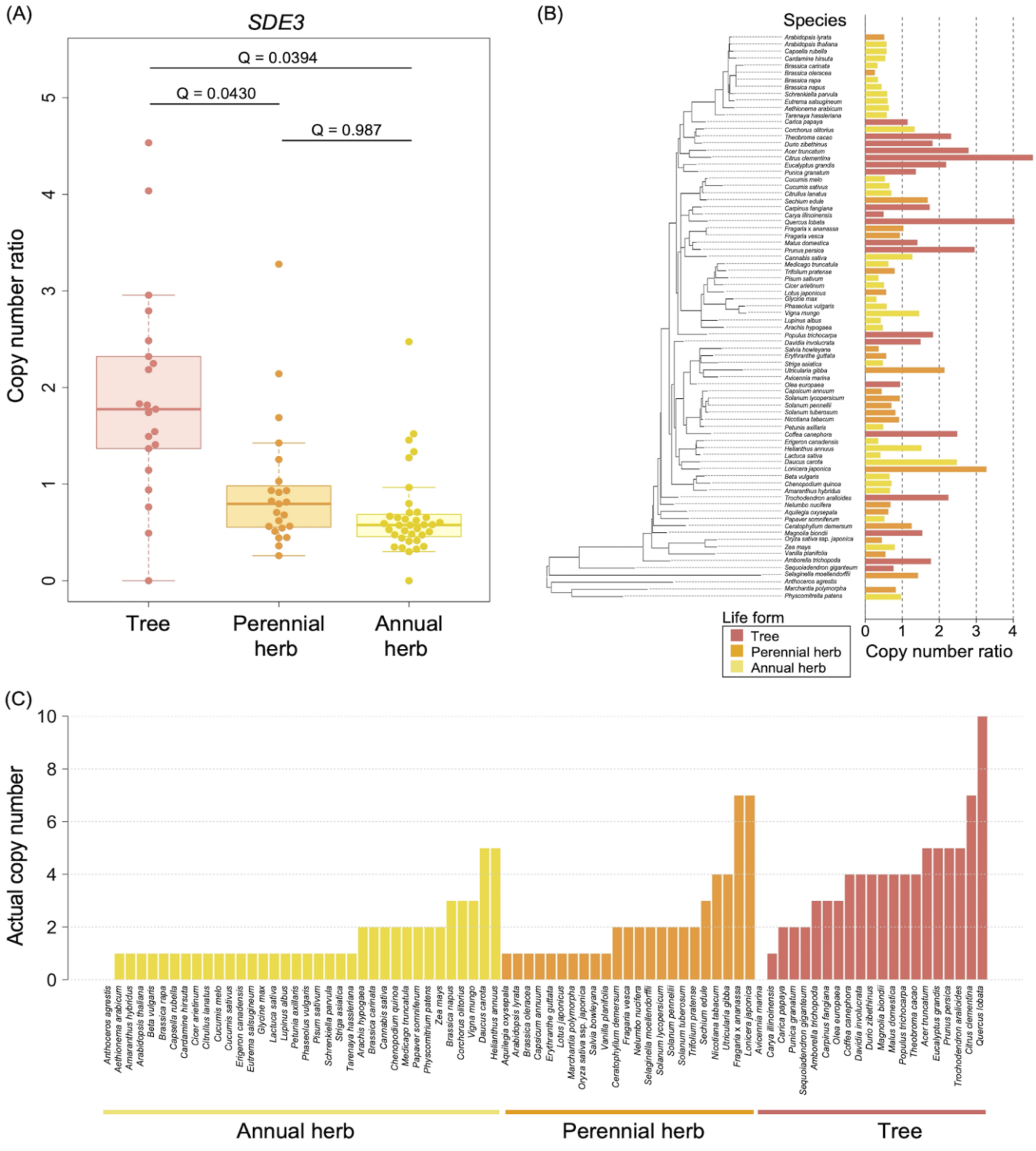



**Figure 3**. Results of the phylogenetic generalized least squares (PGLS) analysis with the Ornstein-Uhlenbech (OU) models for the *SDE3* gene family. **(A)** The copy

number ratios of the *SDE3* gene family in different life forms. The *SDE3* gene family showed a significantly higher copy number ratio in tree species than in both perennial and annual herb species. The horizontal line inside the box shows the median, and the length of the box shows the interquartile range (range between the 25th and 75th percentiles). The whiskers indicate points within 1.5 times the interquartile range. The points beyond the whisker range indicate the outliers. **(B)** Phylogenetic relationships in the copy number ratio of the *SDE3* gene family. The color of each bar indicates the life form of the species. **(C)** The actual copy number of *SDE3* genes within the gene family for the species.

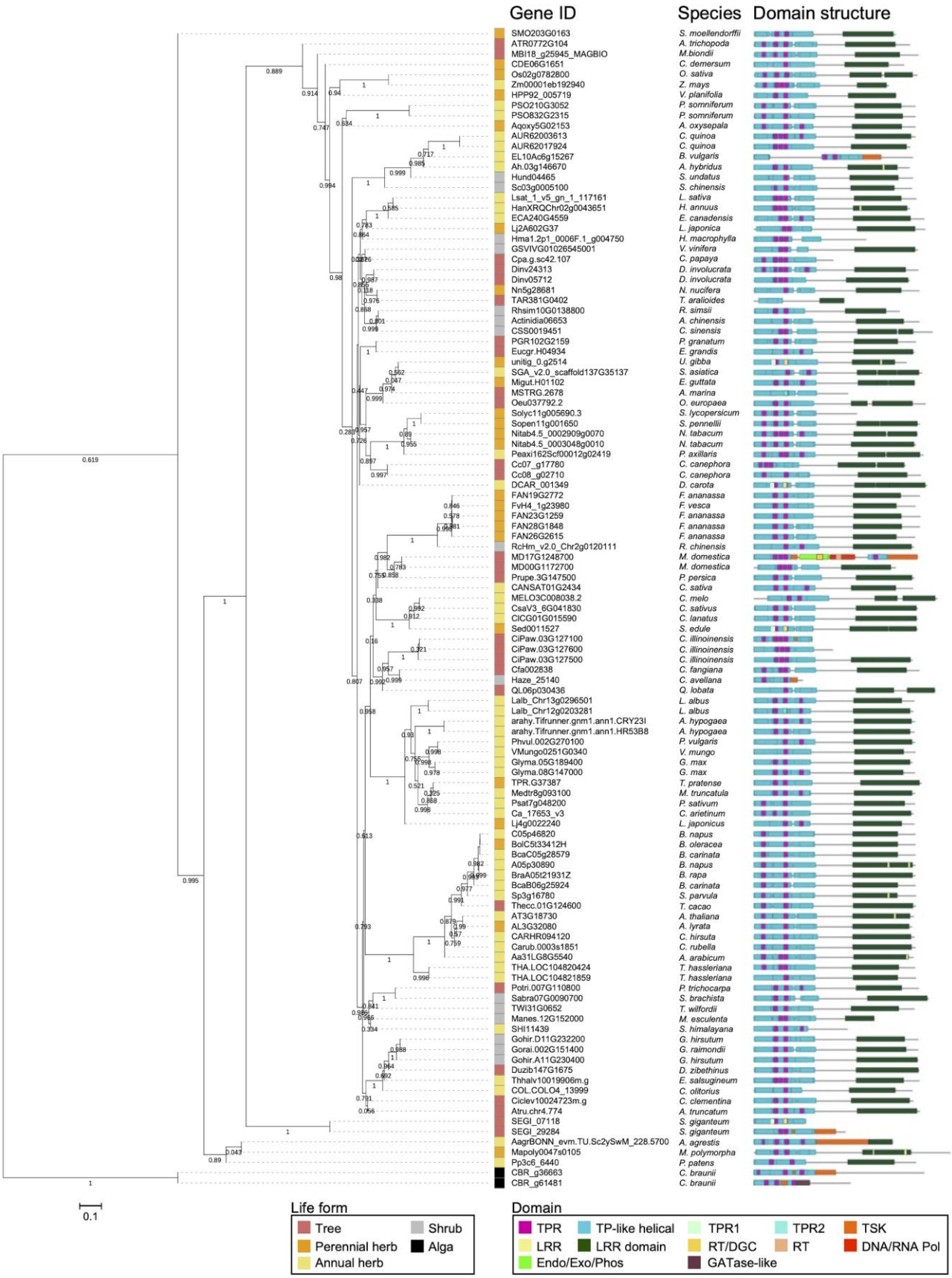

Gene ID
Species
Domain structure
SMO203G0163
ATR0772G104
MBI18_g25945_MAGBIO
CDE06G1651
Os02g0782800
Zm00001eb192940
HPP92_005719
PSO210G3052
PSO832G2315
Aqoxy5G02153
AUR62003613
AUR62017924
EL10Ac6g15267
Ah.03g146670
Hund04465
Sc03g0005100
Lsat_1_v5_gn_1_117161
HanXRQChr02g0043651
ECA240G4559
Lj2A602G37
Hma1.2p1_0006F.1_g004750
GSVIVG01026545001
Cpa.g.sc42.107
Dinv24313
Dinv05712
Nn5g28681
TAR381G0402
Rhsim10G0138800
Actinidia06653
CSS0019451
PGR102G2159
Eucgr.H04934
unlllg_0.g2514
SGA_v2.0_scaffold137G35137
Migut.H01102
MSTRG.2678
Oeu037792.2
Solyc11g005690.3
Sopen11g001650
Nitab4.5_0002909g0070
Nitab4.5_0003048g0010
Peaxi162Scf00012g02419
Cc07_g17780
Cc08_g02710
DCAR_001349
FAN19G2772
FvH4_1g23980
FAN23G1259
FAN28G1848
FAN26G2615
RcHm_v2.0_Chr2g0120111
MD17G1248700
MD00G1172700
Prupe.3G147500
CANSAT01G2434
MELO3C008038.2
CsaV3_6G041830
ClCG01G015590
Sed0011527
ClPaw.03G127100
ClPaw.03G127600
ClPaw.03G127500
Cfa002838
Haze_25140
QL06p030436
Lalb_Chr13g0296501
Lalb_Chr12g0203281
arahy.Tifrunner.gnm1.ann1.CRY23I
arahy.Tifrunner.gnm1.ann1.HR53B8
Phvul.002G270100
VMungo0251G0340
Glyma.05G189400
Glyma.08G147000
TPR.G37387
Medtr8g093100
Psat7g048200
Ca_17653_v3
Lj4g0022240
C05p46820
BolC5t33412H
BcaC05g28579
A05p30890
BraA05t21931Z
BcaB06g25924
Sp3g16780
Thecc.01G124600
AT3G18730
AL3G32080
CARHR094120
Carub.0003s1851
Aa31LG8G5540
THA.LOC104820424
THA.LOC104821859
Potri.007G110800
Sabra07G0090700
TWI31G0652
Manes.12G152000
SHI11439
Gohir.D11G232200
Gorai.002G151400
Gohir.A11G230400
Duzib147G1675
Thhalv10019906m.g
COL.COLO4_13999
Ciclev10024723m.g
Atru.chr4.774
SEGI_07118
SEGI_29284
AagrBONN_evm.TU.Sc2ySwM_228.5700
Mapoly0047s0105
Pp3c6_6440
CBR_g36663
CBR_g61481
S. moellendorffii
A. trichopoda
M.biondii
C. demersum
O. sativa
Z. mays
V. planifolia
P. somniferum
P. somniferum
A. oxysepala
C. quinoa
C. quinoa
B. vulgaris
A. hybridus
S. undatus
S. chinensis
L. sativa
H. annuus
E. canadensis
L. japonica
H. macrophylla
V. vinifera
C. papaya
D. involucrata
D. involucrata
N. nucifera
T. aralioides
R. simsii
A. chinensis
C. sinensis
P. granatum
E. grandis
U. gibba
S. asiatica
E. guttata
A. marina
O. europaea
S. lycopersicum
S. pennellii
N. tabacum
N. tabacum
P. axillaris
C. canephora
C. canephora
D. carota
F. ananassa
F. vesca
F. ananassa
F. ananassa
F. ananassa
R. chinensis
M. domestica
M. domestica
P. persica
C. sativa
C. melo
C. sativus
C. lanatus
S. edule
C. illinoinensis
C. illinoinensis
C. illinoinensis
C. fangiana
C. avellana
Q. lobata
L. albus
L. albus
A. hypogaea
A. hypogaea
P. vulgaris
V. mungo
G. max
G. max
T. pratense
M. truncatula
P. sativum
C. arietinum
L. japonicus
B. napus
B. oleracea
B. carinata
B. napus
B. rapa
B. carinata
S. parvula
T. cacao
A. thaliana
A. lyrata
C. hirsuta
C. rubella
A. arabicum
T. hassleriana
T. hassleriana
P. trichocarpa
S. brachista
T. wilfordii
M. esculenta
S. himalayana
G. hirsutum
G. raimondii
G. hirsutum
D. zibethinus
E. salsugineum
C. olitorius
C. clementina
A. truncatum
S. giganteum
S. giganteum
A. agrestis
M. polymorpha
P. patens
C. braunii
C. braunii
0.1
Life form
Tree
Shrub
Perennial herb
Alga
Annual herb
Domain
TPR
TP-like helical
TPR1
TPR2
TSK
LRR
LRR domain
RT/DGC
RT
DNA/RNA Pol
Endo/Exo/Phos
GATase-like

**Figure 4**. The phylogenetic tree of *BRU1/TSK/MGO3* genes with protein domain structures constructed using the tree explorer tool in Dicots PLAZA 5.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D005030). There were 113 genes within 94 species in the phylogenetic tree. Gene ID of each *BRU1/TSK/MGO3* gene in Dicots PLAZA 5.0 are represented. Species names indicate the species that have the gene, and rectangles to the left of species names indicate the life forms of the species. The numbers under each branch of the phylogenetic tree indicate support values. Protein domains are illustrated by color: TPR, tetratricopeptide repeat; TP-like helical, tetratricopeptide-like helical domain superfamily; TPR1, tetratricopeptide repeat 1; TPR2, tetratricopeptide repeat 2; TSK, TONSOKU; LRR, leucine-rich repeat; LRR domain, leucine-rich repeat domain superfamily; RT/DGC, reverse transcriptase/diguanylate cyclase domain; RT, reverse transcriptase domain; DNA/RNA Pol, DNA/RNA polymerase superfamily; Endo/Exo/Phos, endonuclease/exonuclease/phosphatase superfamily; ANAPC5, anaphase-promoting complex subunit 5 domain; GATase-like, class I glutamine amidotransferase-like.


(A)
Clade1
Gene ID
Species
Domain structure
(B)
Clade2
Gene ID
Species
Domain structure
(C)
Out of clades
Gene ID
Species
Domain structure
Life form
Tree
Shrub
Perennial herb
Alga
Annual herb
Domain
DNA2/NAM7 helic
DNA2/NAM7 helic-like C
DNA2/NAM7-like
Helic SF1/SF2
Helic MOV10
P-loop NTPase
ATPase
PUA-like
RNase H
ZnF C2H2
Ig-like fold

**Figure 5.** The protein domain structures of SDE3 in plants. There were 210 genes within 95 species, including genes within shrub species. Genes were divided into two clades and those outside of clades (Supplementary Figure S4). Seventy-six genes within 35 species were included in clade 1 **(A)**, 122 genes within 84 species were included in clade 2 **(B)**, and 12 genes within seven species were outside of clades **(C)**. Gene ID of each *SDE3* gene in Dicots PLAZA 5.0 are represented. Species names indicate the species that have the gene, and rectangles to the left of species names indicate the life forms of the species. Protein domains are illustrated by colors: DNA2/NAM7 helic, DNA2/NAM7 helicase domain; DNA2/NAM7 helic-like C, DNA2/NAM7 helicase-like at the C-terminal; DNA2/NAM7-like, DNA2/NAM7-like helicase; Helic SF1/SF2, Helicase superfamily 1/2, ATP-binding domain; Helic MOV10, Helicase MOV-10; P-loop NTPase, P-loop containing nucleoside triphosphate hydrolase; ATPase, AAA+ ATPase domain; PUA-like, PUA-like superfamily; RNase H, Ribonuclease H domain; ZnF C2H2, Zinc finger C2H2-type; Ig-like fold, Immunoglobulin-like fold.

## APPENDIXES

**Appendix Table S1.** The species list in Dicots PLAZA 5.0. There were 100 plant species including 21 tree species, 15 shrub species, 23 perennial herb species, 37 annual herb species and four algal species in the dataset of Dicots PLAZA 5.0. Shrub species were eliminated from the analyses.

| Life form | Species name | Reference |
|---|---|---|
| Tree: 21 species | *Acer truncatum* | Plant for a future |
| | *Amborella trichopoda* | Angiosperm Phylogeny Website |
| | *Avicennia marina* | USDA PLANTS database |
| | *Carica papaya* | PLANTS database |
| | *Carpinus fangiana* | Plant of the world online |
| | *Carya illinoinensis* | PLANTS database |
| | *Citrus clementina* | Plants For A Future |
| | *Coffea canephora* | Plants of the World online |
| | *Davidia involucrata* | Plant of the world online |
| | *Durio zibethinus* | Plant for a future |
| | *Eucalyptus grandis* | PLANTS database |
| | *Magnolia biondii* | Dong et al. (2021) |
| | *Malus domestica* | PLANTS database |
| | *Olea europaea* | Plant for a future |
| | *Populus trichocarpa* | PLANTS database |
| | *Prunus persica* | PLANTS database |
| | *Punica granatum* | Plant for a future |
| | *Quercus lobata* | PLANTS database |
| | *Sequoiadendron giganteum* | PLANTS database |
| | *Theobroma cacao* | PLANTS database |
| | *Trochodendron aralioides* | eFloras |
| Shrub: 15 species | *Actinidia chinensis* | PLANTS database |
| | *Camellia sinensis* var. *sinensis* | Plant for a future |
| | *Corylus avellana* | Plant for a future |
| | *Gossypium hirsutum* | Plant for a future |
| | *Gossypium raimondii* | Gotmare, Singh, Tule (2000) |

| | | |
|---|---|---|
| | *Hydrangea macrophylla* | Plants For A Future |
| | *Manihot esculenta* | PLANTS database |
| | *Rhododendron simsii* | eFloras |
| | *Rosa chinensis* | Plant for a future |
| | *Salix brachista* | eFloras |
| | *Selenicereus undatus* | Plant for a future |
| | *Simmondsia chinensis* | PLANTS database |
| | *Tripterygium wilfordii* | eFloras |
| | *Vaccinium macrocarpon* | PLANTS database |
| | *Vitis vinifera* | PLANTS database |
| Perennial herb: 23 species | *Aquilegia oxysepala* | Plant for a future |
| | *Arabidopsis lyrata* | PLANTS database |
| | *Brassica oleracea* | PLANTS database |
| | *Capsicum annuum* | PLANTS database |
| | *Ceratophyllum demersum* | PLANTS database |
| | *Erythranthe guttata* | The University and Jepson Herbaria |
| | *Fragaria vesca* | PLANTS database |
| | *Fragaria* x *ananassa* | Plant for a future |
| | *Lonicera japonica* | PLANTS database |
| | *Lotus japonicus* | eFloras |
| | *Marchantia polymorpha* | University of Massachusetts Weed Herbarium |
| | *Nelumbo nucifera* | PLANTS database |
| | *Nicotiana tabacum* | PLANTS database |
| | *Oryza sativa* ssp. *japonica* | Takasaki et al. (1994) |
| | *Salvia bowleyana* | eFloras |
| | *Sechium edule* | Plant for a future |
| | *Selaginella moellendorffii* | Zhang, Hans, Kato (2013) |
| | *Solanum lycopersicum* | PLANTS database |
| | *Solanum pennellii* | Solanaceae Source |
| | *Solanum tuberosum* | PLANTS database |
| | *Trifolium pratense* | PLANTS database |
| | *Utricularia gibba* | PLANTS database |
| | *Vanilla planifolia* | PLANTS database |
| Annual herb: 37 species | *Aethionema arabicum* | Mérai et al. (2019) |
| | *Amaranthus hybridus* | USDA PLANTS database |
| | *Anthoceros agrestis* | Bisang (2003) |

| | | |
|---|---|---|
| | *Arabidopsis thaliana* | PLANTS database |
| | *Arachis hypogaea* | Plant for a future |
| | *Beta vulgaris* | PLANTS database |
| | *Brassica carinata* | Plant for a future |
| | *Brassica napus* | PLANTS database |
| | *Brassica rapa* | PLANTS database |
| | *Cannabis sativa* | PLANTS database |
| | *Capsella rubella* | PLANTS database |
| | *Cardamine hirsuta* | PLANTS database |
| | *Chenopodium quinoa* | Plants For A Future |
| | *Cicer arietinum* L. | PLANTS database |
| | *Citrullus lanatus* | PLANTS database |
| | *Corchorus olitorius* | PLANTS database |
| | *Cucumis melo* | PLANTS database |
| | *Cucumis sativus* L. | PLANTS database |
| | *Daucus carota* | PLANTS database |
| | *Erigeron canadensis* | Plant for a future |
| | *Eutrema salsugineum* | Yang et al. (2013) |
| | *Glycine max* | PLANTS database |
| | *Helianthus annuus* | PLANTS database |
| | *Lactuca sativa* | Plants For A Future |
| | *Lupinus albus* | PLANTS database |
| | *Medicago truncatula* | Tivoli et al. (2006) |
| | *Papaver somniferum* | PLANTS database |
| | *Petunia axillaris* | PLANTS database |
| | *Phaseolus vulgaris* | PLANTS database |
| | *Physcomitrium patens* | Cove (2005) |
| | *Pisum sativum* | PLANTS database |
| | *Sapria himalayana* | Borah & Ghosh (2018) |
| | *Schrenkiella parvula* | Inan et al. (2004) |
| | *Striga asiatica* | PLANTS database |
| | *Tarenaya hassleriana* | PLANTS database |
| | *Vigna mungo* | PLANTS database |
| | *Zea mays* | PLANTS database |
| Alga: 4 species | *Chara braunii* | Kato et al. (2008) |
| | *Chlamydomonas reinhardtii* | Merchant et al. (2007) |

| | |
|---|---|
| *Micromonas commoda* | van Baren et al. (2016) |
| *Prasinoderma coloniale* | Li et al. (2020) |

**Appendix Table S2.** The list of genes and gene families for analyses. There were 221 genes and 121 gene families in the dataset. Each gene family was categorized into one of five functional groups (DNA modification, histone modification, chromatin formation or chromatin remodeling, Polycomb-group proteins and interacting components, RNA silencing).

| Function group | Gene Family ID in Dicots PLAZA 5.0 | Gene symbol | AT code |
|---|---|---|---|
| Chromatin formation or chromatin remodeling | HOM05D000104 | *SWI2* | AT1G03750 |
| | HOM05D000104 | *CHR5* | AT2G13370 |
| | HOM05D000104 | *CHD3/PKL* | AT2G25170 |
| | HOM05D000104 | *SPD/SYD* | AT2G28290 |
| | HOM05D000104 | *BRM* | AT2G46020 |
| | HOM05D000104 | *AtCHR12* | AT3G06010 |
| | HOM05D000104 | *CHR11* | AT3G06400 |
| | HOM05D000104 | *PIE* | AT3G12810 |
| | HOM05D000104 | *RAD54* | AT3G19210 |
| | HOM05D000104 | *INO80* | AT3G57300 |
| | HOM05D000104 | *PKR2/CHR7* | AT4G31900 |
| | HOM05D000104 | *CHR17* | AT5G18620 |
| | HOM05D000104 | *PKR1/CHR4* | AT5G44800 |
| | HOM05D000104 | *DDM1/CHR1* | AT5G66750 |
| | HOM05D000173 | *ARP4* | AT1G18450 |
| | HOM05D000173 | *ARP8* | AT5G56180 |
| | HOM05D000347 | *MSI1* | AT5G58230 |
| | HOM05D000515 | *SNF2-RING-HELICASE LIKE5* | AT1G11100 |
| | HOM05D000515 | *FRG2/SNF2-RING-HELICASE LIKE2* | AT1G50410 |
| | HOM05D000515 | *SNF2-RING-HELICASE LIKE4* | AT1G61140 |
| | HOM05D000515 | *SNF2-RING-HELICASE LIKE3* | AT3G16600 |
| | HOM05D000515 | *FRG1/SNF2-RING-HELICASE LIKE1* | AT3G20010 |
| | HOM05D000526 | *ARP5* | AT3G12380 |
| | HOM05D000725 | *DRD1* | AT2G16390 |
| | HOM05D000725 | *CLSY1* | AT3G42670 |

| | | |
|---|---|---|
| HOM05D000902 | *DMS11* | AT1G19100 |
| HOM05D001065 | *AtSWI3_C/SWI3C* | AT1G21700 |
| HOM05D001065 | *AtSWI3_B/SWI3B* | AT2G33610 |
| HOM05D001065 | *AtSWI3_A/SWI3A* | AT2G47620 |
| HOM05D001065 | *AtSWI3_D* | AT4G34430 |
| HOM05D001081 | *DMS3/IDN1* | AT3G49250 |
| HOM05D001215 | *AtNAP1_2* | AT2G19480 |
| HOM05D001215 | *AtNAP1_4* | AT3G13782 |
| HOM05D001215 | *AtNAP1_1* | AT4G26110 |
| HOM05D001215 | *AtNAP1_3* | AT5G56950 |
| HOM05D001331 | *AtRad21.1* | AT5G40840 |
| HOM05D001642 | *SPT16* | AT4G10710 |
| HOM05D001674 | *CHR18* | AT1G48310 |
| HOM05D001944 | *MOM1* | AT1G08060 |
| HOM05D002208 | *NRP2* | AT1G18800 |
| HOM05D002208 | *NAP1/NRP1* | AT1G74560 |
| HOM05D002404 | *PCNA1* | AT1G07370 |
| HOM05D002404 | *PCNA2* | AT2G29570 |
| HOM05D002662 | *AtSWP73_A* | AT3G01890 |
| HOM05D002662 | *AtSWP73_B/CHC1* | AT5G14170 |
| HOM05D002664 | *SWI1* | AT5G51330 |
| HOM05D002728 | *RPA2* | AT2G24490 |
| HOM05D002795 | *MGO1* | AT5G55300 |
| HOM05D003239 | *AtASF1a* | AT1G66740 |
| HOM05D003239 | *AtASF1b* | AT5G38110 |
| HOM05D003321 | *SSRP1* | AT3G28730 |
| HOM05D003901 | *HIRA* | AT3G44530 |
| HOM05D004146 | *SWR1* | AT2G47210 |
| HOM05D004178 | *SMC6A* | AT5G07660 |
| HOM05D004178 | *MIM/RAD18/SMC6B* | AT5G61460 |
| HOM05D004779 | *TSL* | AT5G20930 |
| HOM05D005030 | *BRU1/MGO3/TSK* | AT3G18730 |
| HOM05D005087 | *SMC5* | AT5G15920 |
| HOM05D005401 | *FAS2* | AT5G64630 |
| HOM05D005631 | *SEF/SWC6* | AT5G37055 |
| HOM05D005855 | *FAS1* | AT1G65470 |

| | | | |
|---|---|---|---|
| | HOM05D006316 | *BSH* | AT3G17590 |
| | HOM05D006561 | *MMS21* | AT3G15150 |
| | HOM05D007494 | *SWI-SNF-related chromatin binding protein* | AT1G20290 |
| DNA modification | HOM05D000288 | *H3.3, HTR4* | AT4G40030 |
| | HOM05D000288 | *H3.3, HTR5* | AT4G40040 |
| | HOM05D000288 | *H3.3, HTR8* | AT5G10980 |
| | HOM05D000572 | *DDM2/MET1* | AT5G49160 |
| | HOM05D000771 | *MBD10* | AT1G15340 |
| | HOM05D001165 | *CMT3* | AT1G69770 |
| | HOM05D001165 | *CMT2* | AT4G19020 |
| | HOM05D001201 | *DML1/ROS1* | AT2G36490 |
| | HOM05D001201 | *DML2* | AT3G10010 |
| | HOM05D001201 | *DML3* | AT4G34060 |
| | HOM05D001201 | *DME* | AT5G04560 |
| | HOM05D001290 | *MTHFD1* | AT3G12290 |
| | HOM05D001482 | *DRM2* | AT5G14620 |
| | HOM05D001482 | *DRM1* | AT5G15380 |
| | HOM05D001972 | *VIM1* | AT1G57820 |
| | HOM05D001972 | *VIM2* | AT1G66050 |
| | HOM05D001972 | *VIM3* | AT5G39550 |
| | HOM05D002127 | *HOG1* | AT4G13940 |
| | HOM05D004190 | *MBD6* | AT5G59380 |
| | HOM05D005043 | *ZDP* | AT3G14890 |
| | HOM05D005914 | *DNMT2* | AT5G25480 |
| | HOM05D006136 | *ROS3* | AT5G58130 |
| | HOM05D006922 | *DDB2* | AT5G58760 |
| | HOM05D007289 | *XRCC1* | AT1G80420 |
| Histon modification | HOM05D000010 | *SGS1/NAC052* | AT3G10490 |
| | HOM05D000010 | *NAC103* | AT5G64060 |
| | HOM05D000050 | *UBC1* | AT1G14400 |
| | HOM05D000050 | *UBC2* | AT2G02760 |
| | HOM05D000268 | *SUVH6* | AT2G22740 |
| | HOM05D000268 | *SUVH2* | AT2G33290 |
| | HOM05D000268 | *SUVH5* | AT2G35160 |
| | HOM05D000268 | *SUVH4* | AT5G13960 |
| | HOM05D000329 | *MEE27* | AT2G34880 |

| | | |
|---|---|---|
| HOM05D000329 | *REF6* | AT3G48430 |
| HOM05D000329 | *JMJ14* | AT4G20400 |
| HOM05D000329 | *ELF6* | AT5G04240 |
| HOM05D000451 | *IBM1* | AT3G07610 |
| HOM05D000461 | *LDL1* | AT1G62830 |
| HOM05D000461 | *FLD* | AT3G10390 |
| HOM05D000461 | *LDL2* | AT3G13682 |
| HOM05D000590 | *ATX1* | AT2G31650 |
| HOM05D000912 | *HAC12* | AT1G16710 |
| HOM05D000912 | *HAC1* | AT1G79000 |
| HOM05D000912 | *HAC5* | AT3G12980 |
| HOM05D000966 | *EFS/SDG8/ASHH2* | AT1G77300 |
| HOM05D001141 | *HDA1/HDA19/RPD3A* | AT4G38130 |
| HOM05D001141 | *AXE1/HDA6/RPD3B/RTS1/SIL1* | AT5G63110 |
| HOM05D001240 | *HUB2* | AT1G55250 |
| HOM05D001240 | *HUB1* | AT2G44950 |
| HOM05D001451 | *UBC5* | AT1G63800 |
| HOM05D001587 | *ATUBC2-1* | AT1G45050 |
| HOM05D001688 | *HD2d/HDT4* | AT2G27840 |
| HOM05D001688 | *HD2a/HDT1* | AT3G44750 |
| HOM05D001688 | *HD2c/HDT3* | AT5G03740 |
| HOM05D001688 | *HD2b/HDT2* | AT5G22650 |
| HOM05D001734 | *ATM* | AT3G48190 |
| HOM05D001734 | *ATR* | AT5G40820 |
| HOM05D001937 | *HAF1* | AT1G32750 |
| HOM05D001937 | *TAF1* | AT3G19040 |
| HOM05D002415 | *ATXR5* | AT5G09790 |
| HOM05D002415 | *ATXR6* | AT5G24330 |
| HOM05D002939 | *ULT1* | AT4G28190 |
| HOM05D003463 | *ATXR3/SDG2* | AT4G15180 |
| HOM05D004103 | *SRT1* | AT5G55760 |
| HOM05D004180 | *HAG3* | AT5G50320 |
| HOM05D004616 | *OTLD1* | AT2G27350 |
| HOM05D004718 | *SUP32/UBP26* | AT3G49600 |
| HOM05D005044 | *HAG1* | AT3G54610 |
| HOM05D005294 | *HAG2* | AT5G56740 |

| | | | |
|---|---|---|---|
| | HOM05D005757 | *SRT2* | AT5G09230 |
| | HOM05D006833 | *ATXR7* | AT5G42400 |
| Polycomb-group proteins and interacting components | HOM05D000144 | *AtCYP71* | AT3G44600 |
| | HOM05D000319 | *CUL4* | AT5G46210 |
| | HOM05D000809 | *VRN5* | AT3G24440 |
| | HOM05D000809 | *VEL1/VIL2* | AT4G30200 |
| | HOM05D000809 | *VIN3* | AT5G57380 |
| | HOM05D001069 | *AtBMI1a* | AT2G30580 |
| | HOM05D001873 | *LIF2* | AT4G00830 |
| | HOM05D001902 | *FIS1/MEA* | AT1G02580 |
| | HOM05D001902 | *CLF/SET1* | AT2G23380 |
| | HOM05D001902 | *SWN* | AT4G02020 |
| | HOM05D002164 | *FIS2* | AT2G35670 |
| | HOM05D002164 | *VRN2* | AT4G16845 |
| | HOM05D002164 | *EMF2* | AT5G51230 |
| | HOM05D002302 | *MSI4/FVE* | AT2G19520 |
| | HOM05D002302 | *MSI5* | AT4G29730 |
| | HOM05D002349 | *AtRING1b* | AT1G03770 |
| | HOM05D002349 | *AtRING1a* | AT5G44280 |
| | HOM05D003609 | *DDB1A* | AT4G05420 |
| | HOM05D003609 | *DDB1B* | AT4G21100 |
| | HOM05D003719 | *LHP1/TFL2* | AT5G17690 |
| | HOM05D003977 | *RBR* | AT3G12280 |
| | HOM05D004312 | *FIE/FIS3* | AT3G20740 |
| RNA silencing | HOM05D000228 | *POL IV/SMD2* | AT1G63020 |
| | HOM05D000228 | *DRD3/NRPE1* | AT2G40030 |
| | HOM05D000228 | *NRPC1* | AT5G60040 |
| | HOM05D000234 | *AGO2* | AT1G31280 |
| | HOM05D000234 | *AGO3* | AT1G31290 |
| | HOM05D000234 | *AGO1* | AT1G48410 |
| | HOM05D000234 | *AGO7/ZIP* | AT1G69440 |
| | HOM05D000234 | *AGO4* | AT2G27040 |
| | HOM05D000234 | *AGO5* | AT2G27880 |
| | HOM05D000234 | *AGO6* | AT2G32940 |
| | HOM05D000234 | *AGO8* | AT5G21030 |
| | HOM05D000234 | *AGO9* | AT5G21150 |

| | | |
|---|---|---|
| HOM05D000234 | *AGO10/PNH/ZLL* | AT5G43810 |
| HOM05D000399 | *DCL1/EMB76/SIN1/SUS1* | AT1G01040 |
| HOM05D000399 | *DCL2* | AT3G03300 |
| HOM05D000399 | *DCL3* | AT3G43920 |
| HOM05D000399 | *DCL4* | AT5G20320 |
| HOM05D000537 | *FDM4* | AT1G13790 |
| HOM05D000537 | *FDM1* | AT1G15910 |
| HOM05D000537 | *FDM5* | AT1G80790 |
| HOM05D000537 | *FDM3* | AT3G12550 |
| HOM05D000537 | *IDN2/RDM12* | AT3G48670 |
| HOM05D000537 | *FDM2* | AT4G00380 |
| HOM05D000611 | *DRB1/HYL1* | AT1G09700 |
| HOM05D000611 | *DRB2* | AT2G28380 |
| HOM05D000611 | *DRB3* | AT3G26932 |
| HOM05D000611 | *DRB4* | AT3G62800 |
| HOM05D000688 | *NRPD2B* | AT3G18090 |
| HOM05D000688 | *DRD2/NRPD2A/NRPE2* | AT3G23780 |
| HOM05D000688 | *NRPC2* | AT5G45140 |
| HOM05D000822 | *RDR1* | AT1G14790 |
| HOM05D000822 | *RDR6/SDE1/SGS2* | AT3G49500 |
| HOM05D000822 | *RDR2/SMD1* | AT4G11130 |
| HOM05D000917 | *XRN4/EIN5* | AT1G54490 |
| HOM05D000917 | *XRN3* | AT1G75660 |
| HOM05D000917 | *XRN2* | AT5G42540 |
| HOM05D001100 | *FRY1/SAL1* | AT5G63980 |
| HOM05D001296 | *NRPB5/NRPD5* | AT3G22320 |
| HOM05D001296 | *NRPE5* | AT3G57080 |
| HOM05D001495 | *FCA* | AT4G16280 |
| HOM05D001605 | *KTF1/RDM3/SPT5-l* | AT5G04290 |
| HOM05D001613 | *SHH1/DTF1* | AT1G15215 |
| HOM05D002300 | *FPA* | AT2G43410 |
| HOM05D002459 | *SGS3* | AT5G23570 |
| HOM05D002658 | *AtNUC-l1* | AT1G48920 |
| HOM05D002720 | *NRPD7* | AT3G22900 |
| HOM05D002720 | *NRPE7/NRPD7b* | AT4G14660 |
| HOM05D002863 | *SDE3* | AT1G05460 |

| | | |
|---|---|---|
| HOM05D003289 | *ESD7* | AT1G08260 |
| HOM05D003897 | *POL V/NRPE3b* | AT2G15400 |
| HOM05D003897 | *NRPB3/NRPD3/NRPE3a* | AT2G15430 |
| HOM05D004187 | *RDM1* | AT3G22680 |
| HOM05D004205 | *SDE5* | AT3G15390 |
| HOM05D004256 | *HEN1* | AT4G20910 |
| HOM05D004365 | *NRPB9a/NRPD9a/NRPE9a* | AT3G16980 |
| HOM05D004365 | *NRPB9b/NRPD9b/NRPE9b* | AT4G16265 |
| HOM05D004384 | *HST* | AT3G05040 |
| HOM05D004774 | *NRPC7* | AT1G06790 |
| HOM05D005148 | *DDL* | AT3G20550 |
| HOM05D005238 | *SR45* | AT1G16610 |
| HOM05D005600 | *ABH1/CBP80* | AT2G13540 |
| HOM05D006271 | *RDM4/DMS4* | AT2G30280 |
| HOM05D006987 | *WEX* | AT4G13870 |

**Appendix Table S3.** The list of the *BRU1/TSK/MGO3* genes. There were 148 genes. 35 genes were removed from the construction of the phylogenetic tree.

| Gene ID in Dicots PLAZA 5.0 | Species | Life form | Phylogenetic tree |
|---|---|---|---|
| Atru.chr4.774 | *Acer truncatum* | Tree | |
| ATR0772G104 | *Amborella trichopoda* | Tree | |
| MSTRG.2677 | *Avicennia marina* | Tree | removed |
| MSTRG.2678 | *Avicennia marina* | Tree | |
| Cpa.g.sc42.107 | *Carica papaya* | Tree | |
| Cpa.g.sc42.108 | *Carica papaya* | Tree | removed |
| Cfa002838 | *Carpinus fangiana* | Tree | |
| CiPaw.03G127100 | *Carya illinoinensis* | Tree | |
| CiPaw.03G127200 | *Carya illinoinensis* | Tree | removed |
| CiPaw.03G127500 | *Carya illinoinensis* | Tree | |
| CiPaw.03G127600 | *Carya illinoinensis* | Tree | |
| Ciclev10024723m.g | *Citrus clementina* | Tree | |
| Cc07_g17780 | *Coffea canephora* | Tree | |
| Cc08_g02710 | *Coffea canephora* | Tree | |
| Dinv05712 | *Davidia involucrata* | Tree | |
| Dinv24313 | *Davidia involucrata* | Tree | |
| Duzib147G1675 | *Durio zibethinus* | Tree | |
| Eucgr.H04934 | *Eucalyptus grandis* | Tree | |
| MBI18_g25945_MAGBIO | *Magnolia biondii* | Tree | |
| MBI19_g06116_MAGBIO | *Magnolia biondii* | Tree | removed |
| MD00G1172700 | *Malus domestica* | Tree | |
| MD09G1255500 | *Malus domestica* | Tree | removed |
| MD17G1248600 | *Malus domestica* | Tree | removed |
| MD17G1248700 | *Malus domestica* | Tree | |
| Oeu037792.2 | *Olea europaea* | Tree | |
| Potri.007G110800 | *Populus trichocarpa* | Tree | |
| Prupe.3G147500 | *Prunus persica* | Tree | |
| PGR102G2159 | *Punica granatum* | Tree | |
| QL06p030436 | *Quercus lobata* | Tree | |
| SEGI_07118 | *Sequoiadendron giganteum* | Tree | |

| | | | |
|---|---|---|---|
| SEGI_08752 | *Sequoiadendron giganteum* | Tree | removed |
| SEGI_13008 | *Sequoiadendron giganteum* | Tree | removed |
| SEGI_25172 | *Sequoiadendron giganteum* | Tree | removed |
| SEGI_25247 | *Sequoiadendron giganteum* | Tree | removed |
| SEGI_29284 | *Sequoiadendron giganteum* | Tree | |
| SEGI_37730 | *Sequoiadendron giganteum* | Tree | removed |
| Thecc.01G124600 | *Theobroma cacao* | Tree | |
| Thecc.01G125300 | *Theobroma cacao* | Tree | removed |
| TAR376G0051 | *Trochodendron aralioides* | Tree | removed |
| TAR376G0117 | *Trochodendron aralioides* | Tree | removed |
| TAR381G0325 | *Trochodendron aralioides* | Tree | removed |
| TAR381G0402 | *Trochodendron aralioides* | Tree | |
| TAR625G0665 | *Trochodendron aralioides* | Tree | removed |
| TAR625G0866 | *Trochodendron aralioides* | Tree | removed |
| Actinidia06653 | *Actinidia chinensis* | Shrub | |
| Actinidia09824 | *Actinidia chinensis* | Shrub | removed |
| CSS0019451 | *Camellia sinensis* | Shrub | |
| Haze_25135 | *Corylus avellana* | Shrub | removed |
| Haze_25140 | *Corylus avellana* | Shrub | |
| Gohir.A11G230400 | *Gossypium hirsutum* | Shrub | |
| Gohir.D11G232200 | *Gossypium hirsutum* | Shrub | |
| Gorai.002G151400 | *Gossypium raimondii* | Shrub | |
| Hma1.2p1_0006F.1_g004730 | *Hydrangea macrophylla* | Shrub | removed |
| Hma1.2p1_0006F.1_g004750 | *Hydrangea macrophylla* | Shrub | |
| Manes.12G152000 | *Manihot esculenta* | Shrub | |
| Rhsim10G0138800 | *Rhododendron simsii* | Shrub | |
| RcHm_v2.0_Chr2g0120111 | *Rosa chinensis* | Shrub | |
| Sabra05G0049200 | *Salix brachista* | Shrub | removed |
| Sabra07G0090700 | *Salix brachista* | Shrub | |
| Hund04465 | *Selenicereus undatus* | Shrub | |
| Sc03g0005100 | *Simmondsia chinensis* | Shrub | |
| Sc05g0004550 | *Simmondsia chinensis* | Shrub | removed |
| TWI31G0652 | *Tripterygium wilfordii* | Shrub | |
| vmacro12843 | *Vaccinium macrocarpon* | Shrub | removed |
| vmacro12844 | *Vaccinium macrocarpon* | Shrub | removed |
| vmacro12845 | *Vaccinium macrocarpon* | Shrub | removed |

| | | | |
|---|---|---|---|
| GSVIVG01026545001 | *Vitis vinifera* | Shrub | |
| Aqoxy5G02153 | *Aquilegia oxysepala* | Perennial | |
| AL3G32080 | *Arabidopsis lyrata* | Perennial | |
| BolC5t33412H | *Brassica oleracea* | Perennial | |
| CAN.G802.6 | *Capsicum annuum* | Perennial | removed |
| CDE06G1651 | *Ceratophyllum demersum* | Perennial | |
| CDE08G0761 | *Ceratophyllum demersum* | Perennial | removed |
| Migut.H01102 | *Erythranthe guttata* | Perennial | |
| FvH4_1g23980 | *Fragaria vesca* | Perennial | |
| FAN19G2772 | *Fragaria x ananassa* | Perennial | |
| FAN23G1259 | *Fragaria x ananassa* | Perennial | |
| FAN23G1614 | *Fragaria x ananassa* | Perennial | removed |
| FAN26G2615 | *Fragaria x ananassa* | Perennial | |
| FAN28G1848 | *Fragaria x ananassa* | Perennial | |
| Lj2A602G37 | *Lonicera japonica* | Perennial | |
| Lj4g0022240 | *Lotus japonicus* | Perennial | |
| Mapoly0047s0105 | *Marchantia polymorpha* | Perennial | |
| Nn5g28681 | *Nelumbo nucifera* | Perennial | |
| Nitab4.5_0002909g0070 | *Nicotiana tabacum* | Perennial | |
| Nitab4.5_0003048g0010 | *Nicotiana tabacum* | Perennial | |
| Os02g0782800 | *Oryza sativa japonica* | Perennial | |
| Os02g0784100 | *Oryza sativa japonica* | Perennial | removed |
| Sed0011527 | *Sechium edule* | Perennial | |
| SMO203G0163 | *Selaginella moellendorffii* | Perennial | |
| Solyc11g005690.3 | *Solanum lycopersicum* | Perennial | |
| Sopen11g001650 | *Solanum pennellii* | Perennial | |
| PGSC0003DMG400025561 | *Solanum tuberosum* | Perennial | removed |
| TPR.G24704 | *Trifolium pratense* | Perennial | removed |
| TPR.G37387 | *Trifolium pratense* | Perennial | |
| unitig_0.g2514 | *Utricularia gibba* | Perennial | |
| HPP92_005719 | *Vanilla planifolia* | Perennial | |
| Aa31LG8G5540 | *Aethionema arabicum* | Annual | |
| Ah.03g146670 | *Amaranthus hybridus* | Annual | |
| AagrBONN_evm.TU.Sc2ySwM_228.5700 | *Anthoceros agrestis* | Annual | |
| AT3G18730 | *Arabidopsis thaliana* | Annual | |
| arahy.Tifrunner.gnm1.ann1.CRY23I | *Arachis hypogaea* | Annual | |

| | | | |
|---|---|---|---|
| arahy.Tifrunner.gnm1.ann1.HR53B8 | *Arachis hypogaea* | Annual | |
| EL10Ac6g15267 | *Beta vulgaris* | Annual | |
| BcaB06g25924 | *Brassica carinata* | Annual | |
| BcaC05g28579 | *Brassica carinata* | Annual | |
| BcaNung06136 | *Brassica carinata* | Annual | removed |
| A05p30890 | *Brassica napus* | Annual | |
| C05p46820 | *Brassica napus* | Annual | |
| BraA05t21931Z | *Brassica rapa* | Annual | |
| CANSAT01G2434 | *Cannabis sativa* | Annual | |
| Carub.0003s1851 | *Capsella rubella* | Annual | |
| CARHR094120 | *Cardamine hirsuta* | Annual | |
| AUR62003613 | *Chenopodium quinoa* | Annual | |
| AUR62017924 | *Chenopodium quinoa* | Annual | |
| Ca_17653_v3 | *Cicer arietinum* | Annual | |
| ClCG01G015590 | *Citrullus lanatus* | Annual | |
| COL.COLO4_13999 | *Corchorus olitorius* | Annual | |
| MELO3C008038.2 | *Cucumis melo* | Annual | |
| CsaV3_6G041830 | *Cucumis sativus* | Annual | |
| DCAR_001349 | *Daucus carota* | Annual | |
| ECA234G1901 | *Erigeron canadensis* | Annual | removed |
| ECA234G2315 | *Erigeron canadensis* | Annual | removed |
| ECA240G4559 | *Erigeron canadensis* | Annual | |
| Thhalv10019906m.g | *Eutrema salsugineum* | Annual | |
| Glyma.05G189400 | *Glycine max* | Annual | |
| Glyma.08G147000 | *Glycine max* | Annual | |
| HanXRQChr02g0043651 | *Helianthus annuus* | Annual | |
| Lsat_1_v5_gn_1_117161 | *Lactuca sativa* | Annual | |
| Lalb_Chr12g0203281 | *Lupinus albus* | Annual | |
| Lalb_Chr13g0296501 | *Lupinus albus* | Annual | |
| Medtr8g093100 | *Medicago truncatula* | Annual | |
| PSO210G3052 | *Papaver somniferum* | Annual | |
| PSO832G2315 | *Papaver somniferum* | Annual | |
| Peaxi162Scf00012g02419 | *Petunia axillaris* | Annual | |
| Phvul.002G270100 | *Phaseolus vulgaris* | Annual | |
| Pp3c6_6440 | *Physcomitrium patens* | Annual | |
| Psat7g048200 | *Pisum sativum* | Annual | |

| | | | |
|---|---|---|---|
| SHI11439 | *Sapria himalayana* | Annual | |
| Sp3g16780 | *Schrenkiella parvula* | Annual | |
| SGA_v2.0_scaffold137G35137 | *Striga asiatica* | Annual | |
| THA.LOC104820424 | *Tarenaya hassleriana* | Annual | |
| THA.LOC104821859 | *Tarenaya hassleriana* | Annual | |
| VMungo0251G0340 | *Vigna mungo* | Annual | |
| Zm00001eb192940 | *Zea mays* | Annual | |
| CBR_g36663 | *Chara braunii* | alga | |
| CBR_g36665 | *Chara braunii* | alga | removed |
| CBR_g61481 | *Chara braunii* | alga | |

**Appendix Table S4.** The list of the *SDE3* genes. There were 242 genes. 32 genes were removed from the construction of the phylogenetic tree.

| Gene ID in Dicots PLAZA 5.0 | Species | Life form | Phylogenetic tree |
|---|---|---|---|
| Atru.ctg727.2 | *Acer truncatum* | Tree | |
| Atru.chr11.1111 | *Acer truncatum* | Tree | |
| Atru.chr11.1104 | *Acer truncatum* | Tree | |
| Atru.chr11.1113 | *Acer truncatum* | Tree | |
| Atru.chr4.2430 | *Acer truncatum* | Tree | |
| ATR0618G110 | *Amborella trichopoda* | Tree | |
| ATR0618G119 | *Amborella trichopoda* | Tree | |
| ATR0665G272 | *Amborella trichopoda* | Tree | |
| Cpa.g.sc60.77 | *Carica papaya* | Tree | |
| Cpa.g.sc117.40 | *Carica papaya* | Tree | |
| Cfa008768 | *Carpinus fangiana* | Tree | |
| Cfa003109 | *Carpinus fangiana* | Tree | |
| Cfa008472 | *Carpinus fangiana* | Tree | |
| CiPaw.01G198300 | *Carya illinoinensis* | Tree | |
| Ciclev10030791m.g | *Citrus clementina* | Tree | |
| Ciclev10033534m.g | *Citrus clementina* | Tree | |
| Ciclev10031085m.g | *Citrus clementina* | Tree | |
| Ciclev10033310m.g | *Citrus clementina* | Tree | |
| Ciclev10030734m.g | *Citrus clementina* | Tree | |
| Ciclev10004283m.g | *Citrus clementina* | Tree | |
| Ciclev10000353m.g | *Citrus clementina* | Tree | |
| Cc11_g01050 | *Coffea canephora* | Tree | |
| Cc11_g01040 | *Coffea canephora* | Tree | |
| Cc03_g04040 | *Coffea canephora* | Tree | |
| Cc02_g31310 | *Coffea canephora* | Tree | |
| Dinv20810 | *Davidia involucrata* | Tree | |
| Dinv17129 | *Davidia involucrata* | Tree | |
| Dinv17741 | *Davidia involucrata* | Tree | |
| Dinv14015 | *Davidia involucrata* | Tree | |
| Duzib052G0279 | *Durio zibethinus* | Tree | |

| | | | |
|---|---|---|---|
| Duzib052G1003 | *Durio zibethinus* | Tree | |
| Duzib151G1175 | *Durio zibethinus* | Tree | |
| Duzib177G0218 | *Durio zibethinus* | Tree | |
| Eucgr.J03176 | *Eucalyptus grandis* | Tree | |
| Eucgr.J00884 | *Eucalyptus grandis* | Tree | |
| Eucgr.J00890 | *Eucalyptus grandis* | Tree | |
| Eucgr.H03436 | *Eucalyptus grandis* | Tree | |
| Eucgr.H04399 | *Eucalyptus grandis* | Tree | |
| MBI03_g13827_MAGBIO | *Magnolia biondii* | Tree | |
| MBI03_g01415_MAGBIO | *Magnolia biondii* | Tree | |
| MBI05_g12094_MAGBIO | *Magnolia biondii* | Tree | |
| MBI03_g01414_MAGBIO | *Magnolia biondii* | Tree | removed |
| MD13G1238300 | *Malus domestica* | Tree | |
| MD13G1196500 | *Malus domestica* | Tree | |
| MD16G1243100 | *Malus domestica* | Tree | removed |
| MD16G1243200 | *Malus domestica* | Tree | removed |
| Oeu001544.1 | *Olea europaea* | Tree | |
| Oeu001547.1 | *Olea europaea* | Tree | |
| Oeu018974.1 | *Olea europaea* | Tree | |
| Potri.005G089800 | *Populus trichocarpa* | Tree | |
| Potri.007G074070 | *Populus trichocarpa* | Tree | |
| Potri.005G047500 | *Populus trichocarpa* | Tree | |
| Potri.008G155400 | *Populus trichocarpa* | Tree | |
| Prupe.1G070100 | *Prunus persica* | Tree | |
| Prupe.1G070300 | *Prunus persica* | Tree | |
| Prupe.1G070200 | *Prunus persica* | Tree | |
| Prupe.1G070400 | *Prunus persica* | Tree | |
| Prupe.1G017300 | *Prunus persica* | Tree | |
| PGR004G1808 | *Punica granatum* | Tree | |
| PGR004G0456 | *Punica granatum* | Tree | |
| QL04p079513 | *Quercus lobata* | Tree | |
| QL04p079501 | *Quercus lobata* | Tree | |
| QL04p079558 | *Quercus lobata* | Tree | |
| QL04p079654 | *Quercus lobata* | Tree | |
| QL04p079624 | *Quercus lobata* | Tree | |
| QL04p079664 | *Quercus lobata* | Tree | |

| | | | |
|---|---|---|---|
| QL06p001686 | *Quercus lobata* | Tree | |
| QL04p079604 | *Quercus lobata* | Tree | removed |
| QL04p079598 | *Quercus lobata* | Tree | removed |
| QL04p079673 | *Quercus lobata* | Tree | removed |
| SEGI_07136 | *Sequoiadendron giganteum* | Tree | |
| SEGI_10221 | *Sequoiadendron giganteum* | Tree | removed |
| Thecc.02G156400 | *Theobroma cacao* | Tree | |
| Thecc.04G235200 | *Theobroma cacao* | Tree | |
| Thecc.05G291200 | *Theobroma cacao* | Tree | |
| Thecc.02G155900 | *Theobroma cacao* | Tree | removed |
| TAR622G0765 | *Trochodendron aralioides* | Tree | |
| Hund10646 | *Trochodendron aralioides* | Tree | |
| TAR636G1187 | *Trochodendron aralioides* | Tree | |
| TAR260G0006 | *Trochodendron aralioides* | Tree | removed |
| TAR719G0001 | *Trochodendron aralioides* | Tree | removed |
| TAR719G0002 | *Trochodendron aralioides* | Tree | removed |
| vmacro19508 | *Vaccinium macrocarpon* | Tree | |
| vmacro14855 | *Vaccinium macrocarpon* | Tree | |
| Actinidia04334 | *Actinidia chinensis* | Shrub | |
| CSS0037612 | *Camellia sinensis* | Shrub | |
| CSS0007282 | *Camellia sinensis* | Shrub | |
| CSS0000686 | *Camellia sinensis* | Shrub | |
| CSS0027055 | *Camellia sinensis* | Shrub | |
| CSS0024605 | *Camellia sinensis* | Shrub | |
| CSS0031002 | *Camellia sinensis* | Shrub | |
| CSS0010776 | *Camellia sinensis* | Shrub | removed |
| Haze_11780 | *Corylus avellana* | Shrub | |
| Haze_11201 | *Corylus avellana* | Shrub | |
| Gohir.D01G013400 | *Gossypium hirsutum* | Shrub | |
| Gohir.A01G012200 | *Gossypium hirsutum* | Shrub | |
| Gohir.D09G219800 | *Gossypium hirsutum* | Shrub | |
| Gohir.A09G217400 | *Gossypium hirsutum* | Shrub | |
| Gohir.D11G266600 | *Gossypium hirsutum* | Shrub | |
| Gohir.A11G256800 | *Gossypium hirsutum* | Shrub | |
| Gorai.002G015200 | *Gossypium raimondii* | Shrub | |
| Gorai.006G240800 | *Gossypium raimondii* | Shrub | |

| | | | |
|---|---|---|---|
| Gorai.007G286400 | *Gossypium raimondii* | Shrub | |
| Hma1.2p1_0262F.1_g105390 | *Hydrangea macrophylla* | Shrub | |
| Hma1.2p1_0262F.1_g105420 | *Hydrangea macrophylla* | Shrub | |
| Manes.07G109504 | *Manihot esculenta* | Shrub | |
| Manes.07G109600 | *Manihot esculenta* | Shrub | |
| Manes.07G109900 | *Manihot esculenta* | Shrub | |
| Manes.03G175900 | *Manihot esculenta* | Shrub | |
| Rhsim04G0216200 | *Rhododendron simsii* | Shrub | |
| Rhsim04G0216300 | *Rhododendron simsii* | Shrub | |
| Rhsim05G0149800 | *Rhododendron simsii* | Shrub | |
| Rhsim04G0200000 | *Rhododendron simsii* | Shrub | |
| RcHm_v2.0_Chr1g0318111 | *Rosa chinensis* | Shrub | |
| RcHm_v2.0_Chr1g0318071 | *Rosa chinensis* | Shrub | |
| RcHm_v2.0_Chr4g0446861 | *Rosa chinensis* | Shrub | |
| RcHm_v2.0_Chr4g0446881 | *Rosa chinensis* | Shrub | |
| RcHm_v2.0_Chr4g0446891 | *Rosa chinensis* | Shrub | |
| RcHm_v2.0_Chr4g0415581 | *Rosa chinensis* | Shrub | |
| Sabra05G0073900 | *Salix brachista* | Shrub | |
| PGSC0003DMG400016310 | *Selenicereus undatus* | Shrub | |
| Sc13g0004580 | *Simmondsia chinensis* | Shrub | |
| TWI53G1427 | *Tripterygium wilfordii* | Shrub | |
| TWI53G0907 | *Tripterygium wilfordii* | Shrub | |
| TWI12G1273 | *Tripterygium wilfordii* | Shrub | |
| GSVIVG01018011001 | *Vitis vinifera* | Shrub | |
| Aqoxy5G00438 | *Aquilegia oxysepala* | Perennial | |
| AL1G15300 | *Arabidopsis lyrata* | Perennial | |
| AT1G05460 | *Arabidopsis thaliana* | Perennial | |
| BolC5t29158H | *Brassica oleracea* | Perennial | |
| CAN.G231.13 | *Capsicum annuum* | Perennial | |
| CDE07G0808 | *Ceratophyllum demersum* | Perennial | |
| CDE06G1398 | *Ceratophyllum demersum* | Perennial | removed |
| Migut.G00361 | *Erythranthe guttata* | Perennial | |
| FvH4_4g37100 | *Fragaria vesca* | Perennial | |
| FvH4_4g15972 | *Fragaria vesca* | Perennial | |
| FAN12G3862 | *Fragaria x ananassa* | Perennial | |
| FAN07G0901 | *Fragaria x ananassa* | Perennial | |

| | | | |
|---|---|---|---|
| FAN17G2872 | *Fragaria x ananassa* | Perennial | |
| FAN22G0261 | *Fragaria x ananassa* | Perennial | |
| FAN07G0972 | *Fragaria x ananassa* | Perennial | |
| FAN12G0500 | *Fragaria x ananassa* | Perennial | |
| FAN17G1329 | *Fragaria x ananassa* | Perennial | |
| Lj9C239T5 | *Lonicera japonica* | Perennial | |
| Lj9C239G4 | *Lonicera japonica* | Perennial | |
| Lj6A714G14 | *Lonicera japonica* | Perennial | |
| Lj3A732T73 | *Lonicera japonica* | Perennial | |
| Lj3E732T0 | *Lonicera japonica* | Perennial | removed |
| Lj6C25T8 | *Lonicera japonica* | Perennial | removed |
| Lj9C239T4 | *Lonicera japonica* | Perennial | removed |
| LjContig00222g0014636 | *Lotus japonicus* | Perennial | |
| Mapoly0078s0007 | *Marchantia polymorpha* | Perennial | |
| Nn1g02374 | *Nelumbo nucifera* | Perennial | |
| Nn3g18604 | *Nelumbo nucifera* | Perennial | removed |
| Nitab4.5_0000321g0190 | *Nicotiana tabacum* | Perennial | |
| Nitab4.5_0006783g0030 | *Nicotiana tabacum* | Perennial | |
| Nitab4.5_0002840g0170 | *Nicotiana tabacum* | Perennial | |
| Nitab4.5_0002361g0230 | *Nicotiana tabacum* | Perennial | |
| Os03g0160400 | *Oryza sativa japonica* | Perennial | |
| Sed0017560 | *Sechium edule* | Perennial | |
| Sed0008697 | *Sechium edule* | Perennial | |
| Sed0023639 | *Sechium edule* | Perennial | |
| SMO223G0021 | *Selaginella moellendorffii* | Perennial | |
| SMO223G0037 | *Selaginella moellendorffii* | Perennial | |
| Solyc06g054050.3 | *Solanum lycopersicum* | Perennial | |
| Solyc06g054020.3 | *Solanum lycopersicum* | Perennial | |
| Sopen06g019210 | *Solanum pennellii* | Perennial | |
| Sopen06g019190 | *Solanum pennellii* | Perennial | |
| PGSC0003DMG400016265 | *Solanum tuberosum* | Perennial | |
| TPR.G38022 | *Trifolium pratense* | Perennial | |
| TPR.G37897 | *Trifolium pratense* | Perennial | removed |
| unitig_41.g31654 | *Utricularia gibba* | Perennial | |
| unitig_899.g14832 | *Utricularia gibba* | Perennial | |
| unitig_899.g14831 | *Utricularia gibba* | Perennial | removed |

| | | | |
|---|---|---|---|
| unitig_899.g14833 | *Utricularia gibba* | Perennial | removed |
| HPP92_006123 | *Vanilla planifolia* | Perennial | |
| Aa31LG1G19180 | *Aethionema arabicum* | Annual | |
| Ah.07g204230 | *Amaranthus hybridus* | Annual | |
| arahy.Tifrunner.gnm1.ann1.IF8NHZ | *Arachis hypogaea* | Annual | |
| arahy.Tifrunner.gnm1.ann1.SX262Z | *Arachis hypogaea* | Annual | |
| EL10Ac4g09337 | *Beta vulgaris* | Annual | |
| BcaC05g24383 | *Brassica carinata* | Annual | |
| BcaB02g06857 | *Brassica carinata* | Annual | |
| C05p03810 | *Brassica napus* | Annual | |
| A10p04180 | *Brassica napus* | Annual | |
| A08p08300 | *Brassica napus* | Annual | removed |
| BraA10t42687Z | *Brassica rapa* | Annual | |
| CANSAT78G1480 | *Cannabis sativa* | Annual | |
| CANSAT78G2191 | *Cannabis sativa* | Annual | |
| Carub.0001s0479 | *Capsella rubella* | Annual | |
| CARHR004950 | *Cardamine hirsuta* | Annual | |
| AUR62005129 | *Chenopodium quinoa* | Annual | |
| AUR62000846 | *Chenopodium quinoa* | Annual | |
| Ca_24420_v3 | *Cicer arietinum* | Annual | |
| ClCG06G006340 | *Citrullus lanatus* | Annual | |
| COL.COLO4_33089 | *Corchorus olitorius* | Annual | |
| COL.COLO4_21516 | *Corchorus olitorius* | Annual | |
| COL.COLO4_13640 | *Corchorus olitorius* | Annual | |
| MELO3C012664.2 | *Cucumis melo* | Annual | |
| CsaV3_7G022380 | *Cucumis sativus* | Annual | |
| DCAR_002601 | *Daucus carota* | Annual | |
| DCAR_002600 | *Daucus carota* | Annual | |
| DCAR_023470 | *Daucus carota* | Annual | |
| DCAR_002375 | *Daucus carota* | Annual | |
| DCAR_018493 | *Daucus carota* | Annual | removed |
| ECA236G4457 | *Erigeron canadensis* | Annual | |
| Thhalv10006684m.g | *Eutrema salsugineum* | Annual | |
| Glyma.01G235200 | *Glycine max* | Annual | |
| HanXRQChr02g0033141 | *Helianthus annuus* | Annual | |
| HanXRQChr04g0095851 | *Helianthus annuus* | Annual | removed |

| | | | |
|---|---|---|---|
| HanXRQChr04g0095861 | *Helianthus annuus* | Annual | removed |
| HanXRQChr04g0115171 | *Helianthus annuus* | Annual | removed |
| HanXRQChr04g0115181 | *Helianthus annuus* | Annual | removed |
| Lsat_1_v5_gn_9_80360 | *Lactuca sativa* | Annual | |
| Lalb_Chr16g0376481 | *Lupinus albus* | Annual | |
| Medtr5g006890 | *Medicago truncatula* | Annual | |
| Medtr2g049990 | *Medicago truncatula* | Annual | |
| PSO832G1692 | *Papaver somniferum* | Annual | |
| PSO210G3945 | *Papaver somniferum* | Annual | |
| Peaxi162Scf00131g00027 | *Petunia axillaris* | Annual | |
| Phvul.002G042600 | *Phaseolus vulgaris* | Annual | |
| Pp3c12_8920 | *Physcomitrium patens* | Annual | |
| Pp3c4_21380 | *Physcomitrium patens* | Annual | removed |
| Psat2g190040 | *Pisum sativum* | Annual | |
| SalBow2G0283 | *Salvia bowleyana* | Annual | removed |
| Sp1g04340 | *Schrenkiella parvula* | Annual | |
| SGA_v2.0_scaffold191G40493 | *Striga asiatica* | Annual | |
| THA.LOC104800956 | *Tarenaya hassleriana* | Annual | |
| VMungo1215G2311 | *Vigna mungo* | Annual | |
| VMungo0251G0587 | *Vigna mungo* | Annual | removed |
| VMungo0251G2242 | *Vigna mungo* | Annual | removed |
| Zm00001eb004540 | *Zea mays* | Annual | |
| Zm00001eb403310 | *Zea mays* | Annual | |
| CBR_g44368 | *Chara braunii* | Alga | |
| CBR_g44365 | *Chara braunii* | Alga | |
| CBR_g10879 | *Chara braunii* | Alga | |
| CBR_g34008 | *Chara braunii* | Alga | |
| CBR_g44358 | *Chara braunii* | Alga | removed |
| Cre12.g542450 | *Chlamydomonas reinhardtii* | Alga | |
| Cre15.g641650 | *Chlamydomonas reinhardtii* | Alga | |
| MCO13G517 | *Micromonas commoda* | Alga | removed |
| PRCOL_00005108 | *Prasinoderma coloniale* | Alga | |

**Appendix Table S5.** The results of Fisher exact test to test enrichment or dilution of each life form in each of significantly different cluster.

| Cluster | Life form | The number of target life forms in target cluster | The number of target life forms in all species | p-values | Q-values |
|---|---|---|---|---|---|
| Cluster1 | Tree | 0 | 21 | 1 | 1 |
| | Perennial herb | 2 | 23 | 1 | 1 |
| | Annual herb | 0 | 37 | 0.503 | 1 |
| | Alga | 0 | 4 | 0.00168 | 0.00672 |
| Cluster2 | Tree | 0 | 21 | 0.0581 | 0.116 |
| | Perennial herb | 2 | 23 | 0.719 | 0.959 |
| | Annual herb | 9 | 37 | 0.00846 | 0.0338 |
| | Alga | 0 | 4 | 1 | 1 |
| Cluster3 | Tree | 21 | 21 | 0.032 | 0.128 |
| | Perennial herb | 21 | 23 | 0.499 | 0.499 |
| | Annual herb | 28 | 37 | 0.0665 | 0.233 |
| | Alga | 2 | 4 | 0.109 | 0.145 |

**Appendix Table S6.** The results of Fisher exact test to test enrichment or dilution of each function of the gene family in each of significantly different cluster.

| Cluster | The function group of the gene family | The number of target gene families in target cluster | The number of target gene families in all gene families | p-value | Q-value |
|---|---|---|---|---|---|
| ClusterI | Chromatin formation or chromatin remodeling | 13 | 34 | 0.833 | 1 |
| | DNA modification | 5 | 15 | 1 | 1 |
| | Histon modification | 13 | 28 | 0.183 | 0.691 |
| | Polycomb-group proteins and interacting components | 4 | 13 | 0.77 | 1 |
| | RNA silencing | 8 | 31 | 0.276 | 0.691 |
| ClusterII | Chromatin formation or chromatin remodeling | 20 | 34 | 0.532 | 0.887 |
| | DNA modification | 10 | 15 | 1 | 1 |
| | Histon modification | 15 | 28 | 0.263 | 0.658 |
| | Polycomb-group proteins and interacting components | 9 | 13 | 0.767 | 0.959 |
| | RNA silencing | 23 | 31 | 0.196 | 0.658 |

**Appendix Table S7.** The summary of results of the phylogenetic generalized least squares (PGLS) analyses. The phylogenetic model for each gene family were selected based on AIC value. *a: The estimated value of the phylogenetic correlation parameter in the model: λ in Pagel's lambda model, δ in Pagel's delta model, κ in Pagel's kappa model, α in the Ornstein-Uhlenbeck model. *b: The estimated value of the variance rate $\sigma^2$ in the model.

(A) Tree vs. Annual herb

| Gene Family ID in Dicots PLAZA 5.0 | Coefficient | Standard Error | t-value | p-value | q-value |
|---|---|---|---|---|---|
| HOM05D000010 | -0.155 | 0.079 | -1.962 | 0.053 | 0.466 |
| HOM05D000050 | -0.009 | 0.168 | -0.055 | 0.956 | 0.923 |
| HOM05D000104 | -0.090 | 0.070 | -1.286 | 0.202 | 0.722 |
| HOM05D000144 | -0.088 | 0.097 | -0.914 | 0.364 | 0.823 |
| HOM05D000173 | 0.036 | 0.088 | 0.407 | 0.685 | 0.913 |
| HOM05D000228 | -0.173 | 0.131 | -1.322 | 0.190 | 0.708 |
| HOM05D000234 | -0.013 | 0.102 | -0.123 | 0.902 | 0.917 |
| HOM05D000268 | 0.166 | 0.098 | 1.691 | 0.095 | 0.592 |
| HOM05D000288 | 0.268 | 0.107 | 2.508 | 0.014 | 0.315 |
| HOM05D000319 | -0.076 | 0.099 | -0.766 | 0.446 | 0.823 |
| HOM05D000329 | 0.059 | 0.105 | 0.562 | 0.575 | 0.847 |
| HOM05D000347 | -0.059 | 0.085 | -0.693 | 0.490 | 0.830 |
| HOM05D000399 | -0.176 | 0.196 | -0.900 | 0.371 | 0.823 |
| HOM05D000451 | 0.249 | 0.185 | 1.348 | 0.182 | 0.708 |
| HOM05D000461 | 0.013 | 0.125 | 0.105 | 0.916 | 0.917 |
| HOM05D000515 | 0.096 | 0.120 | 0.800 | 0.426 | 0.823 |
| HOM05D000526 | -0.090 | 0.109 | -0.833 | 0.407 | 0.823 |
| HOM05D000537 | 0.160 | 0.181 | 0.883 | 0.380 | 0.823 |
| HOM05D000572 | -0.034 | 0.107 | -0.320 | 0.750 | 0.913 |
| HOM05D000590 | -0.003 | 0.094 | -0.037 | 0.970 | 0.923 |
| HOM05D000611 | -0.222 | 0.101 | -2.192 | 0.031 | 0.449 |
| HOM05D000688 | 0.060 | 0.136 | 0.439 | 0.662 | 0.913 |
| HOM05D000725 | 0.200 | 0.196 | 1.021 | 0.310 | 0.823 |
| HOM05D000771 | 0.020 | 0.126 | 0.158 | 0.875 | 0.917 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D000809 | 0.062 | 0.094 | 0.661 | 0.511 | 0.834 |
| HOM05D000822 | -0.445 | 0.159 | -2.795 | 0.007 | 0.250 |
| HOM05D000902 | 0.184 | 0.109 | 1.680 | 0.097 | 0.592 |
| HOM05D000912 | -0.135 | 0.158 | -0.858 | 0.393 | 0.823 |
| HOM05D000917 | -0.266 | 0.212 | -1.252 | 0.214 | 0.743 |
| HOM05D000966 | -0.014 | 0.093 | -0.152 | 0.879 | 0.917 |
| HOM05D001065 | -0.044 | 0.132 | -0.335 | 0.739 | 0.913 |
| HOM05D001069 | 0.054 | 0.106 | 0.509 | 0.612 | 0.864 |
| HOM05D001081 | 0.103 | 0.144 | 0.718 | 0.475 | 0.830 |
| HOM05D001100 | -0.017 | 0.084 | -0.204 | 0.839 | 0.917 |
| HOM05D001141 | -0.061 | 0.108 | -0.565 | 0.574 | 0.847 |
| HOM05D001165 | 0.230 | 0.163 | 1.407 | 0.164 | 0.708 |
| HOM05D001201 | 0.378 | 0.166 | 2.273 | 0.026 | 0.421 |
| HOM05D001215 | 0.032 | 0.198 | 0.162 | 0.872 | 0.917 |
| HOM05D001240 | -0.168 | 0.109 | -1.540 | 0.128 | 0.592 |
| HOM05D001290 | -0.104 | 0.102 | -1.020 | 0.311 | 0.823 |
| HOM05D001296 | 0.078 | 0.088 | 0.880 | 0.382 | 0.823 |
| HOM05D001331 | -0.118 | 0.119 | -0.992 | 0.324 | 0.823 |
| HOM05D001451 | -0.103 | 0.094 | -1.101 | 0.274 | 0.823 |
| HOM05D001482 | 0.295 | 0.141 | 2.100 | 0.039 | 0.466 |
| HOM05D001495 | 0.020 | 0.121 | 0.166 | 0.868 | 0.917 |
| HOM05D001587 | -0.044 | 0.102 | -0.432 | 0.667 | 0.913 |
| HOM05D001605 | -0.036 | 0.153 | -0.237 | 0.814 | 0.917 |
| HOM05D001613 | -0.084 | 0.126 | -0.672 | 0.504 | 0.834 |
| HOM05D001642 | -0.439 | 0.444 | -0.990 | 0.325 | 0.823 |
| HOM05D001674 | 0.005 | 0.161 | 0.029 | 0.977 | 0.923 |
| HOM05D001688 | 0.074 | 0.119 | 0.623 | 0.535 | 0.847 |
| HOM05D001734 | -0.174 | 0.184 | -0.947 | 0.347 | 0.823 |
| HOM05D001873 | -0.281 | 0.148 | -1.900 | 0.061 | 0.466 |
| HOM05D001902 | 0.181 | 0.118 | 1.532 | 0.130 | 0.592 |
| HOM05D001937 | -0.311 | 0.189 | -1.652 | 0.103 | 0.592 |
| HOM05D001944 | -0.058 | 0.178 | -0.322 | 0.748 | 0.913 |
| HOM05D001972 | 0.266 | 0.163 | 1.634 | 0.106 | 0.592 |
| HOM05D002127 | -0.215 | 0.164 | -1.316 | 0.192 | 0.708 |
| HOM05D002164 | -0.026 | 0.126 | -0.207 | 0.836 | 0.917 |
| HOM05D002208 | -0.029 | 0.154 | -0.186 | 0.853 | 0.917 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D002300 | 0.009 | 0.140 | 0.067 | 0.946 | 0.923 |
| HOM05D002302 | 0.046 | 0.418 | 0.109 | 0.913 | 0.917 |
| HOM05D002349 | 0.124 | 0.195 | 0.634 | 0.528 | 0.847 |
| HOM05D002404 | -0.490 | 0.246 | -1.991 | 0.050 | 0.466 |
| HOM05D002415 | 0.158 | 0.223 | 0.708 | 0.481 | 0.830 |
| HOM05D002459 | -0.069 | 0.203 | -0.339 | 0.736 | 0.913 |
| HOM05D002658 | 0.394 | 0.161 | 2.450 | 0.017 | 0.315 |
| HOM05D002662 | 0.005 | 0.109 | 0.047 | 0.962 | 0.923 |
| HOM05D002664 | 0.076 | 0.136 | 0.563 | 0.575 | 0.847 |
| HOM05D002720 | -0.296 | 0.146 | -2.027 | 0.046 | 0.466 |
| HOM05D002728 | -0.175 | 0.156 | -1.120 | 0.266 | 0.823 |
| HOM05D002795 | 0.116 | 0.168 | 0.687 | 0.494 | 0.830 |
| HOM05D002863 | -0.821 | 0.232 | -3.537 | 0.001 | 0.039 |
| HOM05D002939 | -0.165 | 0.234 | -0.708 | 0.481 | 0.830 |
| HOM05D003239 | -0.118 | 0.153 | -0.777 | 0.440 | 0.823 |
| HOM05D003289 | 0.135 | 0.191 | 0.705 | 0.483 | 0.830 |
| HOM05D003321 | 0.127 | 0.152 | 0.837 | 0.405 | 0.823 |
| HOM05D003463 | 0.446 | 0.232 | 1.917 | 0.059 | 0.466 |
| HOM05D003609 | 0.253 | 0.163 | 1.551 | 0.125 | 0.592 |
| HOM05D003719 | 0.306 | 0.194 | 1.573 | 0.120 | 0.592 |
| HOM05D003897 | -0.019 | 0.188 | -0.102 | 0.919 | 0.917 |
| HOM05D003901 | -0.128 | 0.165 | -0.774 | 0.441 | 0.823 |
| HOM05D003977 | -0.084 | 0.147 | -0.567 | 0.572 | 0.847 |
| HOM05D004103 | -0.341 | 0.178 | -1.922 | 0.058 | 0.466 |
| HOM05D004111 | -0.063 | 0.153 | -0.412 | 0.681 | 0.913 |
| HOM05D004146 | -0.076 | 0.138 | -0.548 | 0.585 | 0.847 |
| HOM05D004178 | -0.032 | 0.233 | -0.136 | 0.892 | 0.917 |
| HOM05D004180 | -0.530 | 0.339 | -1.564 | 0.122 | 0.592 |
| HOM05D004187 | 0.043 | 0.220 | 0.193 | 0.847 | 0.917 |
| HOM05D004190 | -0.184 | 0.302 | -0.608 | 0.545 | 0.847 |
| HOM05D004205 | -0.255 | 0.250 | -1.021 | 0.310 | 0.823 |
| HOM05D004256 | 0.140 | 0.126 | 1.106 | 0.272 | 0.823 |
| HOM05D004312 | 0.224 | 0.132 | 1.692 | 0.095 | 0.592 |
| HOM05D004365 | -0.059 | 0.191 | -0.308 | 0.759 | 0.913 |
| HOM05D004384 | 0.101 | 0.183 | 0.550 | 0.584 | 0.847 |
| HOM05D004616 | -0.043 | 0.165 | -0.260 | 0.795 | 0.917 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D004718 | 0.162 | 0.179 | 0.907 | 0.367 | 0.823 |
| HOM05D004774 | -0.053 | 0.161 | -0.330 | 0.743 | 0.913 |
| HOM05D004779 | -0.033 | 0.165 | -0.201 | 0.841 | 0.917 |
| HOM05D005030 | -0.636 | 0.175 | -3.633 | 0.001 | 0.039 |
| HOM05D005043 | 0.094 | 0.116 | 0.810 | 0.421 | 0.823 |
| HOM05D005044 | 0.049 | 0.124 | 0.396 | 0.693 | 0.913 |
| HOM05D005087 | -0.052 | 0.147 | -0.354 | 0.724 | 0.913 |
| HOM05D005148 | -0.125 | 0.164 | -0.767 | 0.445 | 0.823 |
| HOM05D005238 | 0.027 | 0.164 | 0.167 | 0.867 | 0.917 |
| HOM05D005294 | 0.253 | 0.239 | 1.062 | 0.292 | 0.823 |
| HOM05D005401 | -0.280 | 0.237 | -1.183 | 0.241 | 0.809 |
| HOM05D005600 | -0.374 | 0.152 | -2.455 | 0.016 | 0.315 |
| HOM05D005631 | -0.068 | 0.128 | -0.532 | 0.597 | 0.852 |
| HOM05D005757 | 0.027 | 0.273 | 0.098 | 0.923 | 0.917 |
| HOM05D005855 | -0.050 | 0.160 | -0.313 | 0.755 | 0.913 |
| HOM05D005914 | -0.177 | 0.182 | -0.973 | 0.334 | 0.823 |
| HOM05D006136 | -0.016 | 0.121 | -0.129 | 0.897 | 0.917 |
| HOM05D006271 | 0.192 | 0.144 | 1.338 | 0.185 | 0.708 |
| HOM05D006316 | 0.082 | 0.227 | 0.359 | 0.720 | 0.913 |
| HOM05D006561 | -0.174 | 0.151 | -1.154 | 0.252 | 0.823 |
| HOM05D006833 | -0.023 | 0.137 | -0.164 | 0.870 | 0.917 |
| HOM05D006922 | 0.167 | 0.121 | 1.377 | 0.173 | 0.708 |
| HOM05D006987 | -0.115 | 0.123 | -0.937 | 0.352 | 0.823 |
| HOM05D007289 | 0.007 | 0.123 | 0.058 | 0.954 | 0.923 |
| HOM05D007494 | 0.587 | 0.578 | 1.015 | 0.313 | 0.823 |

(B) Tree vs. Perennial herb

| Gene Family ID in Dicots PLAZA 5.0 | Coefficient | Standard Error | t-value | p-value | q-value |
|---|---|---|---|---|---|
| HOM05D000010 | -0.202 | 0.081 | -2.491 | 0.015 | 0.243 |
| HOM05D000050 | -0.138 | 0.170 | -0.811 | 0.420 | 0.698 |
| HOM05D000104 | -0.049 | 0.074 | -0.664 | 0.509 | 0.748 |
| HOM05D000144 | -0.095 | 0.098 | -0.966 | 0.337 | 0.616 |
| HOM05D000173 | 0.039 | 0.095 | 0.411 | 0.682 | 0.784 |
| HOM05D000228 | -0.216 | 0.134 | -1.615 | 0.110 | 0.541 |
| HOM05D000234 | -0.048 | 0.106 | -0.448 | 0.656 | 0.784 |
| HOM05D000268 | 0.195 | 0.105 | 1.855 | 0.067 | 0.477 |
| HOM05D000288 | 0.166 | 0.117 | 1.418 | 0.160 | 0.552 |
| HOM05D000319 | 0.003 | 0.107 | 0.024 | 0.981 | 0.888 |
| HOM05D000329 | -0.013 | 0.112 | -0.120 | 0.905 | 0.877 |
| HOM05D000347 | -0.104 | 0.089 | -1.170 | 0.246 | 0.583 |
| HOM05D000399 | -0.117 | 0.212 | -0.553 | 0.582 | 0.748 |
| HOM05D000451 | 0.238 | 0.192 | 1.236 | 0.220 | 0.552 |
| HOM05D000461 | 0.072 | 0.135 | 0.532 | 0.596 | 0.748 |
| HOM05D000515 | 0.047 | 0.125 | 0.373 | 0.710 | 0.784 |
| HOM05D000526 | -0.115 | 0.118 | -0.980 | 0.330 | 0.614 |
| HOM05D000537 | 0.134 | 0.184 | 0.727 | 0.469 | 0.748 |
| HOM05D000572 | -0.099 | 0.112 | -0.882 | 0.380 | 0.672 |
| HOM05D000590 | 0.009 | 0.096 | 0.091 | 0.927 | 0.877 |
| HOM05D000611 | -0.295 | 0.104 | -2.845 | 0.006 | 0.166 |
| HOM05D000688 | 0.065 | 0.141 | 0.460 | 0.646 | 0.783 |
| HOM05D000725 | 0.567 | 0.212 | 2.674 | 0.009 | 0.197 |
| HOM05D000771 | -0.101 | 0.136 | -0.737 | 0.464 | 0.748 |
| HOM05D000809 | -0.126 | 0.097 | -1.299 | 0.198 | 0.552 |
| HOM05D000822 | -0.132 | 0.163 | -0.809 | 0.421 | 0.698 |
| HOM05D000902 | 0.255 | 0.113 | 2.255 | 0.027 | 0.291 |
| HOM05D000912 | -0.343 | 0.159 | -2.156 | 0.034 | 0.335 |
| HOM05D000917 | -0.332 | 0.214 | -1.553 | 0.125 | 0.552 |
| HOM05D000966 | -0.058 | 0.101 | -0.572 | 0.569 | 0.748 |
| HOM05D001065 | -0.220 | 0.134 | -1.634 | 0.106 | 0.541 |
| HOM05D001069 | 0.207 | 0.116 | 1.782 | 0.079 | 0.477 |
| HOM05D001081 | 0.094 | 0.156 | 0.602 | 0.549 | 0.748 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D001100 | 0.013 | 0.090 | 0.143 | 0.887 | 0.872 |
| HOM05D001141 | -0.035 | 0.114 | -0.305 | 0.761 | 0.813 |
| HOM05D001165 | 0.220 | 0.177 | 1.243 | 0.218 | 0.552 |
| HOM05D001201 | 0.198 | 0.180 | 1.097 | 0.276 | 0.583 |
| HOM05D001215 | 0.307 | 0.215 | 1.427 | 0.158 | 0.552 |
| HOM05D001240 | -0.170 | 0.117 | -1.448 | 0.152 | 0.552 |
| HOM05D001290 | -0.063 | 0.111 | -0.565 | 0.574 | 0.748 |
| HOM05D001296 | 0.035 | 0.093 | 0.380 | 0.705 | 0.784 |
| HOM05D001331 | -0.075 | 0.129 | -0.578 | 0.565 | 0.748 |
| HOM05D001451 | -0.149 | 0.101 | -1.479 | 0.143 | 0.552 |
| HOM05D001482 | 0.282 | 0.152 | 1.849 | 0.068 | 0.477 |
| HOM05D001495 | -0.180 | 0.123 | -1.467 | 0.147 | 0.552 |
| HOM05D001587 | -0.041 | 0.108 | -0.380 | 0.705 | 0.784 |
| HOM05D001605 | -0.228 | 0.166 | -1.377 | 0.172 | 0.552 |
| HOM05D001613 | -0.175 | 0.129 | -1.357 | 0.179 | 0.552 |
| HOM05D001642 | -0.500 | 0.445 | -1.123 | 0.265 | 0.583 |
| HOM05D001674 | -0.190 | 0.169 | -1.120 | 0.266 | 0.583 |
| HOM05D001688 | -0.159 | 0.124 | -1.282 | 0.204 | 0.552 |
| HOM05D001734 | -0.195 | 0.196 | -0.995 | 0.323 | 0.614 |
| HOM05D001873 | -0.174 | 0.152 | -1.147 | 0.255 | 0.583 |
| HOM05D001902 | 0.010 | 0.128 | 0.080 | 0.936 | 0.877 |
| HOM05D001937 | -0.360 | 0.198 | -1.816 | 0.073 | 0.477 |
| HOM05D001944 | 0.025 | 0.181 | 0.139 | 0.890 | 0.872 |
| HOM05D001972 | 0.338 | 0.179 | 1.886 | 0.063 | 0.477 |
| HOM05D002127 | -0.223 | 0.177 | -1.254 | 0.214 | 0.552 |
| HOM05D002164 | 0.002 | 0.127 | 0.018 | 0.986 | 0.888 |
| HOM05D002208 | -0.034 | 0.161 | -0.211 | 0.834 | 0.864 |
| HOM05D002300 | -0.002 | 0.152 | -0.014 | 0.989 | 0.888 |
| HOM05D002302 | -0.072 | 0.423 | -0.170 | 0.865 | 0.872 |
| HOM05D002349 | -0.097 | 0.204 | -0.477 | 0.635 | 0.777 |
| HOM05D002404 | -0.370 | 0.250 | -1.480 | 0.143 | 0.552 |
| HOM05D002415 | 0.306 | 0.242 | 1.265 | 0.210 | 0.552 |
| HOM05D002459 | 0.040 | 0.208 | 0.194 | 0.847 | 0.869 |
| HOM05D002658 | 0.410 | 0.167 | 2.463 | 0.016 | 0.243 |
| HOM05D002662 | -0.071 | 0.112 | -0.635 | 0.527 | 0.748 |
| HOM05D002664 | 0.079 | 0.141 | 0.557 | 0.579 | 0.748 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D002720 | -0.048 | 0.148 | -0.324 | 0.747 | 0.813 |
| HOM05D002728 | -0.333 | 0.161 | -2.077 | 0.041 | 0.370 |
| HOM05D002795 | 0.070 | 0.174 | 0.405 | 0.687 | 0.784 |
| HOM05D002863 | -0.818 | 0.234 | -3.492 | 0.001 | 0.043 |
| HOM05D002939 | -0.214 | 0.253 | -0.844 | 0.401 | 0.698 |
| HOM05D003239 | -0.006 | 0.156 | -0.038 | 0.970 | 0.888 |
| HOM05D003289 | 0.194 | 0.206 | 0.939 | 0.351 | 0.630 |
| HOM05D003321 | 0.245 | 0.156 | 1.569 | 0.121 | 0.552 |
| HOM05D003463 | 0.357 | 0.235 | 1.516 | 0.134 | 0.552 |
| HOM05D003609 | 0.122 | 0.179 | 0.678 | 0.500 | 0.748 |
| HOM05D003719 | 0.226 | 0.201 | 1.124 | 0.264 | 0.583 |
| HOM05D003897 | -0.073 | 0.199 | -0.369 | 0.713 | 0.784 |
| HOM05D003901 | -0.103 | 0.179 | -0.573 | 0.568 | 0.748 |
| HOM05D003977 | 0.046 | 0.150 | 0.305 | 0.761 | 0.813 |
| HOM05D004103 | -0.230 | 0.182 | -1.264 | 0.210 | 0.552 |
| HOM05D004111 | 0.024 | 0.165 | 0.148 | 0.883 | 0.872 |
| HOM05D004146 | 0.092 | 0.145 | 0.638 | 0.526 | 0.748 |
| HOM05D004178 | 0.133 | 0.241 | 0.552 | 0.582 | 0.748 |
| HOM05D004180 | -0.196 | 0.360 | -0.544 | 0.588 | 0.748 |
| HOM05D004187 | 0.011 | 0.238 | 0.045 | 0.964 | 0.888 |
| HOM05D004190 | 0.132 | 0.305 | 0.433 | 0.666 | 0.784 |
| HOM05D004205 | -0.413 | 0.252 | -1.640 | 0.105 | 0.541 |
| HOM05D004256 | 0.185 | 0.133 | 1.391 | 0.168 | 0.552 |
| HOM05D004312 | -0.150 | 0.138 | -1.085 | 0.281 | 0.583 |
| HOM05D004365 | 0.111 | 0.196 | 0.567 | 0.572 | 0.748 |
| HOM05D004384 | 0.465 | 0.194 | 2.391 | 0.019 | 0.243 |
| HOM05D004616 | -0.045 | 0.171 | -0.264 | 0.792 | 0.837 |
| HOM05D004718 | 0.154 | 0.186 | 0.826 | 0.411 | 0.698 |
| HOM05D004774 | 0.466 | 0.165 | 2.817 | 0.006 | 0.166 |
| HOM05D004779 | -0.096 | 0.179 | -0.533 | 0.595 | 0.748 |
| HOM05D005030 | -0.690 | 0.188 | -3.674 | 0.000 | 0.043 |
| HOM05D005043 | -0.019 | 0.128 | -0.152 | 0.880 | 0.872 |
| HOM05D005044 | -0.088 | 0.135 | -0.649 | 0.518 | 0.748 |
| HOM05D005087 | -0.013 | 0.155 | -0.086 | 0.932 | 0.877 |
| HOM05D005148 | -0.205 | 0.171 | -1.197 | 0.235 | 0.576 |
| HOM05D005238 | -0.091 | 0.172 | -0.531 | 0.597 | 0.748 |

| | | | | | |
|---|---|---|---|---|---|
| HOM05D005294 | 0.000 | 0.242 | -0.001 | 0.999 | 0.890 |
| HOM05D005401 | -0.422 | 0.241 | -1.751 | 0.084 | 0.477 |
| HOM05D005600 | -0.292 | 0.165 | -1.768 | 0.081 | 0.477 |
| HOM05D005631 | -0.056 | 0.141 | -0.398 | 0.691 | 0.784 |
| HOM05D005757 | -0.183 | 0.296 | -0.617 | 0.539 | 0.748 |
| HOM05D005855 | -0.171 | 0.162 | -1.054 | 0.295 | 0.597 |
| HOM05D005914 | -0.251 | 0.189 | -1.325 | 0.189 | 0.552 |
| HOM05D006136 | 0.147 | 0.133 | 1.107 | 0.272 | 0.583 |
| HOM05D006271 | 0.014 | 0.156 | 0.087 | 0.931 | 0.877 |
| HOM05D006316 | 0.051 | 0.232 | 0.218 | 0.828 | 0.864 |
| HOM05D006561 | -0.162 | 0.155 | -1.045 | 0.299 | 0.597 |
| HOM05D006833 | -0.070 | 0.143 | -0.488 | 0.627 | 0.776 |
| HOM05D006922 | 0.122 | 0.124 | 0.979 | 0.331 | 0.614 |
| HOM05D006987 | -0.315 | 0.133 | -2.370 | 0.020 | 0.243 |
| HOM05D007289 | -0.134 | 0.132 | -1.019 | 0.311 | 0.610 |
| HOM05D007494 | -0.724 | 0.584 | -1.239 | 0.219 | 0.552 |

(C) Model parameter estimation

| Gene Family ID in Dicots PLAZA 5.0 | parameter (*a) | Sigma squared (*b) | Log likelihood | AIC |
|---|---|---|---|---|
| HOM05D000010 | 0.906 | 0.337 | -3.598 | 17.196 |
| HOM05D000050 | 3.000 | 2.377 | -61.670 | 133.339 |
| HOM05D000104 | 0.665 | 0.153 | 1.167 | 7.666 |
| HOM05D000144 | 10.924 | 2.341 | -10.095 | 30.191 |
| HOM05D000173 | 66.152 | 13.171 | -17.701 | 45.401 |
| HOM05D000228 | 15.648 | 6.186 | -38.434 | 86.869 |
| HOM05D000234 | 26.394 | 6.759 | -24.619 | 59.238 |
| HOM05D000268 | 0.448 | 0.233 | -28.578 | 67.155 |
| HOM05D000288 | 0.000 | 0.205 | -38.421 | 86.843 |
| HOM05D000319 | 66.152 | 16.792 | -27.415 | 64.831 |
| HOM05D000329 | 45.707 | 12.867 | -30.604 | 71.207 |
| HOM05D000347 | 0.734 | 0.252 | -13.309 | 36.618 |
| HOM05D000399 | 66.152 | 65.798 | -82.044 | 174.088 |
| HOM05D000451 | 23.235 | 19.365 | -71.122 | 152.244 |
| HOM05D000461 | 66.152 | 26.729 | -46.010 | 102.020 |
| HOM05D000515 | 0.770 | 0.537 | -40.262 | 90.524 |
| HOM05D000526 | 66.152 | 20.288 | -34.981 | 79.963 |
| HOM05D000537 | 0.996 | 3.107 | -64.417 | 138.834 |
| HOM05D000572 | 27.081 | 7.677 | -28.821 | 67.641 |
| HOM05D000590 | 0.904 | 0.469 | -17.270 | 44.541 |
| HOM05D000611 | 0.927 | 0.602 | -22.701 | 55.402 |
| HOM05D000688 | 0.824 | 0.782 | -49.392 | 108.783 |
| HOM05D000725 | 66.152 | 65.832 | -82.065 | 174.129 |
| HOM05D000771 | 66.152 | 27.231 | -46.755 | 103.510 |
| HOM05D000809 | 17.688 | 3.717 | -14.167 | 38.334 |
| HOM05D000822 | 0.915 | 1.415 | -59.386 | 128.772 |
| HOM05D000902 | 0.823 | 0.501 | -31.644 | 73.288 |
| HOM05D000912 | 8.404 | 5.004 | -47.167 | 104.335 |
| HOM05D000917 | 6.194 | 7.450 | -69.743 | 149.485 |
| HOM05D000966 | 66.152 | 14.813 | -22.400 | 54.800 |
| HOM05D001065 | 0.988 | 1.490 | -40.203 | 90.405 |
| HOM05D001069 | 0.000 | 0.201 | -37.625 | 85.250 |
| HOM05D001081 | 66.152 | 35.701 | -57.588 | 125.175 |

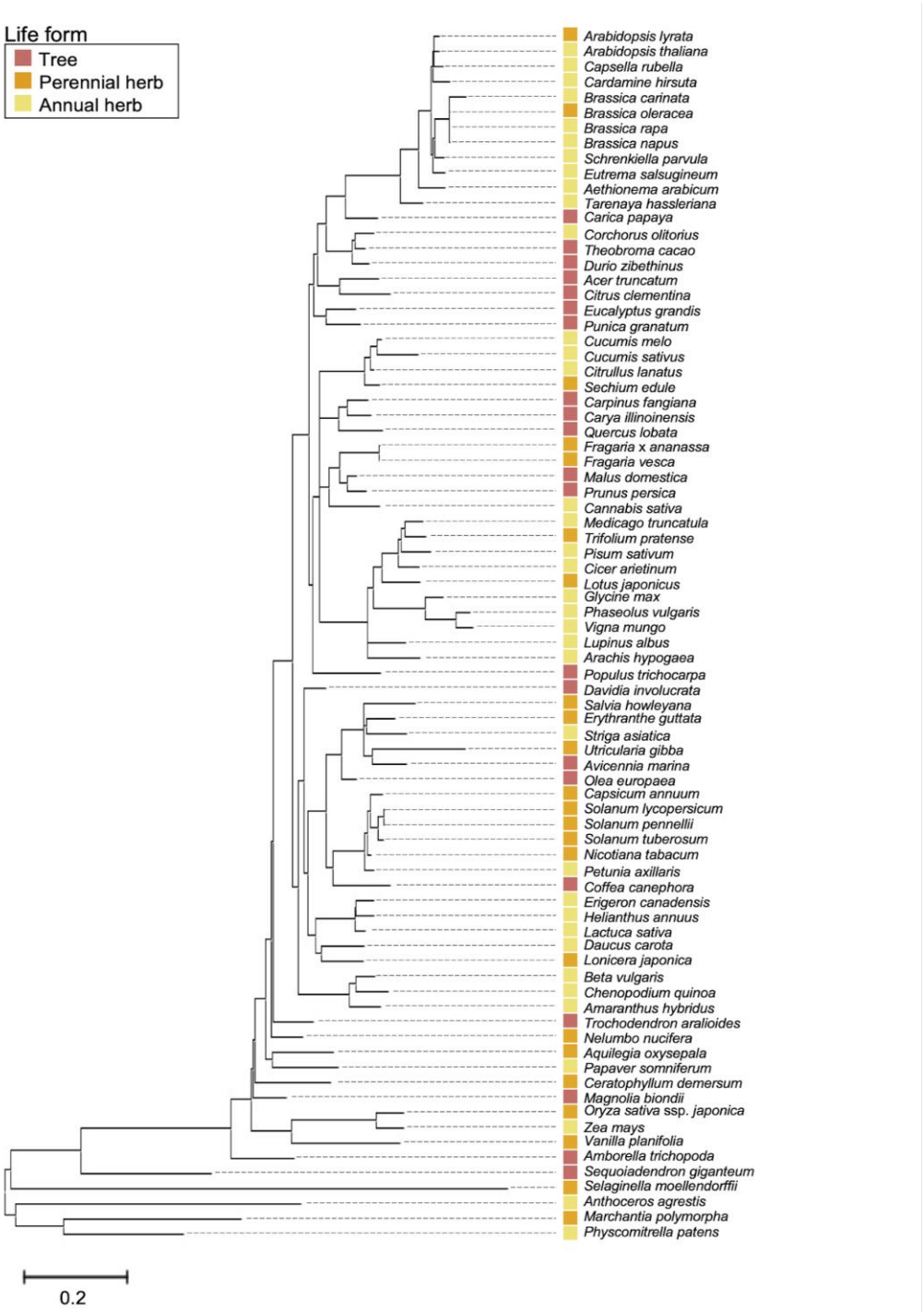
| | | | | |
|---|---|---|---|---|
| HOM05D001100 | 0.470 | 0.176 | -16.298 | 42.596 |
| HOM05D001141 | 34.267 | 10.101 | -31.440 | 72.879 |
| HOM05D001165 | 66.152 | 45.942 | -67.676 | 145.351 |
| HOM05D001201 | 66.152 | 47.512 | -69.020 | 148.040 |
| HOM05D001215 | 66.152 | 67.613 | -83.132 | 176.264 |
| HOM05D001240 | 50.552 | 15.562 | -34.471 | 78.942 |
| HOM05D001290 | 66.152 | 17.956 | -30.096 | 70.192 |
| HOM05D001296 | 0.600 | 0.222 | -18.238 | 46.476 |
| HOM05D001331 | 66.152 | 24.471 | -42.480 | 94.961 |
| HOM05D001451 | 0.378 | 0.201 | -25.765 | 61.530 |
| HOM05D001482 | 66.152 | 34.006 | -55.642 | 121.285 |
| HOM05D001495 | 12.963 | 4.318 | -29.742 | 69.484 |
| HOM05D001587 | 0.618 | 0.305 | -29.717 | 69.434 |
| HOM05D001605 | 66.152 | 40.186 | -62.321 | 134.641 |
| HOM05D001613 | 17.579 | 6.523 | -36.866 | 83.733 |
| HOM05D001642 | -3.969 | 296.761 | -132.435 | 274.871 |
| HOM05D001674 | 28.454 | 18.428 | -62.110 | 134.221 |
| HOM05D001688 | 0.767 | 0.523 | -39.449 | 88.898 |
| HOM05D001734 | 38.731 | 33.544 | -75.013 | 160.025 |
| HOM05D001873 | 0.960 | 1.528 | -51.805 | 113.611 |
| HOM05D001902 | 66.152 | 24.060 | -41.802 | 93.603 |
| HOM05D001937 | 28.843 | 25.639 | -74.842 | 159.683 |
| HOM05D001944 | 0.333 | 0.402 | -65.152 | 140.304 |
| HOM05D001972 | 0.000 | 0.477 | -72.170 | 154.340 |
| HOM05D002127 | 66.152 | 46.073 | -67.789 | 145.579 |
| HOM05D002164 | 0.000 | 0.068 | -42.691 | 95.383 |
| HOM05D002208 | 27.667 | 16.236 | -58.030 | 126.060 |
| HOM05D002300 | 66.152 | 33.862 | -55.472 | 120.943 |
| HOM05D002302 | 3.000 | 14.707 | -134.567 | 279.133 |
| HOM05D002349 | 24.647 | 23.143 | -76.227 | 162.453 |
| HOM05D002404 | 11.878 | 16.461 | -85.785 | 181.569 |
| HOM05D002415 | 66.152 | 85.568 | -92.553 | 195.106 |
| HOM05D002459 | 0.970 | 3.070 | -76.492 | 162.985 |
| HOM05D002658 | 21.532 | 13.435 | -59.078 | 128.155 |
| HOM05D002662 | 0.918 | 0.675 | -29.080 | 68.159 |
| HOM05D002664 | 0.775 | 0.694 | -49.961 | 109.922 |

| | | | | |
|---|---|---|---|---|
| HOM05D002720 | 0.111 | 0.131 | -52.083 | 114.167 |
| HOM05D002728 | 18.017 | 10.353 | -54.547 | 119.095 |
| HOM05D002795 | 0.839 | 1.239 | -65.861 | 141.723 |
| HOM05D002863 | 7.182 | 9.737 | -77.375 | 164.749 |
| HOM05D002939 | 66.152 | 93.830 | -96.240 | 202.479 |
| HOM05D003239 | 0.961 | 1.637 | -54.188 | 118.377 |
| HOM05D003289 | 56.564 | 53.588 | -79.729 | 169.459 |
| HOM05D003321 | 0.910 | 1.273 | -55.980 | 121.959 |
| HOM05D003463 | 9.430 | 11.931 | -79.095 | 168.190 |
| HOM05D003609 | 0.000 | 0.479 | -72.330 | 154.661 |
| HOM05D003719 | 0.813 | 1.548 | -77.975 | 165.949 |
| HOM05D003897 | 33.583 | 30.089 | -75.827 | 161.654 |
| HOM05D003901 | 66.152 | 46.827 | -68.439 | 146.877 |
| HOM05D003977 | 0.990 | 1.920 | -48.996 | 107.992 |
| HOM05D004103 | 0.980 | 2.539 | -65.014 | 140.029 |
| HOM05D004111 | 52.054 | 31.586 | -61.694 | 133.388 |
| HOM05D004146 | 0.659 | 0.588 | -53.171 | 116.342 |
| HOM05D004178 | 0.819 | 2.253 | -92.324 | 194.647 |
| HOM05D004180 | 0.518 | 2.980 | -126.940 | 263.881 |
| HOM05D004187 | 58.810 | 74.021 | -91.189 | 192.378 |
| HOM05D004190 | 0.005 | 0.397 | -112.675 | 235.350 |
| HOM05D004205 | 3.000 | 5.231 | -93.222 | 196.444 |
| HOM05D004256 | 0.638 | 0.480 | -46.527 | 103.054 |
| HOM05D004312 | 26.269 | 11.386 | -45.647 | 101.293 |
| HOM05D004365 | 0.959 | 2.535 | -72.155 | 154.311 |
| HOM05D004384 | 0.566 | 0.921 | -77.216 | 164.433 |
| HOM05D004616 | 0.783 | 1.042 | -65.342 | 140.684 |
| HOM05D004718 | 23.878 | 18.685 | -68.758 | 147.515 |
| HOM05D004774 | 18.268 | 11.158 | -57.094 | 124.189 |
| HOM05D004779 | 66.152 | 47.103 | -68.674 | 147.348 |
| HOM05D005030 | 45.998 | 36.348 | -71.908 | 153.815 |
| HOM05D005043 | 0.000 | 0.244 | -45.333 | 100.666 |
| HOM05D005044 | 66.152 | 26.563 | -45.761 | 101.522 |
| HOM05D005087 | 30.962 | 16.842 | -55.521 | 121.041 |
| HOM05D005148 | 26.726 | 17.736 | -62.775 | 135.550 |
| HOM05D005238 | 0.685 | 0.860 | -66.405 | 142.810 |

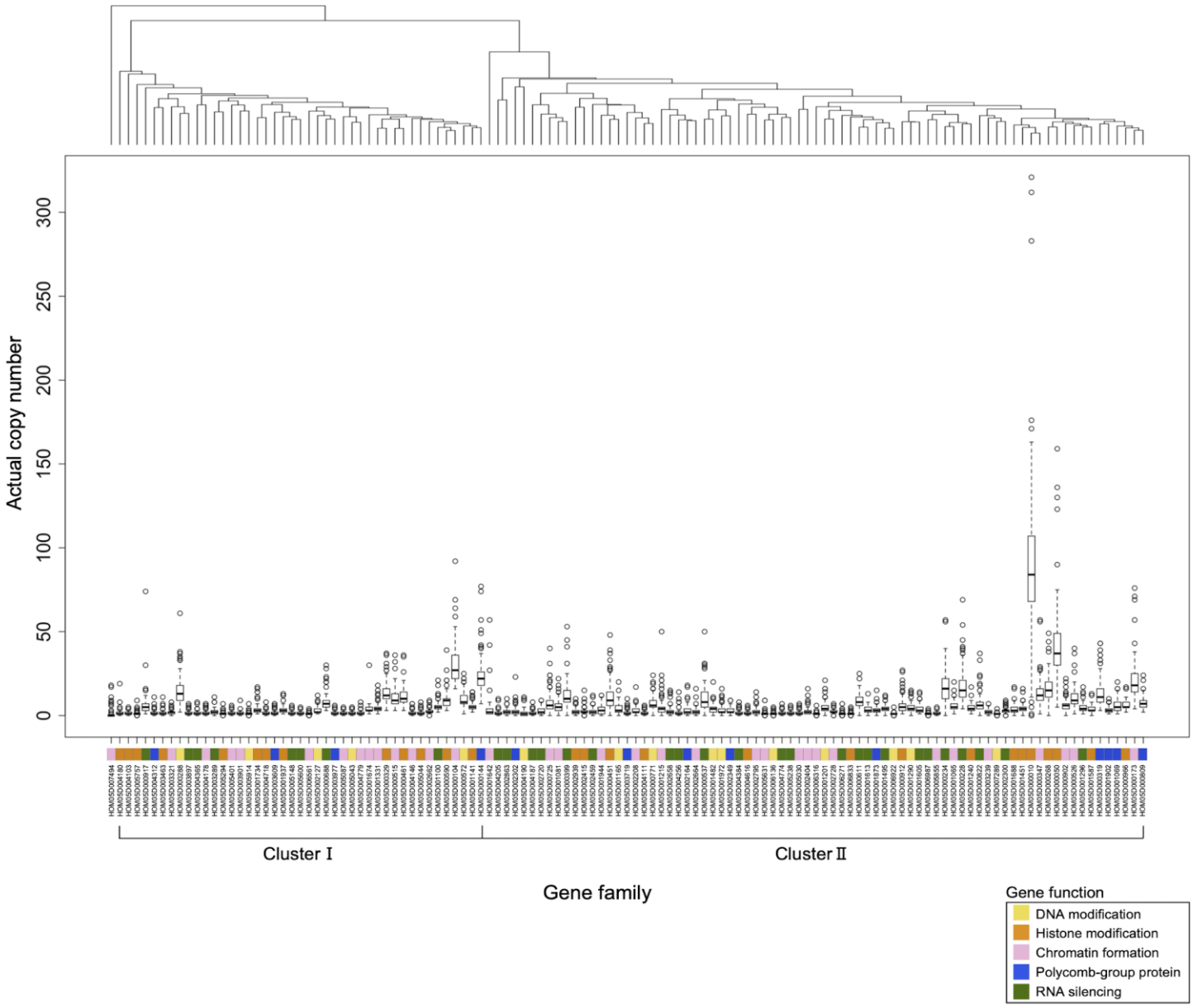
| | | | | |
|---|---|---|---|---|
| HOM05D005294 | 10.082 | 13.302 | -81.722 | 173.443 |
| HOM05D005401 | 12.660 | 16.255 | -83.455 | 176.911 |
| HOM05D005600 | 66.152 | 39.823 | -61.958 | 133.916 |
| HOM05D005631 | 0.000 | 0.295 | -52.894 | 115.788 |
| HOM05D005757 | 63.820 | 123.740 | -108.663 | 227.326 |
| HOM05D005855 | 9.872 | 5.901 | -49.764 | 109.527 |
| HOM05D005914 | 24.682 | 20.021 | -70.381 | 150.763 |
| HOM05D006136 | 0.000 | 0.264 | -48.528 | 107.057 |
| HOM05D006271 | 66.152 | 35.603 | -57.477 | 124.955 |
| HOM05D006316 | 13.766 | 16.313 | -81.133 | 172.266 |
| HOM05D006561 | 0.907 | 1.231 | -55.307 | 120.614 |
| HOM05D006833 | 0.800 | 0.752 | -50.514 | 111.029 |
| HOM05D006922 | 0.964 | 1.059 | -35.783 | 81.565 |
| HOM05D006987 | 54.460 | 21.435 | -44.499 | 98.998 |
| HOM05D007289 | 50.091 | 19.451 | -43.735 | 97.470 |
| HOM05D007494 | 3.000 | 28.100 | -160.465 | 330.930 |

**Supplementary Table S8.** The results of the phylogenetic generalized least squares (PGLS) analyses on dataset with angiosperm species. The phylogenetic model for each gene family were selected based on AIC value. *a: The estimated value of the phylogenetic correlation parameter α in the Ornstein-Uhlenbeck model. *b: The estimated value of the variance rate σ2 in the OU model.

| Symbol of gene family | Trees vs. Annual herbs | | | | | Trees vs. Perennial herbs | | | | | Phylogenetic model | Parameter | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | *t*-value | P-value | Q-value | Coefficient | Standard Error | *t*-value | P-value | Q-value | | | |
| *BRU1/TSK/MGO3* | –0.545 | 0.168 | –3.24 | 0.00183 | 0.0526 | –0.619 | 0.184 | –3.36 | 0.00124 | 0.0586 | Ornstein-Uhlenbech | 62.91 a | 44.01 b |
| *SDE3* | –0.828 | 0.239 | –3.47 | 0.00088 | 0.0382 | –0.839 | 0.241 | –3.48 | 0.00085 | 0.0586 | Ornstein-Uhlenbech | 5.17 a | 8.21 b |

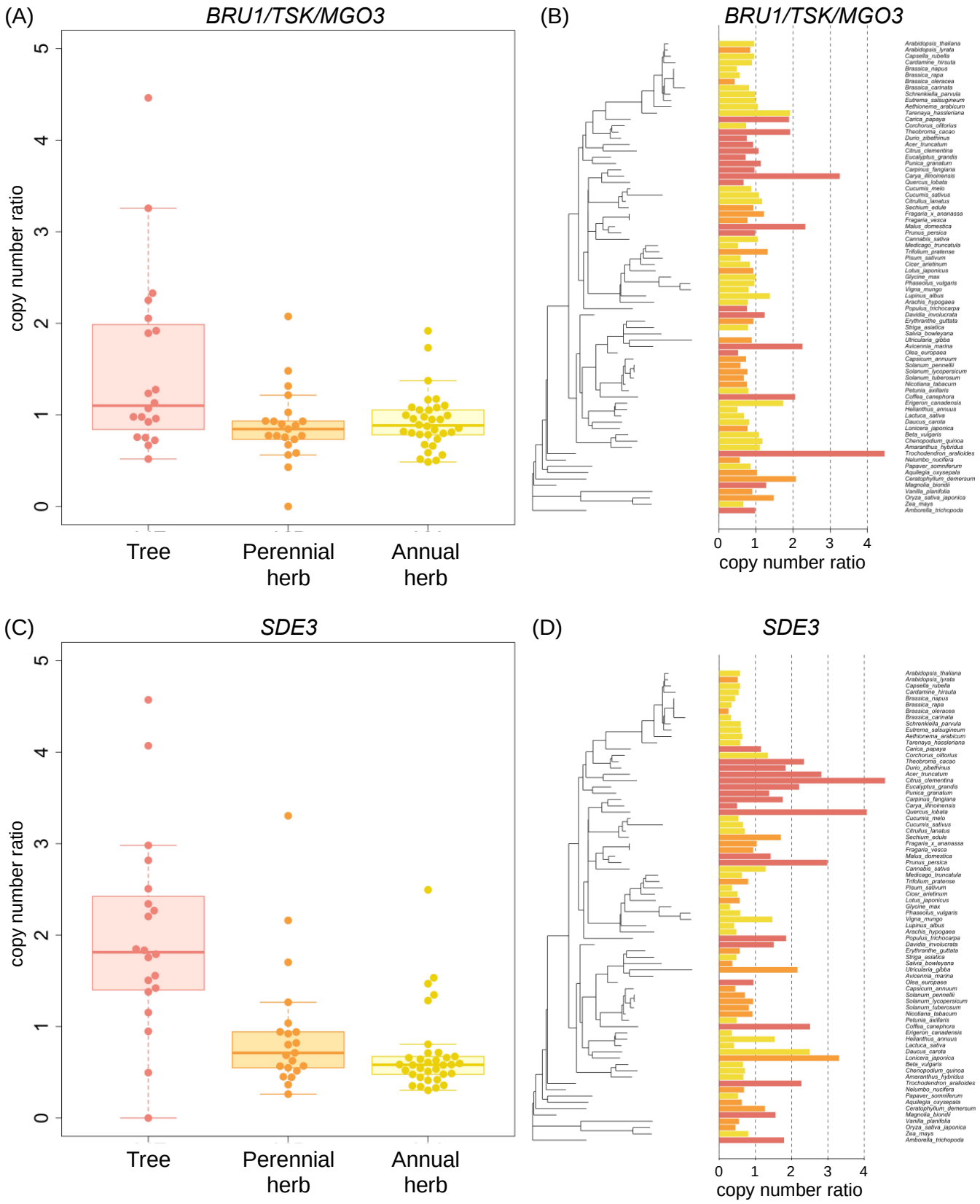



**Appendix Figure S1.** The phylogenetic tree of species for analyses considering the phylogenetic relationships. There were 80 species including 21 tree species, 23 perennial herb species, 36 annual herb species.

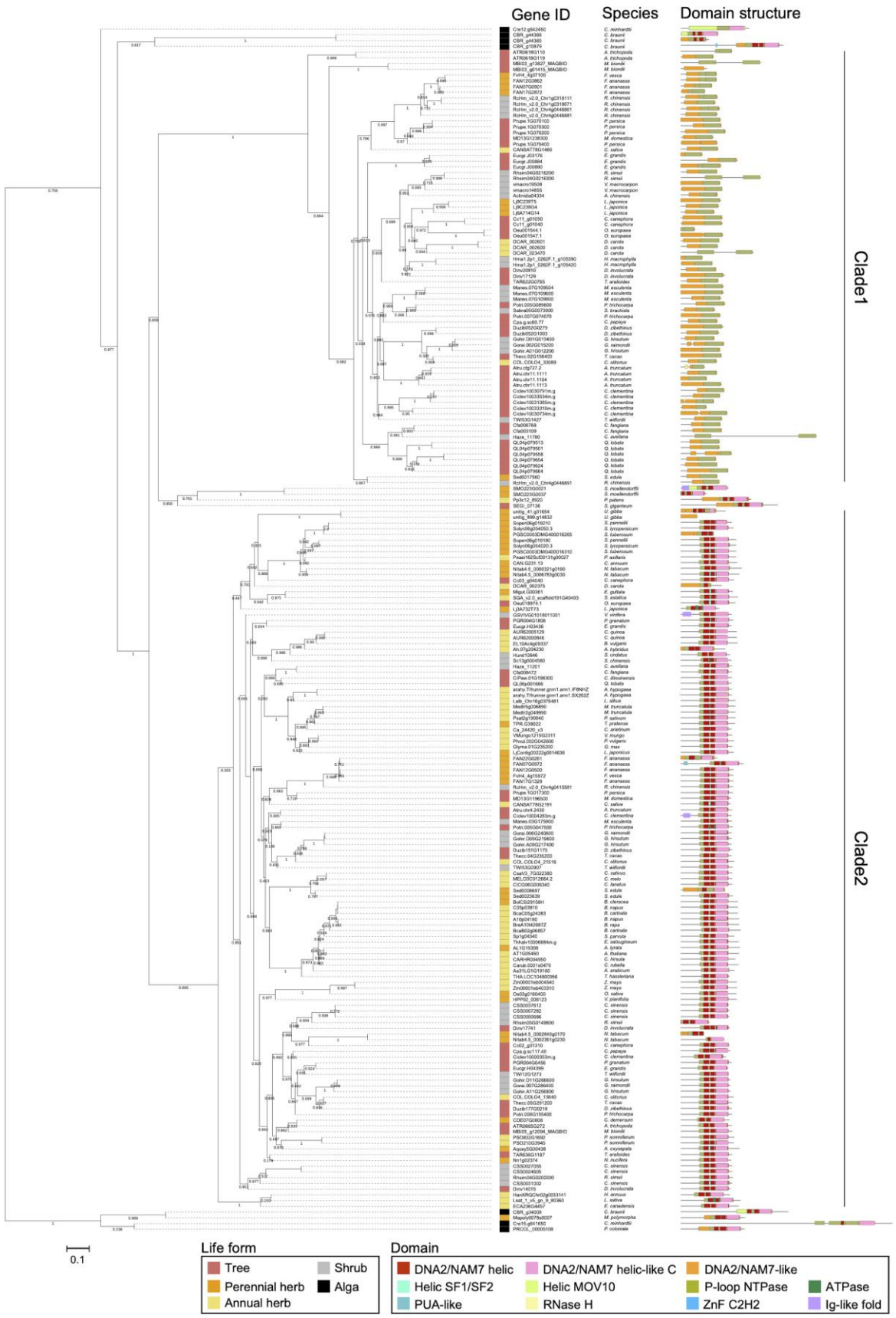



**Appendix Figure S2.** The actual copy number of 121 gene families associated with epigenetic regulation. The gene family IDs in Dicots PLAZA 5.0 database were shown on the horizontal axis. The gene families were ordered according to the result of hierarchical clustering, and the dendrogram was shown above the plot. The order of gene families corresponded to the order of gene families in Figure 1 (A). Each gene family was categorized into one of five groups: DNA modification, Histone modification, Chromatin formation, Polycomb-group proteins, RNA silencing. The

horizontal line inside each box shows the median, and the length of box shows the interquartile range (range between the 25th to 75th percentiles). The whiskers indicate points within 1.5 times the interquartile rage. The points beyond the whisker range indicated the outliers.


(A)
BRU1/TSK/MGO3
copy number ratio
0
1
2
3
4
5
Tree
Perennial herb
Annual herb
(B)
BRU1/TSK/MGO3
0
1
2
3
4
copy number ratio
(C)
SDE3
copy number ratio
0
1
2
3
4
5
Tree
Perennial herb
Annual herb
(D)
SDE3
0
1
2
3
4
copy number ratio

**Supplementary Figure S3.** Results of the phylogenetic generalized least squares (PGLS) analysis on dataset with angiosperm species. The copy number ratios of the *BRU1/TSK/MGO3* gene family (A) and *SDE3* gene family (C) in different life forms. The horizontal line inside each box shows the median, and the length of the box shows the interquartile range (range between the 25th and 75th percentiles). The whiskers indicate points within 1.5 times the interquartile range. The points beyond the whisker range indicate the outliers. Phylogenetic relationships of copy number ratio of the *BRU1/TSK/MGO3* gene family (B) and *SDE3* gene family (D). The color of each bar indicates the life form of the species.


Gene ID
Species
Domain structure
Clade1
Clade2
0.1
Life form
Tree
Shrub
Perennial herb
Alga
Annual herb
Domain
DNA2/NAM7 helic
DNA2/NAM7 helic-like C
DNA2/NAM7-like
Helic SF1/SF2
Helic MOV10
P-loop NTPase
ATPase
PUA-like
RNase H
ZnF C2H2
Ig-like fold

**Supplementary Figure S4.** The phylogenetic tree of *SDE3* genes with protein domain structures constructed using the tree explorer tool in Dicots PLAZA 5.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_dicots/gene_families/explore_trees/HOM05D002863). There were 210 genes within 95 species in the phylogenetic tree. Gene ID of each *SDE3* gene in Dicots PLAZA 5.0 are represented. Species names indicate the species that have the gene, and rectangles to the left of species names indicate the life forms of the species. The numbers under each branch of the phylogenetic tree indicate support values. Protein domains are illustrated by colored: DNA2/NAM7 helic, DNA2/NAM7 helicase, helicase domain; DNA2/NAM7 helic-like C, DNA2/NAM7 helicase-like, C-terminal; DNA2/NAM7-like, DNA2/NAM7-like helicase; Helic SF1/SF2, Helicase superfamily 1/2, ATP-binding domain; Helic MOV10, Helicase MOV-10; P-loop NTPase, P-loop containing nucleoside triphosphate hydrolase; ATPase, AAA+ ATPase domain; PUA-like, PUA-like superfamily; RNase H, Ribonuclease H domain; ZnF C2H2, Zinc finger C2H2-type; Ig-like fold, Immunoglobulin-like fold.

# Chapter 3: Seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation in *Quercus glauca* and *Lithocarpus edulis* under natural conditions

The study in this chapter, done in collaboration with Professor Akiko Satake, is in preparation.

**ABSTRACT**

Living organisms are exposed many types of stresses including biotic and abiotic stresses. To suppress damage due to stresses and maintain to survive for a long time, it is necessary to respond appropriately to stresses that change over time. In the present study, to examine and compare the seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation, we analyzed time-series transcriptome data collected throughout about two years from individuals of different tree species, *Quercus glauca* and *Lithocarpus edulis*, growing in natural environments. The present study demonstrated similar and different seasonal expression dynamics of DNA repair genes and epigenetic regulatory genes among species. Results of the present study suggest that a large number of genes associated with DNA repair and epigenetic regulation exhibit similar seasonal expression patterns among species. In addition, genes with different seasonal expression dynamics are associated with multiple functions and involved in plant development, growth, and reproduction, which is likely to reflect the difference in vegetative and reproductive schedules among species.

**INTRODUCTION**

Living organisms are exposed many types of exogenous stresses (i.e., ultraviolet [UV] radiation, high/low temperature, pathogen infection), and such stresses can cause damage and disrupt homeostasis. The types and amount of stress vary according to seasons (i.e., UV radiation is high in summer but low in winter [Beckmann et al., 2014]). Therefore, in order to suppress damage and maintain homeostasis for a long time, it is necessary for long-lived organisms to respond appropriately to stresses that change over time.

A growing number of studies have revealed that DNA repair and epigenetic regulation have an essential role in genome integrity and normal gene expression,

resulting in maintaining homeostasis under stresses. UV radiation cause DNA damage such as cyclobutane pyrimidine dimers, and such damage can be repaired by nucleotide excision repair (Sinha & Häder, 2002). Reactive oxygen species are generated through metabolic reactions in mitochondria, chloroplasts and peroxisomes in plants (Foyer & Noctor, 2003) and induce oxidative DNA damage such as single- and double-strand breaks (Roldán-Arjona & Ariza, 2009). DNA double-strand breaks can be repaired by two different repair pathways: homologous recombination repair (Puchta, 2005) or non-homologous end-joining repair (Lees-Miller & Meek, 2003). Histone modification and chromatin remodeling are required for regulation, and the regulation of genes involved in stress response under stress conditions often depends on histone modification and chromatin remodeling (Chinnusamy & Zhu, 2009; Kim et al., 2010). RNA silencing inhibits replications of exogenous genetic elements such as viral genes and plays an important role in protection against viruses (Al-Kaff et al., 1998; Ruiz-Ferrer & Voinnet, 2009).

Although many studies have explored expressions and functions of DNA repair and epigenetic regulatory genes in stress response, most of studies have been performed under controlled laboratory conditions with a constant environment. However, organisms live in natural environments with various types of stresses that change over time. To understand how long-lived trees respond stresses and survive under natural environments, it is necessary to monitor expressions of genes associated with DNA repair and epigenetic regulation in individuals growing under natural conditions for long period. In addition, comparisons of seasonal expression dynamics of DNA repair and epigenetic regulatory genes among different species under similar conditions could reveal similarities and differences in responses to stresses among species. Therefore, in the present study, to

examine and compare the seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation and functions of genes with similar or different seasonal expression pattern among species, we analyzed time-series transcriptome data collected throughout about two years from individuals of different tree species, *Quercus glauca* and *Lithocarpus edulis*, growing in natural environments.

## MATERIALS AND MEYHODS

### Study species and study site

*Quercus glauca* and *Lithocalpus edulis* are evergreen tree species. Flowers are self-incompatible and wind-pollinated in *Q. glauca* while they are animal pollinated in *L. edulis*. *Q. glauca* usually start to bloom in April and fruit in the autumn in the same year of anthesis. *L. edulis* begins flowering in June and fruit in the second year after flowering. This fruiting habit is known as biannual fruiting (Borgardt & Nixon, 2003). The scientific names and characteristics are shown in Table 1.

The study site is in the biodiversity reserve of Ito campus of Kyushu University (33°35′ 47.5″ N, 130°12′ 50.0″ E) situated in Fukuoka, southern Japan. The biodiversity reserve of Ito campus occupies an area of about 37 ha at an elevation from 20 to 57 m a.s.l. Mean annual precipitation and temperature near the site were 1677.0 mm and 16.1 °C, respectively (1981–2010; Meteorological Observation System at the NARO Hokkaido Agricultural Research Center).

We collected a pair of a leaf and a bud from each of three current-year shoots per tree every month from April 2017 to March 2019. Samples were taken from the sun-exposed crown (approximately 4m from the ground) using long pruning shears from 11:30 to 12:30 h. For each pair of leaf and bud samples, 0.1–0.3g of leaves and bud tissue

were preserved in a 2ml micro tube containing 1.5ml of RNA stabilizing reagent (RNAlater; Ambion, Austin, TX, USA) immediately after harvesting. Samples were transferred to the laboratory within 3hr after sampling and stored at 4°C overnight and then stored at −20°C until RNA extraction. During the transport to the laboratory, samples were *kept in a cooler box* with ice to maintain low temperature.

The mean (±SD) height and diameter at breast height (DBH) of three individuals were 11.7 m (±2.5) and 36.0 cm (±10.2), respectively.

**RNA extraction**

The extraction of total RNA was performed in accordance with the method described by previous study. RNA was extracted independently from leaf and bud samples from three different branches and pooled at each time point. RNA integrity was examined using the Agilent RNA 6000 Nano kit on a 2100 Bioanalyzer (Agilent Technologies), while the RNA yield was determined on a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific).

**Generation of transcriptome next-generation sequencing (NGS) data**

We obtained transcriptome data from our samples to design DNA microarray probes. We used 8 samples collected monthly from one individual at the study site from May to December 2017 for *Q. glauca* and June to December 2017 for *L. edulis* (Appendix Table S1). Five to six micrograms of total RNA extracted from leaf and bud of each sample was sent to Macrogen (South Korea) where a cDNA library was prepared with Illumina TruSeq Sample Prep Kit and paired-end transcriptome sequencing was conducted using the Illumina Hiseq2000 or NovaSeq6000 sequencer (Illumina, San Diego, CA, USA) for

each sample. A total of 299 and 313 million 100-bp paired-end reads were obtained for each species. The resulting reads shorter than 50 bp were discarded. De novo transcriptome assembly was conducted using Trinity (Ver. 2.0.6). Read quality analysis was performed on the raw data using FastQC v0.11.7 (http://bioinformatics.babraham.ac.uk/projects/fastqc/). Quality trimming and adapter clipping were performed using Trimmomatic version 0.38 (Bolger, Lohse & Usadel, 2014), trimming trailing bases below the average quality 15, minimum length 36 and clipping Illumina adapters. The resulting reads shorter than 50 bp were discarded. De novo transcriptome assembly was conducted using Trinity (Ver. 2.0.6).

**Probe design for DNA microarray**

For custom microarray slides, we used the assembled sequences of the transcripts generated by NGS described above. We selected the assembled sequences for array design based on two steps. We first extracted transcript sequences that showed high homology against *Arabidopsis thaliana* (%Identity >= 40%, qcovhsp >= 40%) by BLASTX searches for each species. For each extracted transcript sequence, top hit *A. thaliana* gene ID was selected. If multiple transcript sequences were annotated for the same *A. thaliana* gene ID, the transcript sequence showing the longest annotation was selected. As a result, we obtained 19,290 and 19,426 transcript sequences for *Q. glauca* and *L. edulis*, respectively. At the second step, we extracted transcript sequences that were eliminated from the homology selection but sequence homology to *F. crenata* transcript sequences used for DNA microarray (Sateke et al. 2019) is high (%Identity >= 60%, qcovhsp >= 60%, e-value cut-off: $10^{-5}$) by BLASTX searches for each species. From the selection of step 2, we obtained 3,474 and 4,357 transcript sequences for *Q.*

*glauca* and *L. edulis*, respectively. We pooled these transcript sequences for each species, and designed the array using the e-array portal for array design hosted by Agilent (https://earray.chem.agilent.com/earray/) based on the total of 22,765 and 23,784 transcript sequences for *Q. glauca* and *L. edulis*, respectively. Two probes were designed for each transcript sequences. After removing probes with the same sequence, 42,121 and 42,436 probes were installed in the 8×60K array format.

**Microarray analysis**

One hundred nanograms of total RNA extracted from leaf and bud of each sample was amplified, labeled, and hybridized to a 60K Agilent 60-mer oligomicroarray, in accordance with the manufacturer's instructions, for each sample for each time point based on the *one-color* method. Hybridized microarray slides were scanned by an Agilent scanner. Relative hybridization intensities and background hybridization values were calculated using Agilent Feature Extraction Software (9.5.1.1). Among two probes designed for each transcript sequences, we selected the probe with larger median. We also removed probes with low signal and low correlation between individuals using following three criteria–(1) no signal over all time points, (2) mean signal value over all time points is lower than 0.05, (3) mean of correlation between each pair of individuals is smaller than 0.2. Finally, we obtained time-series data of 15,451 and 15,182 independent probes for *Q. glauca* and *L. edulis*, respectively.

**Prediction of orthologous genes**

To identify orthologous genes across *Q. glauca* and *L. edulis*, we first used TransDecoder (http://transdecoder.sourceforge.net/) for detecting coding regions from the assembled

contigs. In order to maximize sensitivity for capturing coding regions with functional significance, we scanned all coding regions detected by TransDecoder for the blastp or pfam searches. We used protein sequence database of green plants (Viridiplantae) for the homology searches with E-value < 1E-5. Among the assembled contigs of *Q. glauca* and *L. edulis*, TransDecorder identified 101,371 and 86,128 contigs containing candidate coding regions with homology to known proteins. The longest predicted protein sequences of candidate coding regions were used for subsequent analysis. The construction of groups of orthologous genes (orthogroups, referred to here as gene families including ortholog pair) was performed for 5 plant species: *Q. glauca*, *L. edulis*, two other oak species, *Fagus crenata* (75,926 sequences) and *Quercus robur* (25808 sequences from OAK GENOME SEQUENCING http://www.oakgenome.fr), and *Arabidopsis thaliana* (48,359 sequences from TAIR https://www.arabidopsis.org). The prediction of orthogroups was based on a blastp all-against-all comparison of the protein sequences (E-value < 10−5) of these species, followed by clustering with Ortholog-Finder (Horiike et al., 2016) using default parameters. We obtained 32,149 orthogroups in total. Next, we picked up pairs of orthologous microarray probe for *Q. glauca* and *L. edulis* based on the predicted orthogroups. We considered a pair of the probes of which sequences belongs to an identical orthogroup to be ortholog gene. Some probes could not make orthologous pair because those belong to an orthogroup which lacks either of two species (*Q. glauca* and *L. edulis*). The probes which have multiple partners were excluded from the following analyses, because we could not conclusively identify the best orthologous pair among them. Sequences of such probes generally belong to a large orthogroup. We also excluded orthologous pairs of probes of which sequences belong to an orthogroup lacking *A. thaliana*, because we could not reliably assign their function.

Finally, we could obtain 9,258 pairs of the probes which are predicted to be ortholog genes. GOterms of predicted proteins (orthogroups) were retrieved from annotation data of *A. thaliana*.

**Selection of genes associated with DNA repair and epigenetic regulation for analyses**

Among 9,258 pair of probes, we picked up a total of 264 pairs of probes of ortholog gene for *Q. glauca* and *L. edulis*, which were associated with DNA repair and epigenetic regulation in *A. thaliana*, based on the literature (Singh et al. 2010; Pikaard & Scheid, 2014; Kim, 2019). There were 146 probes associated with DNA repair genes and 118 probes associated with epigenetic regulatory genes. We categorized each probe into one of 16 functional groups (11 groups in DNA repair: base excision repair, nucleotide excision repair, homologous recombination repair, mismatch repair, non-homologous end-joinning repair, editing and processing nuclease, modulation of nucleotide pool, DNA polymerase, Rad6 pathway, direct reversal of damages, DNA damage response; five groups in epigenetic regulations: DNA modification, histone modification, chromatin formation or chromatin remodeling, Polycomb-group proteins and interacting components, RNA silencing) based on the litelature (Singh et al. 2010; Pikaard & Scheid, 2014; Kim, 2019). The selected ortholog genes are shown in Appendix Table S2.

**Statistical analysis**

For statistical analyses described below, we used the data from samples collected from March 2017 to February 2019 (Table 2), and time series data of 264 probes were normalized so that a mean was zero and a standard deviation was one for each probe.

**Principal component analysis and enrichment analysis**

To assess the seasonal expression dynamics of genes, we performed principal component analysis (PCA) for gene expression data from all samples. We performed PCA using the function *prcomp* of the package *stats* in R. To investigate genes and functions that most contribute to each principal component, we picked up the top 13 genes (the top 5% of 264 genes) with the highest absolute values of eigenvectors. Then, to test the enrichment of each functional group in each principal component, we performed Fisher exact tests (two-sided). After the Fisher exact test, we controlled for the false discovery rate using the method of Storey's Q-value (Storey, 2002) and estimated the Q-value of each test using the *qvalue* package (ver. 2.16.0; Storey et al., 2015) in R.

To perform all statistical analyses, we used R ver. 3.4.1 (the R project, http://www.r-project.org/).

## RESULTS

### Principal component analysis reveals the similar and different seasonal gene expression dynamics among *Quercus glauca* and *Lithocarpus edulis*

The standard deviation of the first three principal components (PCs) were 8.22, 6.16 and 4.70, respectively. The first three PCs explained 26.5, 14.9 and 8.65% of the variation, respectively (Table 3). In PC1, PC scores were high around winter but low around spring and summer in *Quercus glauca* and *Lithocarpus edulis* (Fig. 1A). In PC2, PC scores were high around summer but low around spring and fall in both species (Fig. 1B). Although the pattens of PC score was almost similar throughout the sampling period for both species, the time when the PC score was lowest differed between species. PC score was lowest on March 9, 2018 in *Q. glauca* and May 1, 2018 in *L. edulis*, respectively (Fig.

1B). In PC3, there were a contrast pattern of PC scores between species (Fig. 1C). For *Q. glauca*, PC scores were high around spring and summer but low around winter. In contrast, for *L. edulis*, PC scores were high around winter but low around summer. PC4 and PC5 explained minor parts of the total variance in the data (Table 3), and seasonal patterns of PC scores were similar among species in each PC4 and PC5. Although PC6 explained a minor part of the total variance in the data (Table 3), it showed the different patterns of PC scores between species (Fig. 1D). Both species showed periodic-like patterns of PC scores, but peaks and nadirs differed between species. For *Q. glauca*, PC scores were high around summer and winter but low around apring and fall. In contrast, for *L. edulis*, PC scores were high around spring and fall but low around summer and winter. Based on this result, we focused on PC1 and PC2 with similar patterns among species, and PC3 and PC6 with different patterns between species.

**Genes that most contribute to a principal component and seasonal expression dynamics**

To investigate genes and functions that most contribute to each principal component, we picked up the top 5% genes of 264 genes associated with DNA repair and epigenetic regulation with the highest absolute values of eigenvectors. The top 5% genes of PC1 included one gene associated with nucleotide excision repair, one associated with mismatch repair, four associated with DNA damage response, five associated with chromatin formation and remodeling and two associated with RNA silencing (Table 4). The top 5% genes of PC2 included one gene associated with nucleotide excision repair, two associated with homologous recombination repair, two associated with modulation of nucleotide pool, one associated with DNA polymerase, one associated with DNA

damage response, two associated with chromatin formation and remodeling, three associated with DNA modification and one associated with RNA silencing (Table 5). The top 5% genes of PC3 included three genes associated with base excision repair, three associated with nucleotide excision repair, one associated with non-homologous end-joining repair, two associated with DNA damage response, one associated with chromatin formation and remodeling, two associated with histone modification and one associated with RNA silencing (Table 6). The top 5% genes of PC6 included one gene associated with base excision repair, one associated with nucleotide excision repair, two associated with homologous recombination repair, one associated with non-homologous end-joining repair, one associated with modulation of nucleotide pool, one associated with direct reversal of damage, four associated with polycomb-group proteins and interacting components, and one associated with RNA silencing (Table 7).

As the results of test of the enrichment of each gene functional group in each principal component, a significant large number of genes associated with polycomb-group proteins and interacting components was included in the top 5% genes of PC6 (Fisher exact test; P-value was 0.0010 and Q-value was 0.016) (Appendix Table S3). In PC1, among the top 5% genes, the number of genes associated with DNA damage response and those associated with chromatin formation and remodeling were slightly larger than that of genes in other functional groups, but there were not significant differences (Fisher exact test; P-value was 0.029 and Q-value was 0.23 for genes associated with DNA damage response, and P-value was 0.020 and Q-value was 0.23 for genes associated with chromatin formation and remodeling, respectively) (Appendix Table S3). In PC2, among the top 5% genes, the number of genes associated with modulation of nucleotide pool and those associated with DNA modification were slightly

large, but there were not significant differences (Fisher exact test; P-value was 0.013 and Q-value was 0.10 for genes associated with modulation of nucleotide pool, and P-value was 0.0066 and Q-value was 0.10 for genes associated with DNA modification, respectively) (Appendix Table S3). In PC3, there was no significant enrichment of a certain gene functional group (Appendix Table S3).

The top 5% genes of PC1 showed high expression levels around winter but low expression levels around summer in both species. For example, in *CHROMATIN REMODELING 12* (*CHR12*) gene, which encoded SNF2/Brahma-type chromatin-remodeling protein, and *DE-ETIOLATED 1* (*DET1*), which involved in DNA damage response, expression levels were high around winter but low around summer in *Q. glauca* and *L. edulis* (Fig. 2). The top 5% genes of PC2 showed high expression levels around spring and fall but low around summer in both species. For example, *DECREASED DNA METHYLATION 2*/*METHYLTRANSFERASE 1* (*DDM2/MET1*), which encoded a cytosine methyltransferase, and *RIBONUCLEOTIDE REDUCTASE LARGE SUBUNIT 1* (*RNR1*), which was involved in the production of deoxyribonucleoside triphosphates (dNTPs) for DNA replication and repair, showed high expression levels around spring and fall but low expression levels around summer in *Q. glauca* and *L. edulis* (Fig. 3). Seasonal expression dynamics of the top 5% genes of PC3 differed from species. In *POLY(ADP-RIBOSE) POLYMERASE 2* (*PARP2*) gene, which was involved in catalyzation of poly(ADP-ribosyl)ation and DNA repair including base excision repair, expression levels were high from fall to spring but low around summer in *Q. glauca*, whereas expression levels were high around summer but low around winter in *L. edulis* (Fig. 4A). In contrast, in *RNA-DEPENDENT RNA POLYMERASE 6/SILENCING DEFECTIVE 1/SUPPRESSOR OF GENE SILENCING 2* (*RDR6/SDE1/SGS2*), which

was involved in RNA silencing, expression levels were high around spring but low around winter in *Q. glauca*, whereas expression levels were high around winter but low around fall in *L. edulis* (Fig. 4B). Seasonal expression dynamics of the top 5% genes also differed from species in PC6. Two genes associated with polycomb-group proteins, *VERNALIZATION 5* (*VRN5*) and *MULTICOPY SUPPRESSOR OF IRA1 4* (*MSI4/FVE*), showed different seasonal expression dynamics among genes as well as species. In *VRN5*, expression levels were high around summer but low around spring and fall in *Q. glauca*, whereas and expression levels were high around spring and fall but low around summer and winter in *L. edulis* (Fig. 5A). In *MSI4/FVE*, expression levels were high around spring and fall but low around summer and winter in *Q. glauca*, whereas and expression levels were high around winter but low from spring to summer in *L. edulis* (Fig. 5B). In addition, *BRUSHY1/TONSOKU/MGOUN3* (*BRU1/TSK/MGO3*) gene, which was involved in chromatin formation and remodeling, expression levels were relatively high around summer and winter but relatively low around spring and fall in *Q. glauca*, whereas and expression levels were high around spring and fall but low around summer and winter in *L. edulis* (Fig. 5C).

## DISCUSSION

In the present study, we analyzed time-series transcriptome data collected throughout about two years from individuals of different tree species, *Quercus glauca* and *Lithocarpus edulis*, growing in natural environments, and demonstrated the seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation. Results of the present study suggest that a large number of genes associated with DNA repair and epigenetic regulation exhibit similar seasonal expression patterns among

species. In addition, genes with different seasonal expression dynamics are associated with multiple functions and involved in plant development, growth, and reproduction, which is likely to reflect the difference in vegetative and reproductive schedules among species.

**Genes with similar expression dynamics among species**

PC1 and PC2, which explained major parts of the total variance in the data, showed the similar seasonal patterns of PC scores among species (Fig. 1). This suggests that a large number of genes associated with DNA repair and epigenetic regulation exhibit similar seasonal expression patterns among species. Genes that most contribute to PC1, with high expression levels around winter, included genes associated with chromatin remodeling (e.g., *CHR12*) and histone chaperone (e.g., *HISTONE REGULATOR A* [*HIRA*], *NUCLEOSOME ASSEMBLY PROTEIN 1* [NAP1;2], and *SSRP1* and *SPT16*, subunits of FAcilitates Chromatin Transcription (FACT) complex) (Table 4). CHR12 is ATP-dependent chromatin remodeling factor and involved in growth and stress resistance. Over-expression of *AtCHR12* in *A. thaliana* displays temporary growth arrest of primary buds in response to drought and heat stress (Mlynárová et al., 2007). Histone chaperons, HIRA, NAP1 and FACT complex, are required for gene regulation, DNA replication and DNA repair (Belotserkovskaya et al., 2003; Adam, Polo & Almouzni, 2013; Zhou et al., 2015) and are involved in the control of development, growth and abiotic stress response (Nie et al., 2014; Zhou et al., 2015; Grasser, 2020). In addition, Genes that most contribute to PC1 included genes associated with DNA damage response, such as *DET1*, *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*) and *HYDROXY UREA SENSITIVE 1* (*HUS1*). These genes involved not only in DNA repair but also plant

development. DET1 and COP1 plays a role in response to DNA damage (Dornan et al., 2006; Castells et al., 2011), and photomorphogenic development (Osterlund et al., 2000; Schroeder et al., 2002; Kim et al., 2012). These genes may be involved in DNA damage repair and regulation of genes associated with development and stress response during winter in both species.

Seasonal expression dynamics of genes that most contribute to PC2 were also similar among species but different seasonal patterns of PC scores from PC1 (Fig. 1). The top genes in PC2 exhibited that expression levels were high around spring and slightly high around fall in *Q. glauca* and *L. edulis* (Fig. 3). Among the top 5% genes of PC2, the number of genes associated with modulation of nucleotide pool, such as *RNR1*, and those associated with DNA modification, such as *DDM2/MET1*, tended to be high although there were not significant differences. The Modulation of nucleotide pool and DNA modification process are important during DNA replication and cell division. *RNR1* encodes large subunit of ribonucleotide reductase involved in the production of deoxyribonucleoside triphosphates (dNTPs) for DNA replication and repair (Elledge, Zhou & Allen, 1992). *DDM2/MET1* encodes a cytosine methyltransferase and is involved in maintaining DNA methylation after DNA replication and during cell division (Kankel et al., 2003). Cell divisions actively occur in spring in preparation for defoliation in bud and leaf tissues. It is also possible that a large amount of cell division occurs in the fall because plants sometimes unfold their leaves in the fall. These suggest that genes associated with modulation of nucleotide pool and DNA modification with high expression levels in spring and fall are likely to act during DNA replication and cell division and play a role in control of DNA replication and inheritance of epigenetic states. In addition, the timing of the peak expression in spring was different among the species.

This may be because the timings of defoliation and flowering differ among the species, e.g., the defoliation and flowering occur from April to May in *Q. glauca*, and from May to June in *L. edulis*.

**Genes with different expression dynamics among species**

PC3 and PC6 showed the different seasonal patterns of PC scores among species (Fig. 1). *PARP2* gene, which the copy number was significantly increased in trees than in annual and perennial herbs (see chapter 1), *RDR6/SDE1/SGS2* gene are included in genes that most contribute to PC3 (Table 7). PARP2, a member of poly(ADP-ribose) polymerase, catalyzes the poly(ADP-ribosyl)ation, and is involved in multiple biological pathways, such as DNA damage response and repair including pathogen-induce DNA damage (Song et al., 2015), DNA replication (Messner & Hottiger, 2011), transcription (Messner & Hottiger, 2011), accumulation of anthocyanin (Schulz et al., 2012), and abiotic stress response (De Block et al., 2005; Vanderauwera et al., 2007). RDR6/SDE1/SGS2, RNA-dependent RNA polymerase, is involved in generation of small interfering RNAs (siRNAs) and is required for gene regulation by posttranscriptional gene silencing and inhibition of exogenous genes such as virus gene and transgene (Al-Kaff et al., 1998; Garcia-Ruiz et al., 2010). In addition, RDR is implicated in leaf development (Peragine et al., 2004) and self-incompatibility (Tantikanjana et al., 2009). *Q. glauca* flowers from April to May and fertiles after flowering. *L. edulis* flowers in June and receive pollen by insect pollination and displays delayed fertilization. The Difference in seasonal expression dynamics of RDR6 among species might be related a difference in timing of discrimination of compatible and incompatible pollens for self-incompatibility.

In addition to PC3, PC6 showed the different seasonal patterns of PC scores among species. In genes that most contribute to PC6, there were a significant large number of genes encoding components of POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) (e.g., *VRN5*, *FIE/FIS3*, *MSI4/FVE*) (Table 7). PRC2 repressed gene expression and is involved in the control of development, growth and reproduction (Derkacheva & Henning, 2014). In Arabidopsis thaliana, VRN5 is required for vernalization-mediated repression of *FLOWERING LOCUS C* (*FLC*) gene (Greb et al., 2007). *FIE* is universally expressed in wild-type *A. thaliana* during vegetative and reproductive phases (Köhler & Grossniklaus, 2002) and is involved in seedling development and flowering in *A. thaliana* (Yadegari et al., 2000; Kinoshita et al., 2001). MSI4/FVE is also involved in controlling the transition from vegetative to reproductive phase in *A. thaliana* (Ausin et al., 2004). These suggest that the genes encoding polycomb-group proteins have important roles in control of development and transition from vegetative to reproductive phase in plants, and the difference in seasonal gene expression dynamics is likely to affect schedules of growth and reproductive among species. In addition, *BRU1/TSK/MGO3* gene, which the copy number was significantly increased in trees than in annual and perennial herbs (see chapter 2), was included in the genes that most contribute to PC6 (Table 7) and showed the different seasonal expression dynamics among species (Fig. 5C). *BRU1/TSK/MGO3* is highly expressed in S-phase of the cell cycle (Suzuki et al., 2005), and is involved in DNA damage repair, maintenance of meristems and inheritance of chromatin states through chromatin formation and remodeling in *A. thaliana* (Guyomarc'h et al., 2004; Suzuki et al., 2004; Takeda et al., 2004). BRU1/TSK/MGO3 is also involved in regulation of genes associated with flowering and stress response, such as FLC and heat shock memory genes in *A. thaliana* (Guyomarc'h 2006; Brzezinka et al., 2018). Results of the

present study suggest that genes with different seasonal expression dynamics are associated with multiple functions and involved in plant development, growth and reproduction, which is likely to affect the difference in vegetative and reproductive schedules among species.

**Limitations of the study and future directions**

In the present study, we analyzed seasonal expression dynamics of genes associated with DNA repair and epigenetic regulation among two different species. Genes with different seasonal expression dynamics among species are likely to be associated with development and vegetative and reproductive programs, rather than longevity. This is because the lifespans of *Q. glauca* and *L. edulis* are not sufficiently different. To elucidate the relationship between seasonal expression dynamics of DNA repair and epigenetic regulatory genes and plant longevity, it is necessary to compare species with different lifespans. In addition, genes with increased copy number may have variations in expression levels and functions among copies. Improvements in sequencing and annotation can reveal differences in expression levels and patterns among copies.

**ACKNOWLEDGMENTS**

I would like to thank Kayoko Ohta and Yuta Sawasaki for assistance with collecting sampling and gene expression data. This study was funded by JSPS KAKENHI (JP17H01449) to A.S.

## REFERENCES

Adam, S., Polo, S. E., & Almouzni, G. (2013). XTranscription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. *Cell*, *155*(1), 94. https://doi.org/10.1016/j.cell.2013.08.029

Al-Kaff, N. S., Covey, S. N., Kreike, M. M., Page, A. M., Pinder, R., & Dale, P. J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science*, *279*(5359), 2113–2115. https://doi.org/10.1126/science.279.5359.2113

Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM. 2004. Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nat. Genet. 36:162–66

Beckmann, M., Václavík, T., Manceur, A. M., Šprtová, L., von Wehrden, H., Welk, E., & Cord, A. F. (2014). glUV: A global UV-B radiation data set for macroecological studies. *Methods in Ecology and Evolution*, *5*(4), 372–383. https://doi.org/10.1111/2041-210X.12168

Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science*, *301*(5636), 1090–1093. https://doi.org/10.1126/science.1085703

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borgardt, S. J., & Nixon, K. C. (2003). A comparative flower and fruit anatomical study of Quercus acutissima, a biennial-fruiting oak from the Cerris group (Fagaceae). *American Journal of Botany*, *90*(11), 1567–1584. https://doi.org/10.3732/ajb.90.11.1567

Brzezinka, K., Altmann, S., & Bäurle, I. (2018). BRUSHY1/TONSOKU/MGOUN3 is required for heat stress memory. Plant Cell and Environment, 42(3), 771–781. https://doi.org/10.1111/pce.13365

Castells, E., Molinier, J., Benvenuto, G., Bourbousse, C., Zabulon, G., Zalc, A., Cazzaniga, S., Genschik, P., Barneche, F., & Bowler, C. (2011). The conserved factor DE-ETIOLATED 1 cooperates with CUL4-DDB1 DDB2 to maintain genome integrity upon UV stress. *EMBO Journal*, *30*(6), 1162–1172. https://doi.org/10.1038/emboj.2011.20

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol 12: 133–139

De Block, M., Verduyn, C., De Brouwer, D., & Cornelissen, M. (2005). Poly(ADP-ribose) polymerase in plants affects energy homeotasis, cell death and stress tolerance. Plant Journal, 41(1), 95–106. https://doi.org/10.1111/j.1365-313X.2004.02277.x

Derkacheva, M., & Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *Journal of Experimental Botany*, *65*(10), 2769–2784. https://doi.org/10.1093/jxb/ert410

Dornan, D., Shimizu, H., Mah, A., Dudhela, T., Eby, M., O'rourke, K., Seshagiri, S., & Dixit, V. M. (2006). ATM engages autodegradation of the E3 ubiquitin ligase COP1 after DNA damage. *Science (New York, N.Y.)*, *313*(5790), 1122–1126. https://doi.org/10.1126/science.1127335

Elledge, S. J., Zhou, Z., & Allen, J. B. (1992). Ribonucleotide reductase: regulation, regulation, regulation. *Trends in Biochemical Sciences*, *17*(3), 119–123. https://doi.org/10.1016/0968-0004(92)90249-9

Foyer, C. H., & Noctor, G. (2003). Redox sensing and signalling associated with reactive oxygen in chloroplasts, peroxisomes and mitochondria. *Physiologia Plantarum*, *119*(3), 355–364. https://doi.org/10.1034/j.1399-3054.2003.00223.x

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., & Carrington, J. C. (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip mosaic virus infection. *Plant Cell*, *22*(2), 481–496. https://doi.org/10.1105/tpc.109.073056

Grasser, K. D. (2020). The FACT Histone Chaperone: Tuning Gene Transcription in the Chromatin Context to Modulate Plant Growth and Development. *Frontiers in Plant Science*, *11*(February), 1–8. https://doi.org/10.3389/fpls.2020.00085

Greb, T., Mylne, J. S., Crevillen, P., Geraldo, N., An, H., Gendall, A. R., & Dean, C. (2007). The PHD Finger Protein VRN5 Functions in the Epigenetic Silencing of Arabidopsis FLC. *Current Biology*, *17*(1), 73–78. https://doi.org/10.1016/j.cub.2006.11.052

Guyomarc'h, S., Vernoux, T., Traas, J., Zhou, D. X., & Delarue, M. (2004). MGOUN3, an Arabidopsis gene with TetratricoPeptide-Repeat-related motifs, regulates meristem

cellular organization. Journal of Experimental Botany, 55(397), 673–684. https://doi.org/10.1093/jxb/erh069

Guyomarc'h, S., Benhamed, M., Lemonnier, G., Renou, J. P., Zhou, D. X., & Delarue, M. (2006). MGOUN3: Evidence for chromatin-mediated regulation of FLC expression. Journal of Experimental Botany, 57(9), 2111–2119. https://doi.org/10.1093/jxb/erj169

Horiike, T., Minai, R., Miyata, D., Nakamura, Y., & Tateno, Y. (2016). Ortholog-finder: A tool for constructing an ortholog data set. *Genome Biology and Evolution*, *8*(2), 446–457. https://doi.org/10.1093/gbe/evw005

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L., & Richards, E. J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics*, *163*(3), 1109–1122. https://doi.org/10.1093/genetics/163.3.1109

Kim, J. M., To, T. K., Nishioka, T., & Seki, M. (2010). Chromatin regulation functions in plant abiotic stress responses. *Plant, Cell and Environment*, *33*(4), 604–611. https://doi.org/10.1111/j.1365-3040.2009.02076.x

Kim, E., Ly, V., Hatherell, A., & Schroeder, D. F. (2012). Genetic interactions between arabidopsis DET1 and UVH6 during development and abiotic stress response. *G3: Genes, Genomes, Genetics*, *2*(8), 913–920. https://doi.org/10.1534/g3.112.003368

Kim, J. H. (2019) Chromatin remodeling and epigenetic regulation in plant DNA damage repair, *International Journal of Molecular Sciences*, 20(17). doi: 10.3390/ijms20174093.

Kinoshita, T., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (2001). Polycomb repression of flowering during early plant development. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 14156–14161. https://doi.org/10.1073/pnas.241507798

Köhler, C., & Grossniklaus, U. (2002). Epigenetic inheritance of expression states in plant development: The role of Polycomb group proteins. *Current Opinion in Cell Biology*, *14*(6), 773–779. https://doi.org/10.1016/S0955-0674(02)00394-0

Lees-Miller, S. P., & Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie*, *85*(11), 1161–1173. https://doi.org/10.1016/j.biochi.2003.10.011

Messner, S.; Hottiger, M.O. Histone ADP-ribosylation in DNA repair, replication and transcription. Trends Cell. Biol., 2011, 21, 534-542.

Mlynárová, L., Nap, J. P., & Bisseling, T. (2007). The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in Arabidopsis thaliana upon perceiving environmental stress. *Plant Journal*, *51*(5), 874–885. https://doi.org/10.1111/j.1365-313X.2007.03185.x

Nie, X., Wang, H., Li, J., Holec, S., & Berger, F. (2014). The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. *Biology Open*, *3*(9), 794–802. https://doi.org/10.1242/bio.20148680

Osterlund, M., Hardtke, C., Wei, N. *et al.* Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* **405,** 462–466 (2000). https://doi.org/10.1038/35013076

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., & Poethig, R. S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes and Development*, *18*(19), 2368–2379. https://doi.org/10.1101/gad.1231804

Pikaard, C. S. & Scheid, O. M. (2014) Epigenetic regulation in plants, *Cold Spring Harbor Perspectives in Biology*, 6(12), pp. 1–32. doi: 10.1101/cshperspect.a019315.

Puchta, H. (2005). The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, *56*(409), 1–14. https://doi.org/10.1093/jxb/eri025

Roldán-Arjona, T., & Ariza, R. R. (2009). Repair and tolerance of oxidative DNA damage in plants. *Mutation Research - Reviews in Mutation Research*, *681*(2–3), 169–179. https://doi.org/10.1016/j.mrrev.2008.07.003

Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology*, *60*, 485–510. https://doi.org/10.1146/annurev.arplant.043008.092111

Satake, A., Kawatsu, K., Teshima, K., Kabeya, D., & Han, Q. (2019). Field transcriptome revealed a novel relationship between nitrate transport and flowering in Japanese beech. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-39608-1

Schroeder, D. F., Gahrtz, M., Maxwell, B. B., Cook, R. K. K., Kan, J. M., Alonso, J. M., Ecker, J. R., & Chory, J. (2002). De-Etiolated 1 and Damaged DNA Binding

Protein 1 Interact to Regulate. *Current, 12*(02), 1462–1472. http://www.sciencedirect.com/science/article/pii/S0960982202011065

Schulz, P., Neukermans, J., van der Kelen, K., Mühlenbock, P., van Breusegem, F., Noctor, G., Teige, M., Metzlaff, M., & Hannah, M. A. (2012). Chemical PARP inhibition enhances growth of arabidopsis and reduces anthocyanin accumulation and the activation of stress protective mechanisms. PLoS ONE, 7(5). https://doi.org/10.1371/journal.pone.0037287

Singh, S. K., Roy, S., Choudhury, S. R., & Sengupta, D. N. (2010) DNA repair and recombination in higher plants: Insights from comparative genomics of arabidopsis and rice, *BMC Genomics*. doi: 10.1186/1471-2164-11-443.

Sinha, R. P., & Häder, D. P. (2002). UV-induced DNA damage and repair: A review. *Photochemical and Photobiological Sciences*, *1*(4), 225–236. https://doi.org/10.1039/b201230h

Song, J., Keppler, B. D., Wise, R. R. and Bent, A. F. (2015) PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in Arabidopsis DNA Damage and Immune Responses, *PLoS Genetics*, 11(5), pp. 1–24. doi: 10.1371/journal.pgen.1005200.

Storey, J. D. (2002) A direct approach to false discovery rates, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(3), pp. 479–498. doi: 10.1111/1467-9868.00346.

Storey, J.D., Bass, A.J., Dabney, A., Robinson, D. (2015) qvalue: Q-value estimation for false discovery rate control. R Package. http://github.com/jdstorey/qvalue

Suzuki, T., Inagaki, S., Nakajima, S., Akashi, T., Ohto, M. A., Kobayashi, M., et al. (2004). A novel Arabidopsis gene Tonsoku is required for proper cell arrangement in root and shoot apical meristems, *Plant Journal*, 38(4), 673–684. https://doi.org/10.1111/j.1365-313X.2004.02074.x

Suzuki, T., Nakajima, S., Inagaki, S., Hirano-Nakakita, M., Matsuoka, K., Demura, T., et al. (2005). TONSOKU is expressed in S phase of the cell cycle and its defect delays cell cycle progression in Arabidopsis, *Plant and Cell Physiology*, 46(5), 736–742. https://doi.org/10.1093/pcp/pci082

Takeda, S., Tadele, Z., Hofmann, I., Probst, A. V., Angelis, K. J., Kaya, H., et al. (2004). BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis, *Genes and Development*, 18(7), 782–793. https://doi.org/10.1101/gad.295404

Tantikanjana, T., Rizvi, N., Nasrallah, M. E., & Nasrallah, J. B. (2009). A dual role for the s-locus receptor kinase in self-incompatibility and pistil development revealed by an arabidopsis rdr6 mutation. *Plant Cell*, *21*(9), 2642–2654. https://doi.org/10.1105/tpc.109.067801

Vanderauwera, S., De Block, M., Van De Steene, N., Van De Cotte, B., Metzlaff, M., & Van Breusegem, F. (2007). Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(38), 15150–15155. https://doi.org/10.1073/pnas.0706668104

Yadegari, R., Kinoshita, T., Lotan, O., Cohen, G., Katz, A., Choi, Y., Katz, A., Nakashima, K., Harada, J. J., Goldberg, R. B., Fischer, R. L., & Ohad, N. (2000). Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell*, *12*(12), 2367–2381. https://doi.org/10.1105/tpc.12.12.2367

Zhou, W., Zhu, Y., Dong, A., & Shen, W. H. (2015). Histone H2A/H2B chaperones: From molecules to chromatin-based functions in plant growth and development. *Plant Journal*, *83*(1), 78–95. https://doi.org/10.1111/tpj.12830

## TABLES

Table 1. List of surveyed beech family plants

| Name | Leaves | Seed | Pollen | Flowering |
|---|---|---|---|---|
| *Querucus glauca* | Evergreen | 1 Year | Wind | Apr–May |
| *Lithocarpus edulis* | Evergreen | 2 Years | Insects | Jun |

Table 2. List of sample date for DNA microarray analysis.

| Year | date |
|---|---|
| 2017 | May 3 |
| | June 1 |
| | June 28 |
| | July 26 |
| | August 24 |
| | September 20 |
| | October 18 |
| | November 15 |
| | December 13 |
| 2018 | January 14 |
| | February 8 |
| | March 9 |
| | April 4 |
| | May 1 |
| | May 31 |
| | June 27 |
| | July 25 |
| | August 21 |
| | September 19 |
| | October 17 |
| | November 15 |
| | December 13 |
| 2019 | January 9 |
| | February 4 |

Table 3. The result of principal component analysis (PCA). The results are showed up to PC6.

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|---|---|---|---|---|---|---|
| Standard deviation | 8.22 | 6.16 | 4.7 | 3.99 | 3.52 | 2.89 |
| Proportion of Variance | 0.265 | 0.149 | 0.0865 | 0.0626 | 0.0488 | 0.0327 |
| Cumulative Proportion | 0.265 | 0.414 | 0.5 | 0.563 | 0.612 | 0.645 |

Table 4. List of the top 13 genes (the top 5% of 264 genes) with the highest absolute values of eigenvectors in PC1.

| Function group | Gene symbol | AT code | Eigenvector |
|---|---|---|---|
| Chromatin formation or chromatin remodeling | *HIRA* | AT3G44530 | 0.115 |
| Mismatch repair | *Muts_like* | AT1G65070 | 0.114 |
| Chromatin formation or chromatin remodeling | *AtNAP1_2* | AT2G19480 | 0.113 |
| Chromatin formation or chromatin remodeling | *AtCHR12* | AT3G06010 | 0.112 |
| Chromatin formation or chromatin remodeling | *SSRP1* | AT3G28730 | 0.112 |
| DNA damage response | *HUS1* | AT1G52530 | 0.111 |
| Chromatin formation or chromatin remodeling | *SPT16* | AT4G10710 | 0.110 |
| RNA silencing | *DCL1/EMB76/SIN1/SUS1* | AT1G01040 | 0.110 |
| DNA damage response | *COP1* | AT2G32950 | 0.109 |
| Nucleotide excision repair | *GTF2H3* | AT1G18340 | 0.108 |
| RNA silencing | *ABH1/CBP80* | AT2G13540 | 0.108 |
| DNA damage response | *CHEK2* | AT4G04720 | 0.108 |
| DNA damage response | *DET1* | AT4G10180 | 0.107 |

Table 5. List of the top 13 genes (the top 5% of 264 genes) with the highest absolute values of eigenvectors in PC2.

| Function group | Gene symbol | AT code | Eigenvector |
|---|---|---|---|
| DNA modification | *DDM2/MET1* | AT5G49160 | -0.145 |
| DNA modification | *VIM1* | AT1G57820 | -0.144 |
| RNA silencing | *ESD7* | AT1G08260 | -0.144 |
| Modulation of nucleotide pool | *RNR1* | AT2G21790 | -0.142 |
| DNA damage response | *RECQL5* | AT1G27880 | -0.141 |
| Homologous recombination repair | *RAD54L* | AT3G19210 | -0.141 |
| DNA modification | *CMT3* | AT1G69770 | -0.139 |
| Modulation of nucleotide pool | *TSO2* | AT3G27060 | -0.138 |
| Chromatin formation or chromatin remodeling | *PCNA2* | AT2G29570 | -0.137 |
| Chromatin formation or chromatin remodeling | *PCNA1* | AT1G07370 | -0.137 |
| Homologous recombination repair | *BARD1* | AT1G04020 | -0.134 |
| DNA plymerase | *POLE* | AT5G22110 | -0.134 |
| Nucleotide excision repair | *RPA1* | AT5G08020 | -0.134 |

Table 6. List of the top 13 genes (the top 5% of 264 genes) with the highest absolute values of eigenvectors in PC3.

| Function group | Gene symbol | AT code | Eigenvector |
|---|---|---|---|
| Base excision repair | *Tag* | AT5G57970 | 0.184 |
| Base excision repair | *MPG/MAG* | AT3G12040 | 0.181 |
| Histone modification | *SUVH4* | AT5G13960 | -0.169 |
| Nucleotide excision repair | *RAD23D* | AT5G38470 | 0.164 |
| Non-homologous end-joinning repair | *ATRAD21.3* | AT5G16270 | 0.162 |
| Nucleotide excision repair | *GTF2H1* | AT1G55750 | 0.160 |
| Nucleotide excision repair | *MNAT1* | AT4G30820 | 0.154 |
| Chromatin formation or chromatin remodeling | *DMS11* | AT1G19100 | 0.150 |
| Base excision repair | *PARP2* | AT4G02390 | -0.148 |
| DNA damage response | *AXR1* | AT1G05180 | -0.145 |
| RNA silencing | *RDR6/SDE1/SGS2* | AT3G49500 | 0.143 |
| Histone modification | *HAG1* | AT3G54610 | -0.140 |
| DNA damage response | *CHEK1* | AT2G26980 | 0.139 |

Table 7. List of the top 13 genes (the top 5% of 264 genes) with the highest absolute values of eigenvectors in PC6.

| Function group | Gene symbol | AT code | Eigenvector |
|---|---|---|---|
| Homologous recombination repair | *RAD51A* | AT5G20850 | -0.221 |
| Polycomb-group proteins and interacting components | *VRN5* | AT3G24440 | 0.177 |
| Polycomb-group proteins and interacting components | *VEL1/VIL2* | AT4G30200 | 0.172 |
| Direct reversal of damage | *UVR3* | AT3G15620 | -0.156 |
| Nucleotide excision repair | *RFC1* | AT5G22010 | -0.141 |
| Polycomb-group proteins and interacting components | *FIE/FIS3* | AT3G20740 | 0.141 |
| Homologous recombination repair | *RAD51B* | AT2G28560 | -0.140 |
| Chromatin formation or chromatin remodeling | *BRU1/MGO3/TSK* | AT3G18730 | 0.140 |
| RNA silencing | *FPA* | AT2G43410 | -0.135 |
| Polycomb-group proteins and interacting components | *MSI4/FVE* | AT2G19520 | -0.134 |
| Base excision repair | *PARP2* | AT4G02390 | -0.134 |
| Non-homologous end-joinning repair | *LIG4* | AT5G57160 | 0.129 |
| Modulation of nucleotide pool | *NUDX1* | AT1G68760 | 0.125 |

**FIGURES**

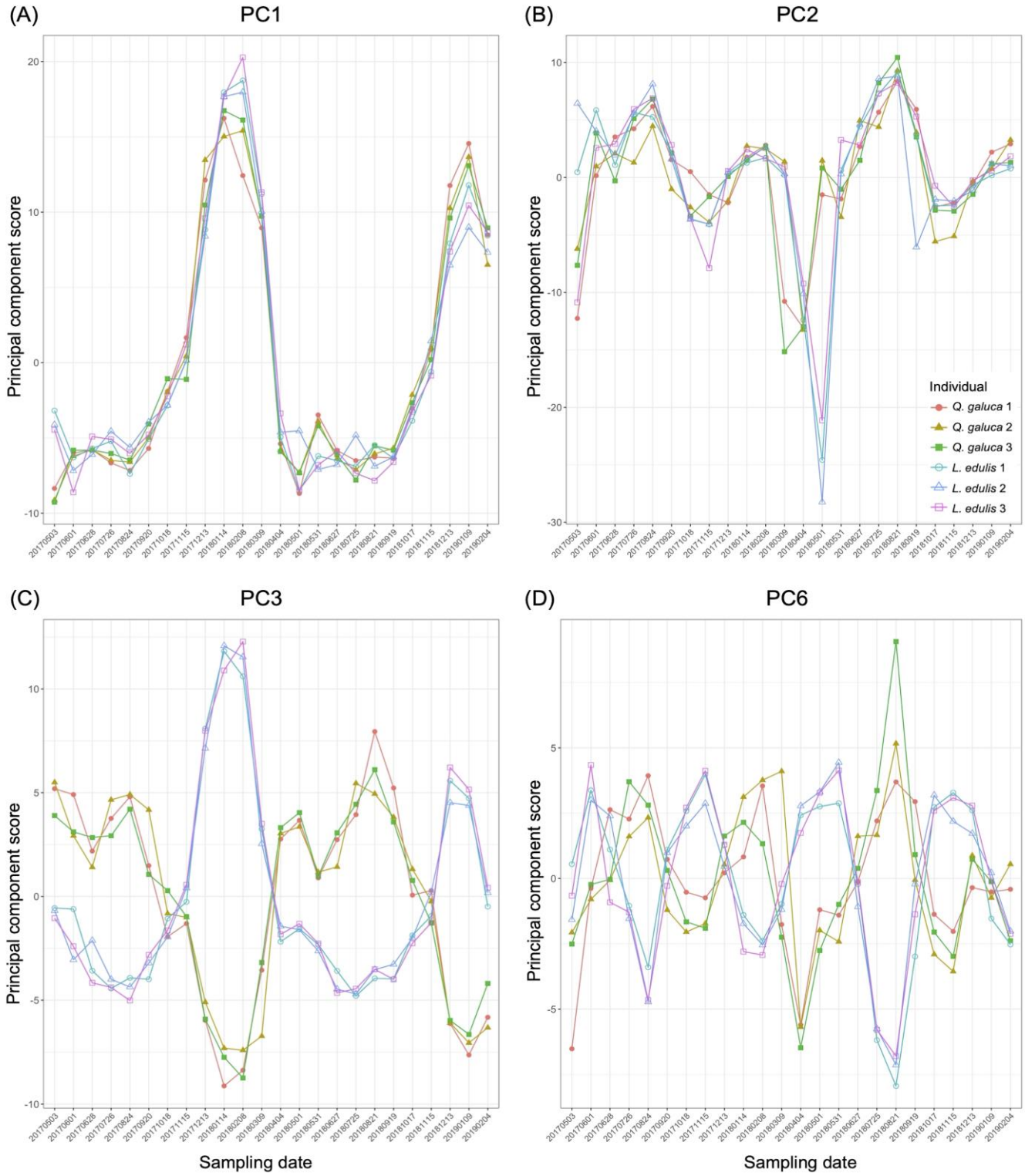

(A)
PC1
(B)
PC2
(C)
PC3
(D)
PC6
Principal component score
Sampling date
Individual
Q. galuca 1
Q. galuca 2
Q. galuca 3
L. edulis 1
L. edulis 2
L. edulis 3

Figure 1. The principal component scores of each sample in PC1 (A), PC2 (B), PC3 (C) and PC6 (D). The vertical axis represents the principal component score, and the horizontal axis indicates the sampling date. Filled marker indicates the data from *Quercus glauca*; open marker indicates the data from *Lithocarpus edulis*. Shapes of makers represent the individuals.

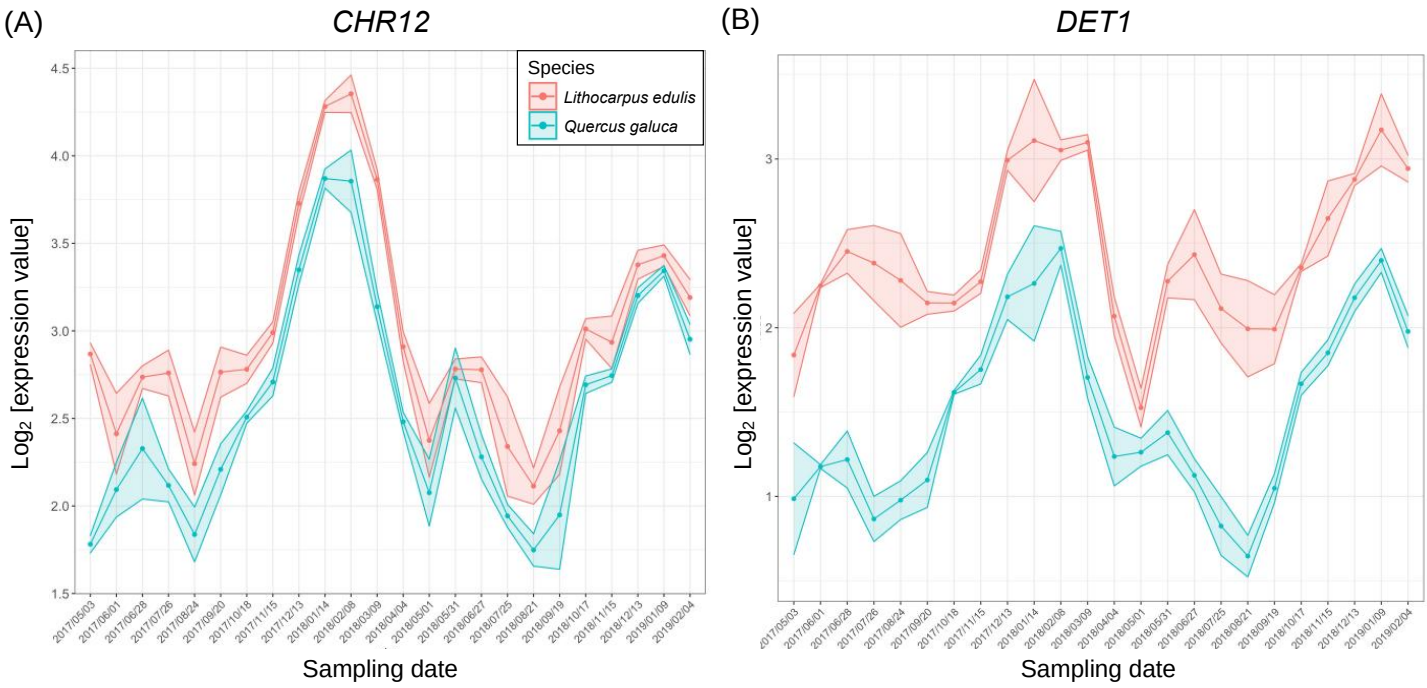



Figure 2. Seasonal gene expression dynamics of *CHR12* (A) and *DET1* (B) in *Quercus glauca* and *Lithocarpus edulis*. Values of gene expression were log2 transformed. Each point represents mean expression value of three individuals in the species, and shaded regions represent standard deviation. Red points and lines indicate *Q. glauca* and blue indicate *L. edulis*.

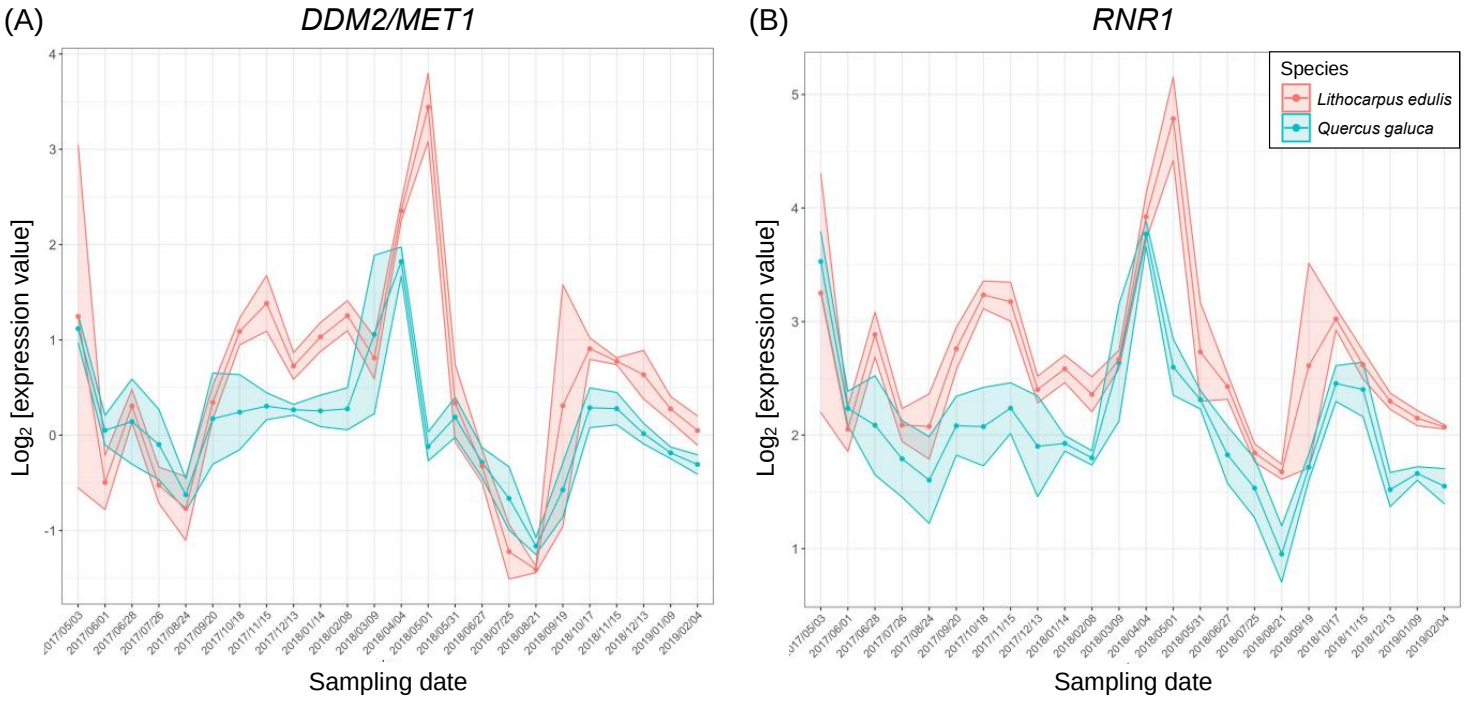



Figure 3. Seasonal gene expression dynamics of *DDM2/MET1* (A) and *RNR1* (B) in *Quercus glauca* and *Lithocarpus edulis*. Values of gene expression were log2 transformed. Each point represents mean expression value of three individuals in the species, and shaded regions represent standard deviation. Red points and lines indicate *Q. glauca* and blue indicate *L. edulis*.

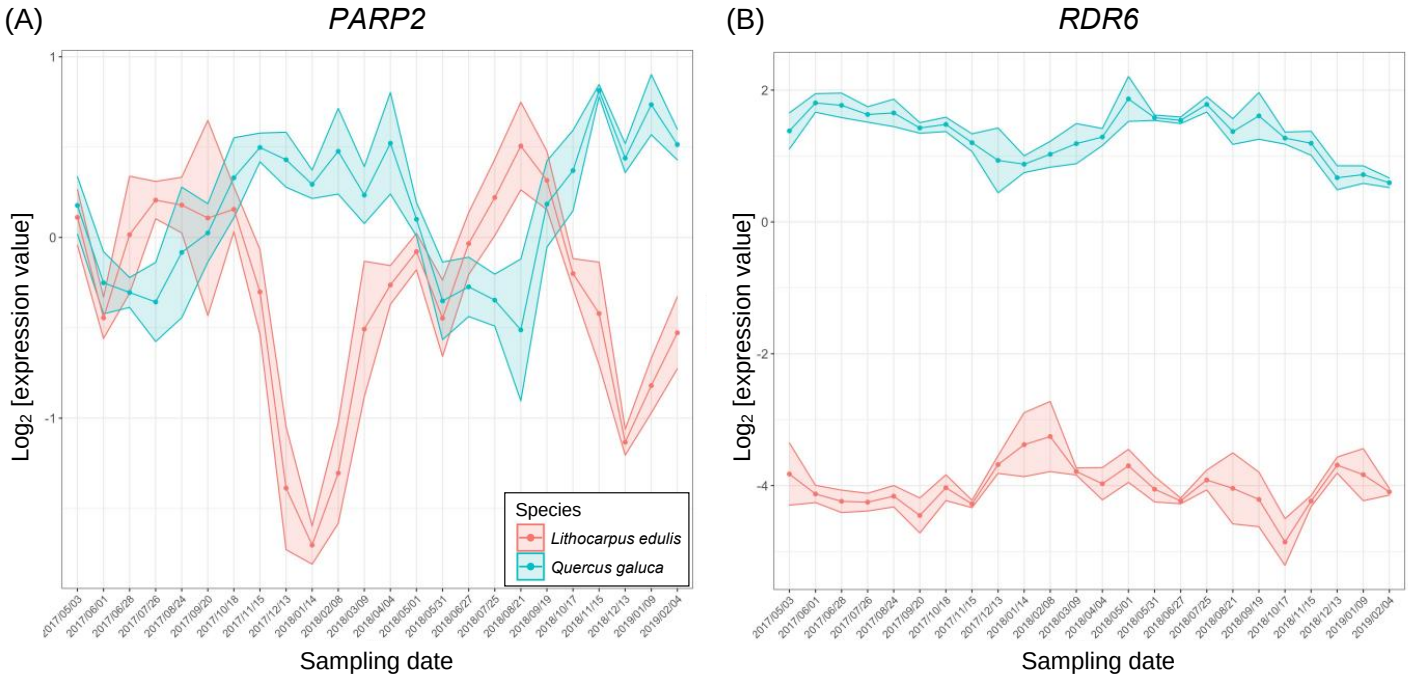



Figure 4. Seasonal gene expression dynamics of *PARP2* (A) and *RDR6* (B) in *Quercus glauca* and *Lithocarpus edulis*. Values of gene expression were log2 transformed. Each point represents mean expression value of three individuals in the species, and shaded regions represent standard deviation. Red points and lines indicate *Q. glauca* and blue indicate *L. edulis*.

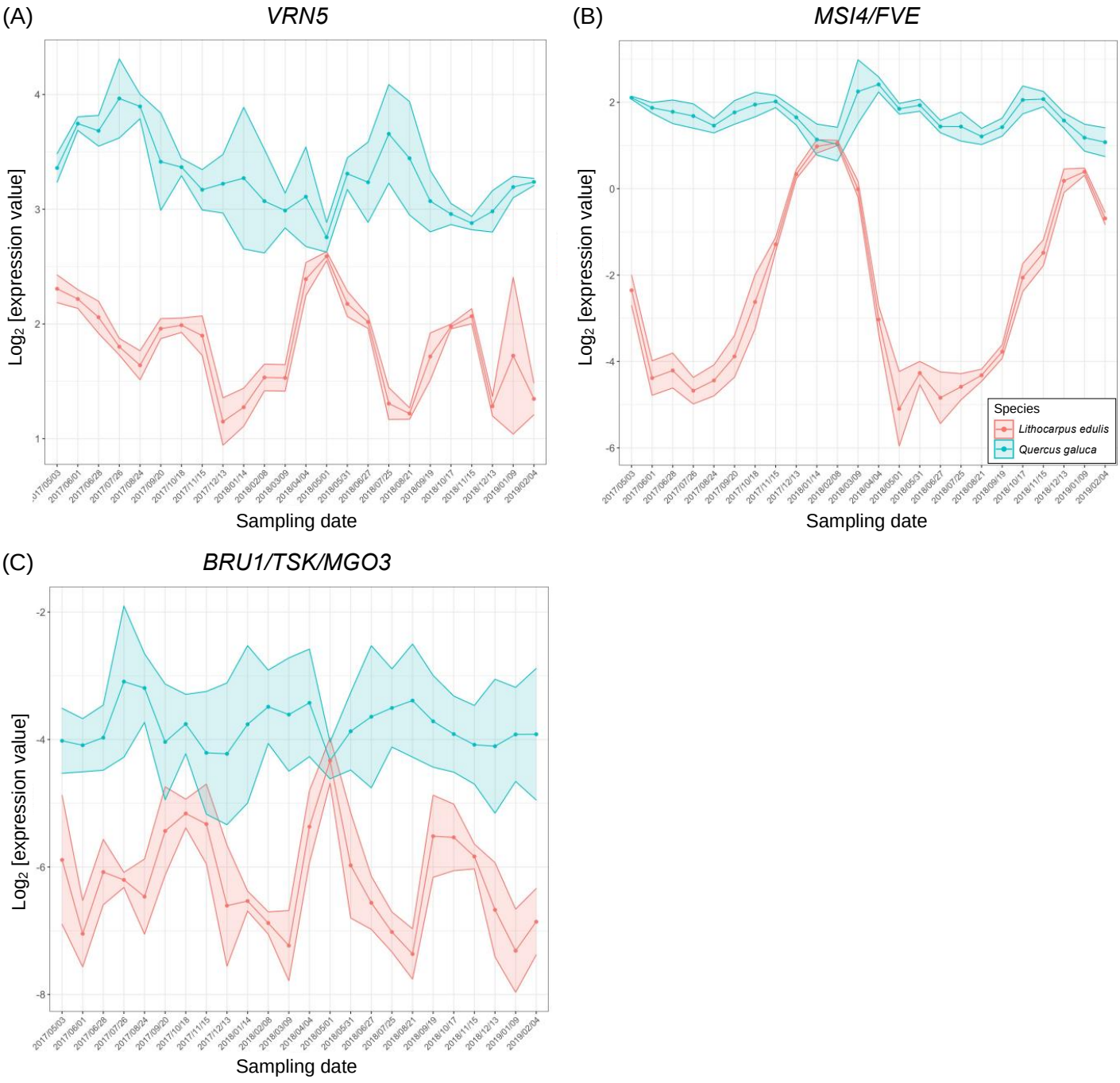



Figure 5. Seasonal gene expression dynamics of *VRN5* (A), *MSI4/FVE* (B) and *BRU1/TSK/MGO3* (C) in *Quercus glauca* and *Lithocarpus edulis*. Values of gene expression were log2 transformed. Each point represents mean expression value of three

individuals in the species, and shaded regions represent standard deviation. Red points and lines indicate *Q. glauca* and blue indicate *L. edulis*.

## APPENDIXES

Appendix Table S1. List of samples used for NGS analysis.

| *Q. glauca* sampling date | *L. edulis* sampling date |
|---|---|
| May 3, 2017 | June 1, 2017 |
| June 1, 2017 | June 28, 2017 |
| July 26, 2017 | July 26, 2017 |
| August 24, 2017 | August 24, 2017 |
| September 20, 2017 | September 20, 2017 |
| October 18, 2017 | October 18, 2017 |
| November 15, 2017 | November 15, 2017 |
| December 13, 2017 | December 13, 2017 |

Appendix Table S2. List of genes associated with DNA repair and epigenetic regulation for expression data analyses.

| Function group | Gene symbol | AT code |
|---|---|---|
| **DNA repair** | | |
| Base excision repair | *Tag* | AT1G13635 |
| | *OGG1* | AT1G21710 |
| | *FPG* | AT1G52500 |
| | *MAGLP/AlkA* | AT1G75230 |
| | *XRCC1* | AT1G80420 |
| | *PARP1* | AT2G31320 |
| | *NTH* | AT2G31450 |
| | *APE1* | AT2G41460 |
| | *DML2* | AT3G10140 |
| | *MPG/MAG* | AT3G12040 |
| | *Tag* | AT3G12710 |
| | *UNG* | AT3G18630 |
| | *APE1L* | AT3G48425 |
| | *MAGLP/AlkA* | AT3G50880 |
| | *PARP2* | AT4G02390 |
| | *MUTY* | AT4G12740 |
| | *APE2* | AT4G36050 |
| | *APTX* | AT5G01310 |
| | *TDP1* | AT5G15170 |
| | *Tag* | AT5G57970 |
| Nucleotide excision repair | *XPD/UVH6/ERCC2* | AT1G03190 |
| | *GTF2H2* | AT1G05055 |
| | *RAD16* | AT1G05120 |
| | *LIG1* | AT1G08130 |
| | *CDK7* | AT1G18040 |
| | *GTF2H3* | AT1G18340 |
| | *RFC2* | AT1G21690 |
| | *CSA* | AT1G27840 |
| | *GTF2H1* | AT1G55750 |

| | | |
|---|---|---|
| | *RFC4* | AT1G63160 |
| | *RFC3* | AT1G77470 |
| | *RAD23A* | AT1G79650 |
| | *Mfd* | AT3G02060 |
| | *CETN2* | AT3G50360 |
| | *DDB1* | AT4G05420 |
| | *GTF2H4* | AT4G17020 |
| | *RPA3* | AT4G18590 |
| | *UvrD* | AT4G25120 |
| | *MNAT1* | AT4G30820 |
| | *RPA1* | AT5G08020 |
| | *XPC* | AT5G16630 |
| | *RBX1* | AT5G20570 |
| | *RFC1* | AT5G22010 |
| | *CCNH* | AT5G27620 |
| | *RFC5* | AT5G27740 |
| | *XAB2* | AT5G28740 |
| | *RAD23D* | AT5G38470 |
| | *RAD1/UVH1/ERCC4/XPF* | AT5G41150 |
| | *XPB/ERCC3* | AT5G41370 |
| | *RPA1* | AT5G45400 |
| | *CUL4* | AT5G46210 |
| | *MMS19* | AT5G48120 |
| Homologous recombination repair | *BARD1* | AT1G04020 |
| | *RAD51D* | AT1G07745 |
| | *BLM/RecQl4* | AT1G10930 |
| | *RecA* | AT1G79050 |
| | *RecG* | AT2G01440 |
| | *RecA* | AT2G19490 |
| | *EME1* | AT2G22140 |
| | *RAD51B* | AT2G28560 |
| | *RAD50* | AT2G31970 |
| | *TOP3* | AT2G32000 |
| | *RAD51C* | AT2G45280 |
| | *NBS1* | AT3G02680 |
| | *RAD54L* | AT3G19210 |

| | | |
|---|---|---|
| | *DMC* | AT3G22880 |
| | *SSB* | AT4G11060 |
| | *BRCA1* | AT4G21070 |
| | *MND1* | AT4G29170 |
| | *BRCA2* | AT5G01630 |
| | *RAD51A* | AT5G20850 |
| | *XRCC3* | AT5G57450 |
| | *TOP3* | AT5G63920 |
| Mismatch repair | *Muts_like* | AT1G65070 |
| | *MSH2* | AT3G18524 |
| | *MSH7* | AT3G24495 |
| | *MSH6* | AT4G02070 |
| | *PMS1* | AT4G02460 |
| | *MLH1* | AT4G09140 |
| | *MSH4* | AT4G17380 |
| | *MSH3* | AT4G25540 |
| | *Muts_like* | AT5G54090 |
| Non-homologous end-joinning repair | *KU70* | AT1G16970 |
| | *KU80* | AT1G48050 |
| | *PRKDC* | AT1G50030 |
| | *XRCC4* | AT3G23100 |
| | *ATRAD21.2* | AT3G59550 |
| | *ATRAD21.3* | AT5G16270 |
| | *LIG4* | AT5G57160 |
| Editing and processing nuclease | *GEN1* | AT1G01880 |
| | *HEX1/EXO1* | AT1G18090 |
| | *HEX1/EXO1* | AT1G29630 |
| | *SPO11-1* | AT3G13170 |
| | *GEN2* | AT3G48900 |
| | *FLJ35220* | AT4G31150 |
| | *FEN1* | AT5G26680 |
| Modulation of nucleotide pool | *NUDX1* | AT1G68760 |
| | *RNR1* | AT2G21790 |
| | *RNR2a* | AT3G23580 |
| | *TSO2* | AT3G27060 |
| DNA plymerase | *POLD4* | AT1G09815 |

| | | |
|---|---|---|
| | *POLL* | AT1G10520 |
| | *REV7* | AT1G16590 |
| | *Polk* | AT1G49980 |
| | *REV3* | AT1G67500 |
| | *POLD3* | AT1G78650 |
| | *POLD2* | AT2G42120 |
| | *POLE* | AT5G22110 |
| | *POLH* | AT5G44750 |
| | *POLD1* | AT5G63960 |
| Rad6 pathway | *UBE2N* | AT1G16890 |
| | *MMS2* | AT1G70660 |
| | *MMS2* | AT3G52560 |
| Direct reversal of damage | *CRY2* | AT1G04400 |
| | *AlkB* | AT1G11780 |
| | *ABH3/AlkB* | AT2G22260 |
| | *PHR2* | AT2G47590 |
| | *UVR3* | AT3G15620 |
| | *CRY1* | AT4G08920 |
| | *CRY3* | AT5G24850 |
| DNA damage response | *PR19B/PUB60-1* | AT1G04510 |
| | *AXR1* | AT1G05180 |
| | *SOG1* | AT1G25580 |
| | *SNM1B* | AT1G27410 |
| | *RECQL5* | AT1G27880 |
| | *DRT111* | AT1G30480 |
| | *RECQl2* | AT1G31360 |
| | *HUS1* | AT1G52530 |
| | *CHEK1* | AT2G26980 |
| | *SMC3* | AT2G27170 |
| | *COP1* | AT2G32950 |
| | *DRT102* | AT3G04880 |
| | *RAD9* | AT3G05480 |
| | *SNM1* | AT3G26680 |
| | *ATM* | AT3G48190 |
| | *SMC1* | AT3G54670 |
| | *CHEK2* | AT4G04720 |

| | | |
|---|---|---|
| | *DET1* | AT4G10180 |
| | *RAD1* | AT4G17760 |
| | *REX1* | AT5G04910 |
| | *DRT101* | AT5G18070 |
| | *SM3L2/RAD5a* | AT5G22750 |
| | *RECQSIM* | AT5G27680 |
| | *SM3L/RAD5b* | AT5G43530 |
| | *SMC2* | AT5G62410 |
| | *RAD17* | AT5G66130 |
| **Epigenetic reguration** | | |
| Chromatin formation or chromatin remodeling | *SWI2* | AT1G03750 |
| | *PCNA1* | AT1G07370 |
| | *ARP4* | AT1G18450 |
| | *DMS11* | AT1G19100 |
| | *AtSWI3_C/SWI3C* | AT1G21700 |
| | *CHR18* | AT1G48310 |
| | *FRG2/SNF2-RING-HELICASE LIKE2* | AT1G50410 |
| | *FAS1* | AT1G65470 |
| | *CHR5* | AT2G13370 |
| | *DRD1* | AT2G16390 |
| | *AtNAP1_2* | AT2G19480 |
| | *RPA2* | AT2G24490 |
| | *CHD3/PKL* | AT2G25170 |
| | *PCNA2* | AT2G29570 |
| | *AtSWI3_B/SWI3B* | AT2G33610 |
| | *SWR1* | AT2G47210 |
| | *AtCHR12* | AT3G06010 |
| | *CHR11* | AT3G06400 |
| | *ARP5* | AT3G12380 |
| | *PIE* | AT3G12810 |
| | *MMS21* | AT3G15150 |
| | *BSH* | AT3G17590 |
| | *BRU1/MGO3/TSK* | AT3G18730 |
| | *SSRP1* | AT3G28730 |
| | *HIRA* | AT3G44530 |
| | *DMS3/IDN1* | AT3G49250 |

| | | |
|---|---|---|
| | *INO80* | AT3G57300 |
| | *SPT16* | AT4G10710 |
| | *AtSWP73_B/CHC1* | AT5G14170 |
| | *SMC5* | AT5G15920 |
| | *TSL* | AT5G20930 |
| | *AtASF1b* | AT5G38110 |
| | *MGO1* | AT5G55300 |
| | *ARP8* | AT5G56180 |
| | *FAS2* | AT5G64630 |
| | *SOM* | AT5G66750 |
| DNA modification | *VIM1* | AT1G57820 |
| | *CMT3* | AT1G69770 |
| | *ZDP* | AT3G14890 |
| | *CMT2* | AT4G19020 |
| | *H3.3, HTR4* | AT4G40030 |
| | *DME* | AT5G04560 |
| | *DNMT2* | AT5G25480 |
| | *DDM2/MET1* | AT5G49160 |
| | *DDB2* | AT5G58760 |
| Histone modification | *HAF1* | AT1G32750 |
| | *HUB2* | AT1G55250 |
| | *LDL1* | AT1G62830 |
| | *EFS/SDG8/ASHH2* | AT1G77300 |
| | *HAC1* | AT1G79000 |
| | *UBC2* | AT2G02760 |
| | *SUVH6* | AT2G22740 |
| | *OTLD1* | AT2G27350 |
| | *SGS1/NAC052* | AT3G10490 |
| | *LDL2* | AT3G13682 |
| | *REF6* | AT3G48430 |
| | *SUP32/UBP26* | AT3G49600 |
| | *HAG1* | AT3G54610 |
| | *ATXR3/SDG2* | AT4G15180 |
| | *ULT1* | AT4G28190 |
| | *ELF6* | AT5G04240 |
| | *ATXR5* | AT5G09790 |

| | | |
|---|---|---|
| | *SUVH4* | AT5G13960 |
| | *ATXR6* | AT5G24330 |
| | *HAG3* | AT5G50320 |
| | *SRT1* | AT5G55760 |
| | *HAG2* | AT5G56740 |
| | *AXE1/HDA6/RPD3B/RTS1/SIL1* | AT5G63110 |
| | *NAC103* | AT5G64060 |
| Polycomb-group proteins and interacting components | *MSI4/FVE* | AT2G19520 |
| | *CLF/SET1* | AT2G23380 |
| | *RBR* | AT3G12280 |
| | *FIE/FIS3* | AT3G20740 |
| | *VRN5* | AT3G24440 |
| | *AtCYP71* | AT3G44600 |
| | *LIF2* | AT4G00830 |
| | *VRN2* | AT4G16845 |
| | *VEL1/VIL2* | AT4G30200 |
| | *LHP1/TFL2* | AT5G17690 |
| | *EMF2* | AT5G51230 |
| RNA silencing | *DCL1/EMB76/SIN1/SUS1* | AT1G01040 |
| | *NRPC7* | AT1G06790 |
| | *ESD7* | AT1G08260 |
| | *FDM4* | AT1G13790 |
| | *RDR1* | AT1G14790 |
| | *SHH1/DTF1* | AT1G15215 |
| | *FDM1* | AT1G15910 |
| | *AGO2* | AT1G31280 |
| | *XRN4/EIN5* | AT1G54490 |
| | *POL IV/SMD2* | AT1G63020 |
| | *AGO7/ZIP* | AT1G69440 |
| | *XRN3* | AT1G75660 |
| | *ABH1/CBP80* | AT2G13540 |
| | *NRPB3/NRPD3/NRPE3a* | AT2G15430 |
| | *AGO4* | AT2G27040 |
| | *AGO5* | AT2G27880 |
| | *DRB2* | AT2G28380 |
| | *RDM4/DMS4* | AT2G30280 |

| | |
|---|---|
| *FPA* | AT2G43410 |
| *DCL2* | AT3G03300 |
| *HST* | AT3G05040 |
| *SDE5* | AT3G15390 |
| *NRPB9a/NRPD9a/NRPE9a* | AT3G16980 |
| *DDL* | AT3G20550 |
| *DRD2/NRPD2A/NRPE2* | AT3G23780 |
| *IDN2/RDM12* | AT3G48670 |
| *RDR6/SDE1/SGS2* | AT3G49500 |
| *NRPE5* | AT3G57080 |
| *DRB4* | AT3G62800 |
| *RDR2/SMD1* | AT4G11130 |
| *WEX* | AT4G13870 |
| *FCA* | AT4G16280 |
| *HEN1* | AT4G20910 |
| *KTF1/RDM3/SPT5-l* | AT5G04290 |
| *DCL4* | AT5G20320 |
| *AGO10/PNH/ZLL* | AT5G43810 |
| *NRPC2* | AT5G45140 |
| *FRY1/SAL1* | AT5G63980 |

Appendix Table S3. Results of Fisher exact test.

| | The function group of the gene | The number of target genes in the principal component | The number of target genes in all genes | p-values | Q-values |
|---|---|---|---|---|---|
| PC1 | Base excision repair | 0 | 20 | 0.6075 | 1.0000 |
| | Nucleotide excision repair | 1 | 32 | 1.0000 | 1.0000 |
| | Homologous recombination repair | 0 | 21 | 0.6082 | 1.0000 |
| | Mismatch repair | 1 | 9 | 0.3698 | 1.0000 |
| | Non-homologous endJoining repair | 0 | 7 | 1.0000 | 1.0000 |
| | Editing and processing nuclease | 0 | 7 | 1.0000 | 1.0000 |
| | Modulation of nucleotide pool | 0 | 4 | 1.0000 | 1.0000 |
| | DNA plymerase | 0 | 10 | 1.0000 | 1.0000 |
| | Rad6 pathway | 0 | 3 | 1.0000 | 1.0000 |
| | Direct reversal of damage | 0 | 7 | 1.0000 | 1.0000 |
| | DNA damage response | 4 | 26 | 0.0288 | 0.2301 |
| | Chromatin formation or chromatin remodeling | 5 | 36 | 0.0204 | 0.2301 |
| | DNA modification | 0 | 9 | 1.0000 | 1.0000 |
| | Histone modification | 0 | 24 | 0.6155 | 1.0000 |
| | Polycomb group proteins and interacting components | 0 | 11 | 1.0000 | 1.0000 |
| | RNA silencing | 2 | 38 | 1.0000 | 1.0000 |
| PC2 | Base excision repair | 0 | 20 | 0.6075 | 1.0000 |
| | Nucleotide excision repair | 1 | 32 | 1.0000 | 1.0000 |
| | Homologous recombination repair | 2 | 21 | 0.2766 | 1.0000 |
| | Mismatch repair | 0 | 9 | 1.0000 | 1.0000 |
| | Non-homologous endJoining repair | 0 | 7 | 1.0000 | 1.0000 |
| | Editing and processing nuclease | 0 | 7 | 1.0000 | 1.0000 |
| | Modulation of nucleotide pool | 2 | 4 | 0.0127 | 0.1019 |
| | DNA plymerase | 1 | 10 | 0.4019 | 1.0000 |
| | Rad6 pathway | 0 | 3 | 1.0000 | 1.0000 |
| | Direct reversal of damage | 0 | 7 | 1.0000 | 1.0000 |
| | DNA damage response | 1 | 26 | 1.0000 | 1.0000 |
| | Chromatin formation or chromatin remodeling | 2 | 36 | 0.6930 | 1.0000 |

| | | | | | |
|---|---|---|---|---|---|
| | DNA modification | 3 | 9 | 0.0066 | 0.1019 |
| | Histone modification | 0 | 24 | 0.6155 | 1.0000 |
| | Polycomb group proteins and interacting components | 0 | 11 | 1.0000 | 1.0000 |
| | RNA silencing | 1 | 38 | 0.6998 | 1.0000 |
| PC3 | Base excision repair | 3 | 20 | 0.0653 | 1.0000 |
| | Nucleotide excision repair | 3 | 32 | 0.1989 | 1.0000 |
| | Homologous recombination repair | 0 | 21 | 0.6082 | 1.0000 |
| | Mismatch repair | 0 | 9 | 1.0000 | 1.0000 |
| | Non-homologous endJoining repair | 1 | 7 | 0.3007 | 1.0000 |
| | Editing and processing nuclease | 0 | 7 | 1.0000 | 1.0000 |
| | Modulation of nucleotide pool | 0 | 4 | 1.0000 | 1.0000 |
| | DNA plymerase | 0 | 10 | 1.0000 | 1.0000 |
| | Rad6 pathway | 0 | 3 | 1.0000 | 1.0000 |
| | Direct reversal of damage | 0 | 7 | 1.0000 | 1.0000 |
| | DNA damage response | 2 | 26 | 0.3729 | 1.0000 |
| | Chromatin formation or chromatin remodeling | 1 | 36 | 1.0000 | 1.0000 |
| | DNA modification | 0 | 9 | 1.0000 | 1.0000 |
| | Histone modification | 2 | 24 | 0.3346 | 1.0000 |
| | Polycomb group proteins and interacting components | 0 | 11 | 1.0000 | 1.0000 |
| | RNA silencing | 1 | 38 | 0.6998 | 1.0000 |
| PC6 | Base excision repair | 1 | 20 | 1.0000 | 1.0000 |
| | Nucleotide excision repair | 1 | 32 | 1.0000 | 1.0000 |
| | Homologous recombination repair | 2 | 21 | 0.2766 | 0.9622 |
| | Mismatch repair | 0 | 9 | 1.0000 | 1.0000 |
| | Non-homologous endJoining repair | 1 | 7 | 0.3007 | 0.9622 |
| | Editing and processing nuclease | 0 | 7 | 1.0000 | 1.0000 |
| | Modulation of nucleotide pool | 1 | 4 | 0.1839 | 0.9622 |
| | DNA plymerase | 0 | 10 | 1.0000 | 1.0000 |
| | Rad6 pathway | 0 | 3 | 1.0000 | 1.0000 |
| | Direct reversal of damage | 1 | 7 | 0.3007 | 0.9622 |
| | DNA damage response | 0 | 26 | 0.6242 | 1.0000 |
| | Chromatin formation or chromatin remodeling | 1 | 36 | 1.0000 | 1.0000 |

| | | | | |
|---|---|---|---|---|
| DNA modification | 0 | 9 | 1.0000 | 1.0000 |
| Histone modification | 0 | 24 | 0.6155 | 1.0000 |
| Polycomb group proteins and interacting components | 4 | 11 | 0.0010 | 0.0157 |
| RNA silencing | 1 | 38 | 0.6998 | 1.0000 |