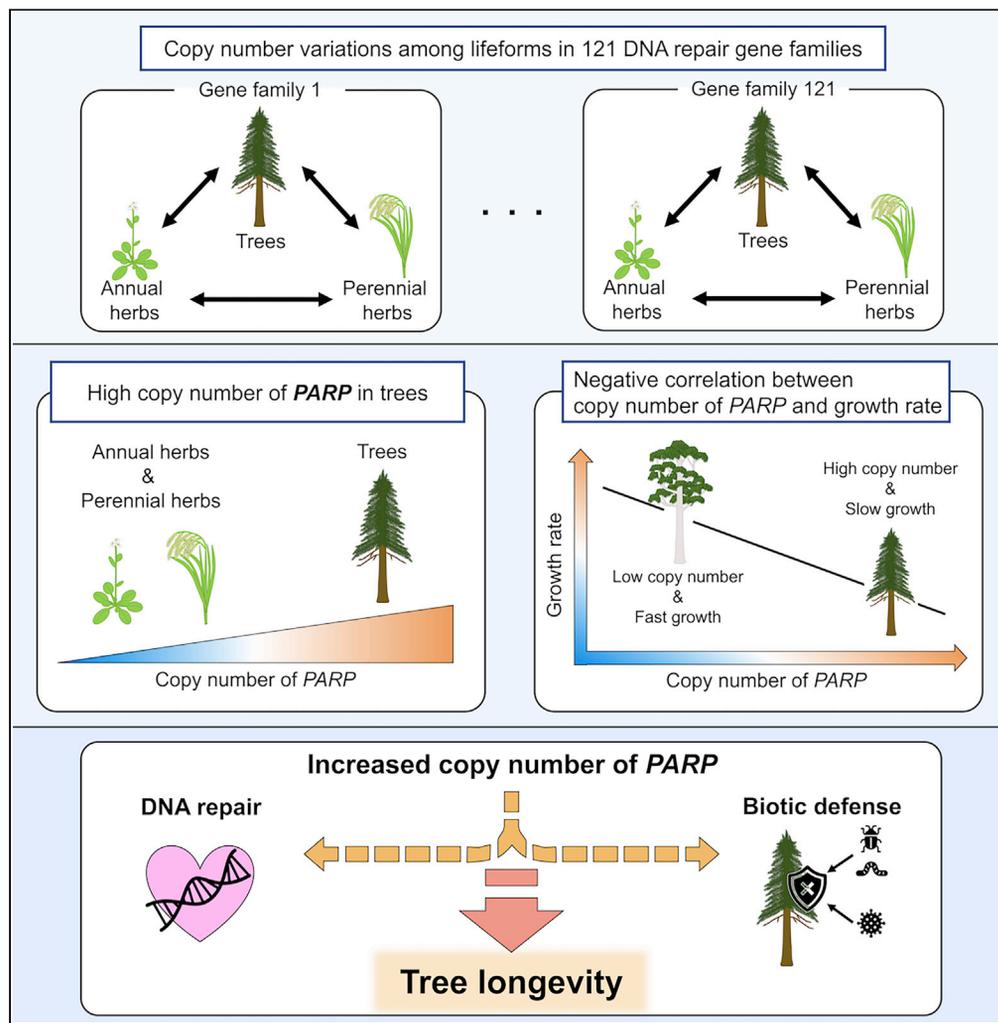


Article

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



Yuta Aoyagi Blue,  
Junko Kusumi,  
Akiko Satake

ao7yuta@gmail.com (Y.A.B.)  
akiko.satake@kyudai.jp (A.S.)

**Highlights**

Comparing the copy number variations of DNA repair genes in diverse plant species

*PARP* gene family showed higher copy number in trees compared to herbs

There was negative correlation between copy number of *PARP1* and growth rate in trees

Increased copy number of *PARP* would be evolutionary favored in plant longevity

Aoyagi Blue et al., iScience 24, 102779  
July 23, 2021 © 2021 The Author(s).  
<https://doi.org/10.1016/j.isci.2021.102779>

## Article

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

Yuta Aoyagi Blue,<sup>1,\*</sup> Junko Kusumi,<sup>2</sup> and Akiko Satake<sup>3,4,\*</sup>

## SUMMARY

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 *PARPs*, *PARP1* showed a close association with growth rate. *PARPs* catalyze poly(ADP-ribosylation) 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; Ciccio 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

<sup>1</sup>Graduate School of Systems Life Sciences, Kyushu University, 744 Motoooka, Fukuoka, 819-0395, Japan

<sup>2</sup>Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, 744 Motoooka, Fukuoka 819-0395, Japan

<sup>3</sup>Department of Biology, Kyushu University, 744 Motoooka, Fukuoka 819-0395, Japan

<sup>4</sup>Lead contact

\*Correspondence: [ao7yuta@gmail.com](mailto:ao7yuta@gmail.com) (Y.A.B.), [akiko.satake@kyudai.jp](mailto:akiko.satake@kyudai.jp) (A.S.)  
<https://doi.org/10.1016/j.isci.2021.102779>



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 thaliana*, 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- $\gamma$  (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.

## 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 STAR Methods section). We selected 171 genes involved in DNA repair within *A. thaliana* (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) (Table S2). The average of the mean copy number ratio over 121 DNA

**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 algae 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.**

	Species name
Tree: 23 species	
Angiosperm	<i>Amborella trichopoda</i>
	<i>Carica papaya</i>
	<i>Citrus clementina</i>
	<i>Coffea canephora</i>
	<i>Eucalyptus grandis</i>
	<i>Hevea brasiliensis</i>
	<i>Malus domestica</i>
	<i>Populus trichocarpa</i>
	<i>Prunus persica</i>
	<i>Pyrus bretschneideri</i>
	<i>Theobroma cacao</i>
	<i>Ziziphus jujuba</i>
Gymnosperm	<i>Cycas micholitzii</i>
	<i>Ginkgo biloba</i>
	<i>Gnetum montanum</i>
	<i>Picea abies</i>
	<i>Picea glauca</i>
	<i>Picea sitchensis</i>
	<i>Pinus pinaster</i>
	<i>Pinus sylvestris</i>
	<i>Pinus taeda</i>
	<i>Pseudotsuga menziesii</i>
	<i>Taxus baccata</i>
Perennial herb: 15 species	
	<i>Arabidopsis lyrata</i>
	<i>Brassica oleracea</i>
	<i>Cajanus cajan</i>
	<i>Capsicum annuum</i>
	<i>Erythranthe guttata</i>
	<i>Fragaria vesca</i>
	<i>Marchantia polymorpha</i>
	<i>Nelumbo nucifera</i>
	<i>Oryza sativa</i> ssp. <i>japonica</i>
	<i>Ricinus communis</i>
	<i>Selaginella moellendorffii</i>
	<i>Solanum lycopersicum</i>
	<i>Solanum tuberosum</i>
	<i>Trifolium pratense</i>
	<i>Utricularia gibba</i>
Annual herb: 21 species	
	<i>Amaranthus hypochondriacus</i>
	<i>Arabidopsis thaliana</i>

(Continued on next page)

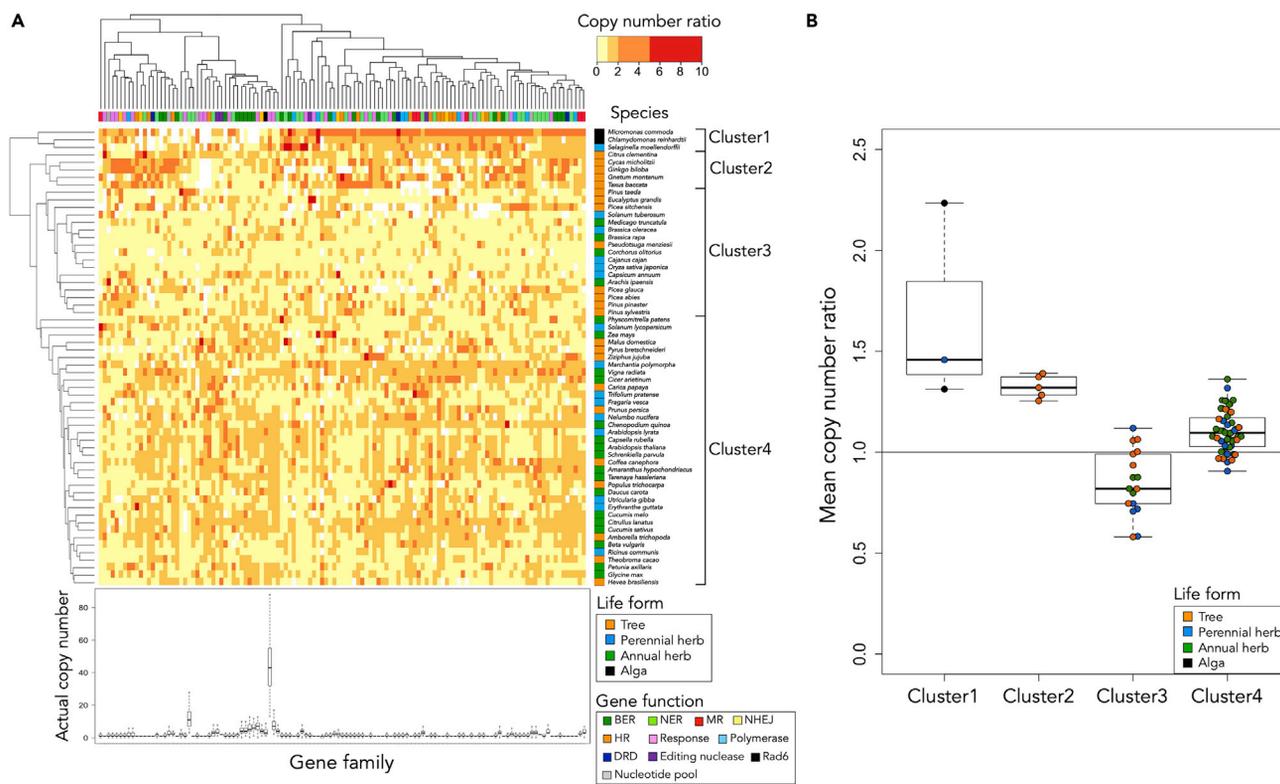
Table 1. Continued

	Species name
	<i>Arachis ipaensis</i>
	<i>Beta vulgaris</i>
	<i>Brassica rapa</i>
	<i>Capsella rubella</i>
	<i>Chenopodium quinoa</i>
	<i>Cicer arietinum</i>
	<i>Citrullus lanatus</i>
	<i>Corchorus olitorius</i>
	<i>Cucumis melo</i>
	<i>Cucumis sativus</i> L.
	<i>Daucus carota</i>
	<i>Glycine max</i>
	<i>Medicago truncatula</i>
	<i>Petunia axillaris</i>
	<i>Physcomitrella patens</i>
	<i>Schrenkiella parvula</i>
	<i>Tarenaya hassleriana</i>
	<i>Vigna radiata</i> var. <i>radiata</i>
	<i>Zea mays</i>
Alga: 2 species	<i>Chlamydomonas reinhardtii</i>
	<i>Micromonas commoda</i>

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) (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 =  $2.32 \times 10^{-4}$ , Q-value =  $4.64 \times 10^{-4}$ ) (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 (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 =  $1.42 \times 10^{-6}$ , Q-value =  $5.69 \times 10^{-6}$ ) (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 (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

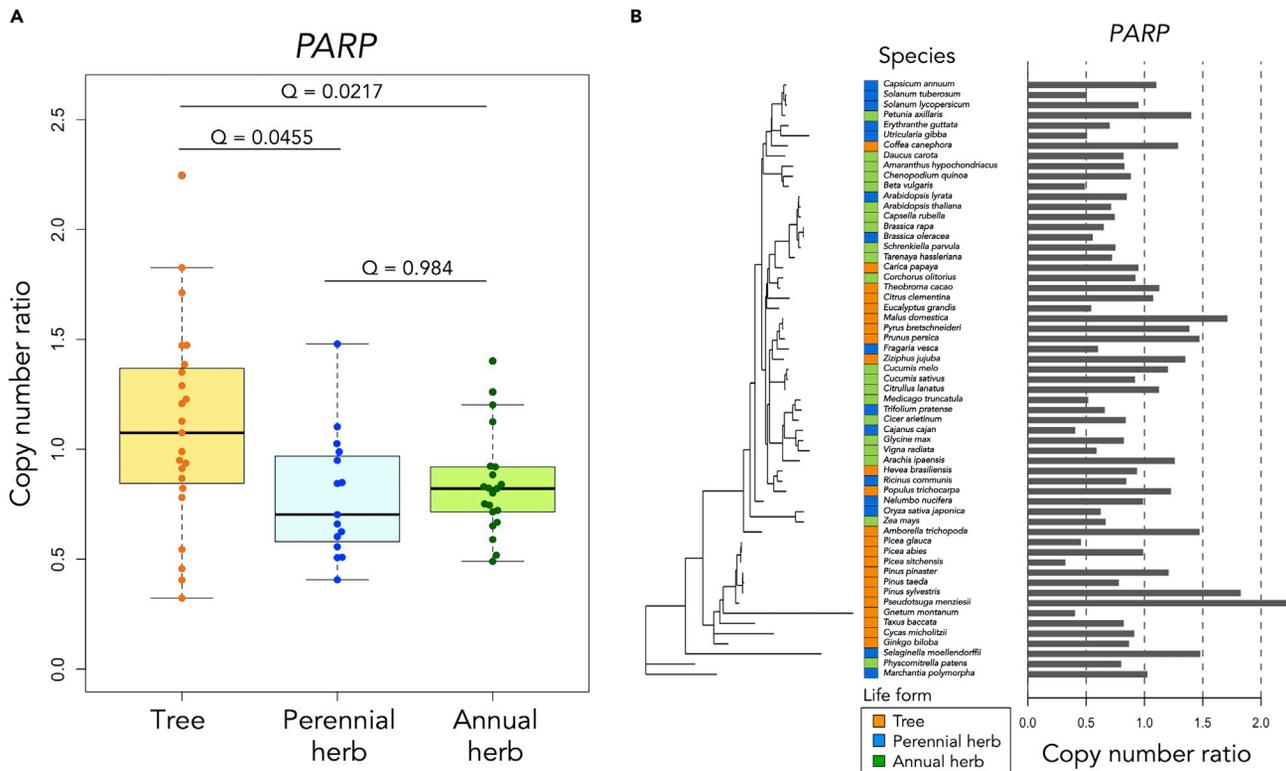


**Figure 1. Interspecies comparison of copy number ratio of 121 DNA repair gene families**  
 (A) Clustered heatmap of the copy number ratio of 121 DNA repair gene families. Hierarchical clustering was performed based on the Euclidean distance of the copy number ratio of each species using the Ward's method. 23 tree species, 15 perennial herb species, and two algae 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, non-homologous 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 range. The colors of the points correspond to the life form of the species.

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 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  $7.55 \times 10^{-5}$ ; 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



**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 range. 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  $4.97 \times 10^{-9}$ .

61 gene families (e.g., DNA glycosylase superfamily protein [*Tag*], replication protein A2 [*RPA2*] and structural maintenance of chromosomes 6 [*SMC6*]) (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 (*PARPs*) (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 (Figure S2). The three species with the highest copy number ratio of *PARPs* 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-ribosylation). Poly(ADP-ribosylation) is a covalent post-translational modification process of proteins via the synthesis and transfer of poly ADP-ribose from  $\text{NAD}^+$  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. *PARPs* are found in all eukaryotic supergroups (Citarelli et al., 2010) and *A. thaliana* encodes three canonical *PARP* proteins (At*PARP1*, At*PARP2*, and At*PARP3*).

Poly(ADP-ribosylation) 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 PARGs 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 (SROs) 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 (Figure S3), suggesting that three paralogs of the PARP genes were present in the common ancestor of angiosperms, gymnosperms, lycophytes, and bryophytes.

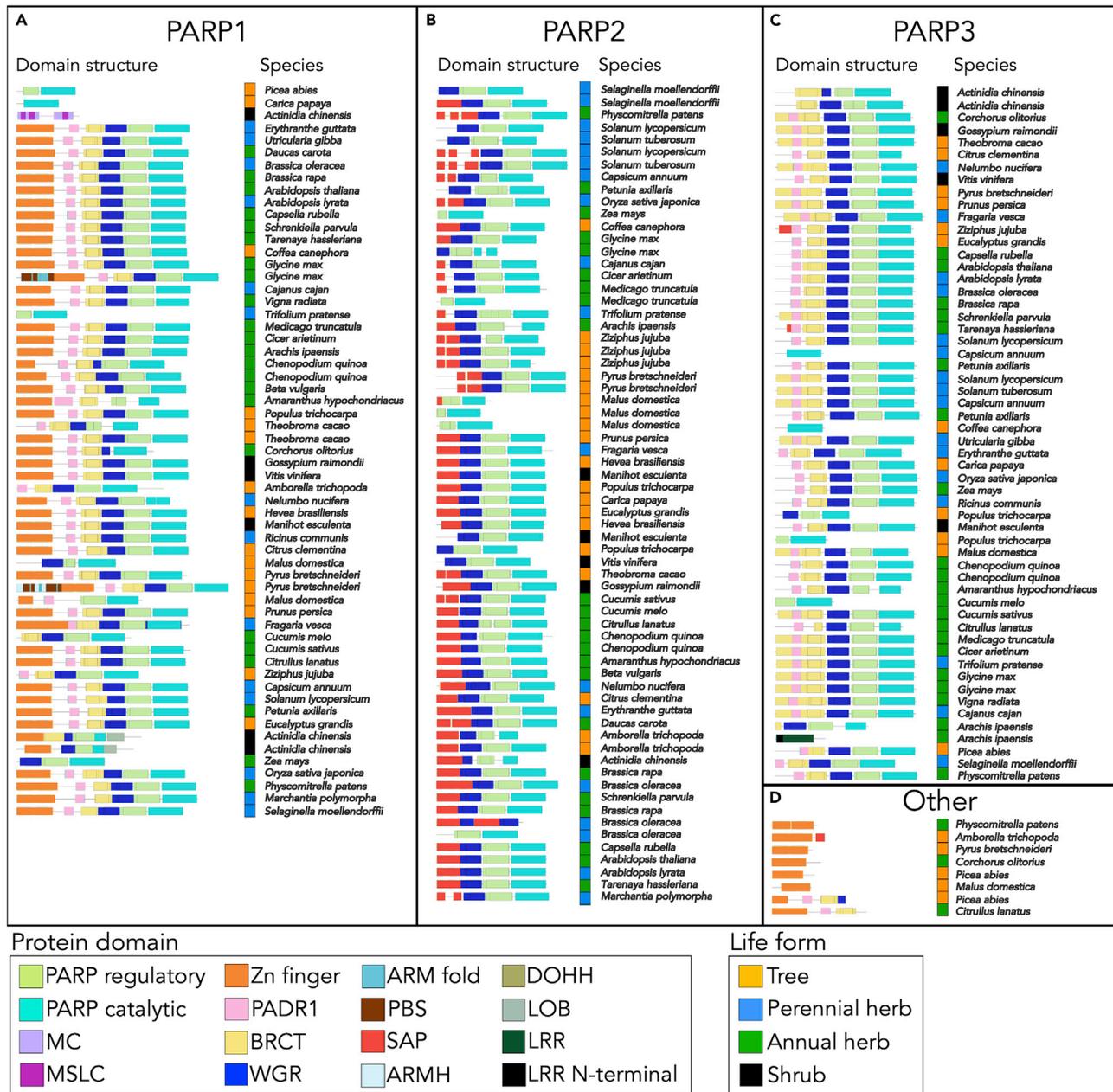
### Tree species have higher copy number ratios in PARP1 and PARP2 but not in PARP3

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 PARPs and growth rate in tree species

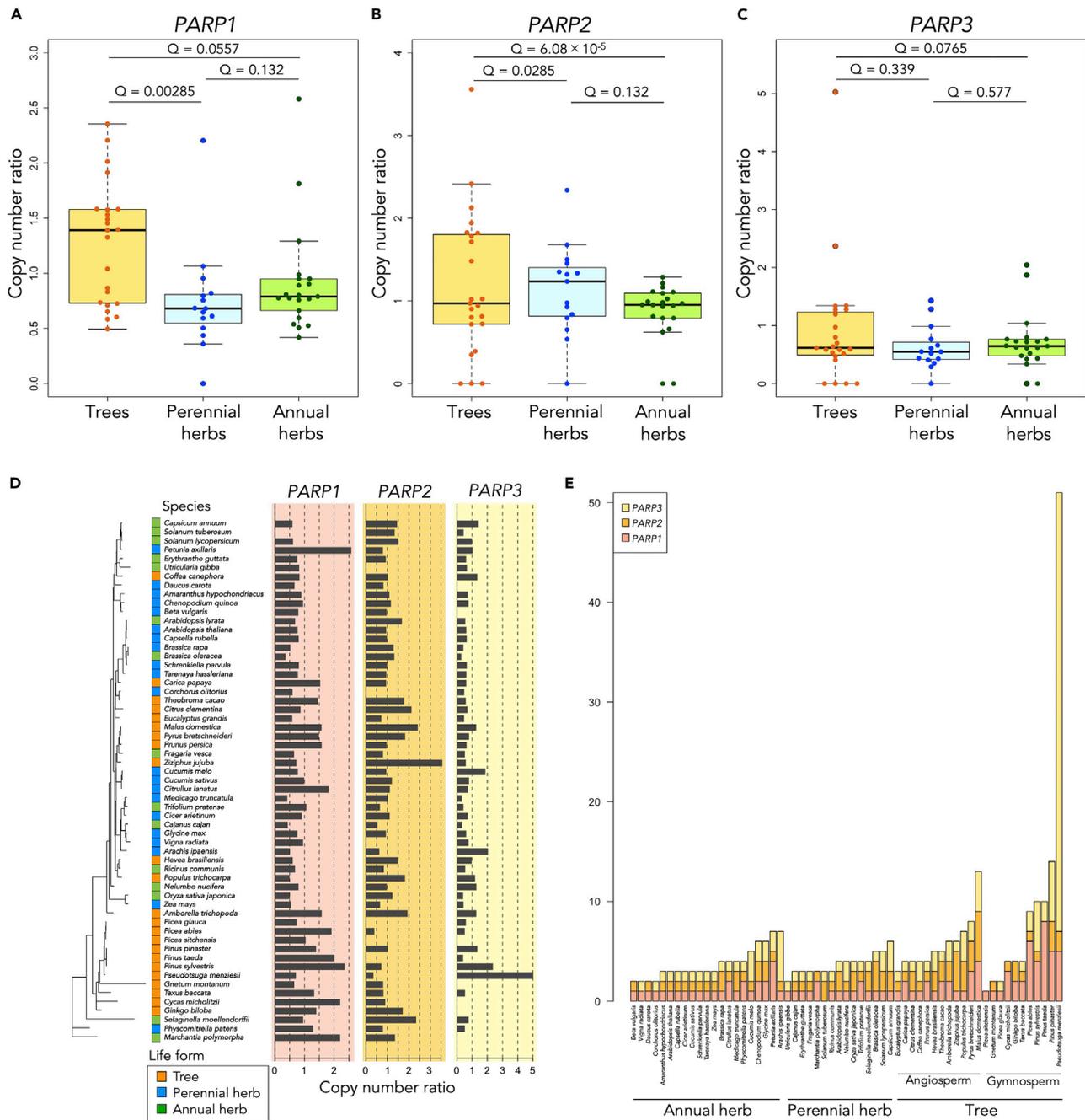
Next, we tested whether there is a significant association between copy number ratio of PARPs and longevity. Because reliable estimation of plant lifespan is difficult and maximum tree lifespans published in prestigious scientific journals are not always supported by scientific evidence (Piovesan and Biondi, 2020), we used growth rate (the rate of height increment) instead of lifespan. In the field of forest ecology,



**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.

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 (Köstler, 1956; Burns and Honkala, 1990a; 1990b; Liebhard et al., 2003;



**Figure 4. Comparison analyses for each type of PARP**

(A–C) 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 range. 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.

**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	$7.54 \times 10^{-9}$
PARP2	-0.973	0.209	-4.655	$6.1 \times 10^{-5}$	-0.686	0.21	-3.263	0.00285	0.878
PARP3	-0.316	0.175	-1.80	0.0765	-0.207	0.215	-0.964	0.339	$8.04 \times 10^{-9}$

Bravo-Oviedo et al., 2004) (Table S4) and investigated the relationship between the growth rate and the copy number ratio of PARPs 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 PARP1 for slow growth and longevity in tree species.

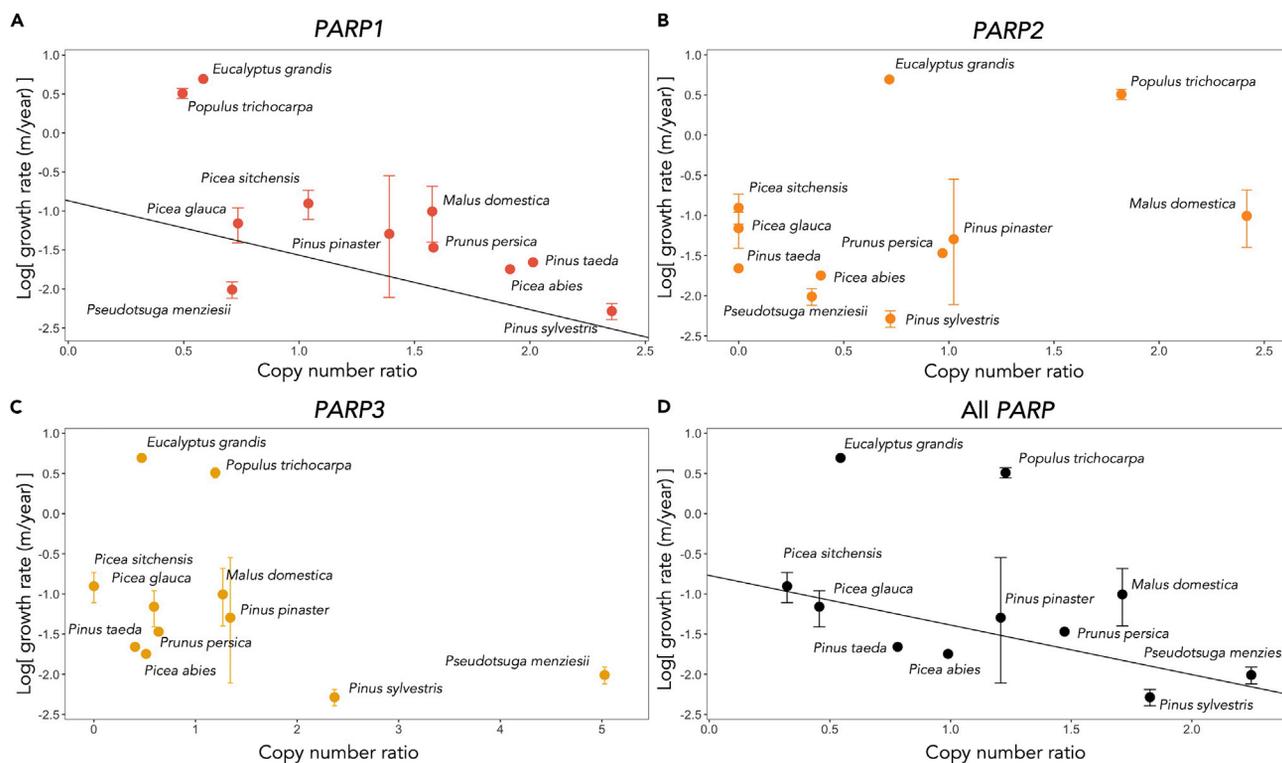
## DISCUSSION

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 PARPs. These results suggest that selection probably promotes convergent evolution of increased copy numbers of PARPs in tree species.

As key enzymes associated with poly(ADP-ribosylation), 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 PARPs 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 binding 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 (SIRT) are



**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).

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<sup>+</sup>) 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 PARPs 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 PARPs 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 PARP1s showed a significant negative correlation (Figure 5).

**Table 3. The result of regressions to investigate the relationships between growth rate and copy number ratio of PARPs 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

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 *PARPs* 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 *PARPs* 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 binding 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 *PARPs* 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 *PARPs* 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 *PARPs*, 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 researches 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.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
  - Selecting genes associated with DNA repair
  - The index of the copy number of genes for analyses
  - Construction of the phylogenetic tree
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - The similarity of the copy number ratio between species
  - PGLS to investigate the relationship between the copy number ratio and life forms
  - The evolutionary history of the *PARP* gene family
  - The relationship between copy number ratio of *PARP* and growth rate

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102779>.

### 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.

### AUTHOR CONTRIBUTIONS

Y.A., J.K., A.S.: Conceptualization. Y.A., J.K., A.S.: Methodology. Y.A.: Data analysis. Y.A., A.S.: Writing—original draft preparation with input from J.K. A.S.: Funding acquisition.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 27, 2020

Revised: May 22, 2021

Accepted: June 22, 2021

Published: July 23, 2021

## REFERENCES

- Amé, J.C., Spellenhauer, C., and de Murcia, G. (2004). The PARP superfamily. *Bioessays* 26, 882–893.
- Amor, Y., Babiychuk, E., Inzé, D., and Levine, A. (1998). The involvement of poly(ADP-ribose) polymerase in the oxidative stress responses in plants. *FEBS Lett.* 440, 1–7. [https://doi.org/10.1016/S0014-5793\(98\)01408-2](https://doi.org/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 Res.* 27, 1223–1242. <https://doi.org/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., and Kushnir, S. (1998). Higher plants possess two structurally different poly(ADP-ribose) polymerases. *Plant J.* 15, 635–645. <https://doi.org/10.1046/j.1365-313X.1998.00240.x>.
- Babiychuk, E., Van Montagu, M., and Kushnir, S. (2001). N-terminal domains of plant poly(ADP-ribose) polymerases define their association with mitotic chromosomes. *Plant J.* 28, 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. *Nat. Rev. Genet.* 8, 619–631. <https://doi.org/10.1038/nrg2158>.
- Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., and Vandepoele, K. (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol.* 158, 590–600. <https://doi.org/10.1104/pp.111.189514>.
- van Baren, M.J., Bachy, C., Reistetter, E.N., Purvine, S.O., Grimwood, J., Sudek, S., Yu, H., Poirier, C., Deerinck, T.J., Kuo, A., et al. (2016). Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* 17, 1–22. <https://doi.org/10.1186/s12864-016-2585-6>.
- Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F., and Vandepoele, K. (2018). Plaza 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46, D1190–D1196. <https://doi.org/10.1093/nar/gkx1002>.
- Bick, M.J., Carroll, J.-W.N., Gao, G., Goff, S.P., Rice, C.M., and MacDonald, M.R. (2003). Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* 77, 11555–11562. <https://doi.org/10.1128/jvi.77.21.11555-11562.2003>.
- Black, B.A., Colbert, J.J., and Pederson, N. (2008). Relationships between radial growth rates and lifespan within North American tree species. *Écoscience* 15, 349–357. <https://doi.org/10.2980/15-3-3149>.
- De Block, M., Verduyn, C., De Brouwer, D., and Cornelissen, M. (2005). Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. *Plant J.* 41, 95–106. <https://doi.org/10.1111/j.1365-313X.2004.02277.x>.
- Bravo-Oviedo, A., Río, M. Del, and Montero, G. (2004). Site index curves and growth model for Mediterranean maritime pine (*Pinus pinaster* Ait.) in Spain. *For. Ecol. Management* 201, 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 Phytol.* 168, 511–528. <https://doi.org/10.1111/j.1469-8137.2005.01548.x>.
- Briggs, A.G., and Bent, A.F. (2011). Poly(ADP-ribose)ylation in plants. *Trends Plant Sci.* 16, 372–380. <https://doi.org/10.1016/j.tplants.2011.03.008>.
- Burns, R.M., and Honkala, B.H. (1990a). *Silvics of North America. In Conifers, Volume 1* Conifers (U.S. Department of Agriculture, Forest Service), *Agriculture Handbook* 654.
- Burns, R.M., and Honkala, B.H. (1990b). *Silvics of North America. In Hardwoods, Volume 2* Hardwoods (U.S. Department of Agriculture, Forest Service), *Agriculture Handbook* 654.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* 40, 179–204. <https://doi.org/10.1016/j.molcel.2010.09.019>.
- Citarella, M., Teotia, S., and Lamb, R.S. (2010). Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evol. Biol.* 10, 1–26. <https://doi.org/10.1186/1471-2148-10-308>.
- Cove, D. (2005). *Themos Physcomitrella patens*. *Annu. Rev. Genet.* 39, 339–358. <https://doi.org/10.1146/annurev.genet.39.073003.110214>.
- Cybulski, C., Górski, B., Huzarski, T., Masojć, B., Mierzejewski, M., Dębniak, T., Teodorczyk, U., Byrski, T., Gronwald, J., Matyjasik, J., et al. (2004). CHEK2 is a multiorgan cancer susceptibility gene. *Am. J. Hum. Genet.* 75, 1131–1135. <https://doi.org/10.1086/426403>.
- Daugherty, M.D., Young, J.M., Kerns, J.A., and Malik, H.S. (2014). Rapid evolution of PARP genes suggests a broad role for ADP-ribosylation in host-virus conflicts. *PLoS Genet.* 10, e1004403. <https://doi.org/10.1371/journal.pgen.1004403>.
- Doucet-Chabeaud, G., Godon, C., Brutescio, C., De Murcia, G., and Kazmaier, M. (2001). Ionising radiation induces the expression of PARP-1 and PARP-2 genes in *Arabidopsis*. *Mol. Genet. Genomics* 265, 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. *Mutat. Res.* 435, 171–213. [https://doi.org/10.1016/S0921-8777\(99\)00050-6](https://doi.org/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. *Mutat. Res.* 728, 12–22. <https://doi.org/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* 297, 1703–1706. <https://doi.org/10.1126/science.1074276>.
- Gibson, B.A., and Kraus, W.L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.* 13, 411–424.
- Gotmare, V.P., Singh, P., and Tule, B.N. (2000). *Wild and cultivated species of cotton (CICR)*, Technical Bulletin No. 5.
- Gouveia-Oliveira, R., Sackett, P.W., and Pedersen, A.G. (2007). MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics* 8, 1–8. <https://doi.org/10.1186/1471-2105-8-312>.
- Grafen, A. (1989). The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 326, 119–157. <https://doi.org/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. *Proc. Natl. Acad. Sci. U S A* 89, 11759–11763. <https://doi.org/10.1073/pnas.89.24.11759>.
- Gu, Z., Pan, W., Chen, W., Lian, Q., Wu, Q., Lv, Z., Cheng, X., and 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 Biol.* 19, 1–18. <https://doi.org/10.1186/s12870-019-1958-9>.
- Hassa, P.O., and Hottiger, M.O. (2008). The diverse biological roles of mammalian PARPs, a small but powerful family of poly-ADP-ribose polymerases. *Front. Biosci.* 13, 3046–3082.

- Hays, J.B. (2002). Arabidopsis thaliana, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair* 1, 579–600. [https://doi.org/10.1016/S1568-7864\(02\)00093-9](https://doi.org/10.1016/S1568-7864(02)00093-9).
- Hottiger, M.O., Hassa, P.O., Lüscher, B., Schuler, H., and Koch-Nolte, F. (2010). Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* 35, 208–219.
- Huang, Z., Whelan, C.V., Dechmann, D., and Teeling, E.C. (2020). Genetic variation between long-lived versus short-lived bats illuminates the molecular signatures of longevity. *Aging* 12, 15962–15977. <https://doi.org/10.18632/aging.103725>.
- Inan, G., Zhang, Q., Li, P., Wang, Z., Coa, Z., Zhang, H., Zhang, C., Quist, T.M., Goodwin, S.M., Zhu, J., et al. (2004). 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, 1718–1737. <https://doi.org/10.1104/pp.104.041723>.
- Jaspers, P., Overmyer, K., Wrzaczek, M., Vainonen, J.P., Blomster, T., Salojärvi, J., and 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, 170.
- Johnson, S.E., and Abrams, M.D. (2009). Age class, longevity and growth rate relationships: protracted growth increases in old trees in the eastern United States. *Tree Physiol.* 29, 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* 8, 275–282. <https://doi.org/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. *Brief. Bioinform.* 20, 1160–1166. <https://doi.org/10.1093/bib/bbx108>.
- Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., et al. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* 10, 112–122. <https://doi.org/10.1016/j.celrep.2014.12.008>.
- Köstler, Josef (1956). *Silviculture (Oliver and Boyd)*, p. 416.
- Krishnakumar, R., and Kraus, W.L. (2010). The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell* 39, 8–24. <https://doi.org/10.1016/j.molcel.2010.06.017>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). Mega X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Kuny, C.V., and Sullivan, C.S. (2016). Virus–host interactions and the ARTD/PARP family of enzymes. *PLoS Pathog.* 12, 1–7. <https://doi.org/10.1371/journal.ppat.1005453>.
- Lamb, R.S., Citarelli, M., and Teotia, S. (2012). Functions of the poly(ADP-ribose) polymerase superfamily in plants. *Cell Mol. Life Sci.* 69, 175–189. <https://doi.org/10.1007/s00018-011-0793-4>.
- Li, G., Nasar, V., Yang, Y., Li, W., Liu, B., Sun, L., Li, D., and Song, F. (2011). Arabidopsis poly(ADP-ribose) glycohydrolase 1 is required for drought, osmotic and oxidative stress responses. *Plant Sci.* 180, 283–291. <https://doi.org/10.1016/j.plantsci.2010.09.002>.
- Liebhart, R., Kellerhals, M., Pfammatter, W., Jertmini, M., and Gessler, C. (2003). Mapping quantitative physiological traits in apple (*Malus x domestica* Borkh.). *Plant Mol. Biol.* 52, 511–526. <https://doi.org/10.1023/A:1024886500979>.
- Lin, Z., Kong, H., Nei, M., and Ma, H. (2006). Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl. Acad. Sci. U S A* 103, 10328–10333. <https://doi.org/10.1073/pnas.0604232103>.
- Loehle, C. (1988). Tree life history strategies: the role of defenses. *Can. J. For. Res.* 18, 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., and Stamato, T.D. (2009). Significant correlation of species longevity with DNA double strand break recognition but not with telomere length. *Mech. Ageing Dev.* 130, 784–792. <https://doi.org/10.1016/j.mad.2009.10.004>.
- MacRae, S.L., Zhang, Q., Lemetre, C., Seim, I., Calder, R.B., Hoeijmakers, J., Suh, Y., Gladyshev, V.N., Seluanov, A., Gorbunova, V., et al. (2015a). Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. *Aging Cell* 14, 288–291. <https://doi.org/10.1111/ace1.12314>.
- MacRae, L., Croken, M.M., Calder, R.B., Aliper, A., Milholland, B., White, R.R., Zhavoronkov, A., Gladyshev, V.N., Seluanov, A., Gorbunova, V., et al. (2015b). DNA repair in species with extreme lifespan differences. *Aging (Albany NY)* 7, 1171–1182.
- Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants – from models to crops. *Front. Plant Sci.* 6, 1–26. <https://doi.org/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., and Guo, H. (2013). Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog.* 9, e1003494. <https://doi.org/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, 507–515. [https://doi.org/10.1016/0092-8674\(95\)90501-4](https://doi.org/10.1016/0092-8674(95)90501-4).
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., et al. (2007). The Chlamydomonas genome reveals the evolution of key. *Natl. Institutes Health* 318, 245–250.
- Molinier, J., Lechner, E., Dumbliauskas, E., and Genschik, P. (2008). Regulation and role of arabidopsis CUL4-DDB1A-DDB2 in maintaining genome integrity upon UV stress. *PLoS Genet.* 4, e1000093. <https://doi.org/10.1371/journal.pgen.1000093>.
- Mouchiroud, L., Houtkooper, R.H., Moullan, N., Katsyuba, E., Ryu, D., Cantó, C., Mottis, A., Jo, Y.S., Viswanathan, M., Schoonjans, K., et al. (2013). XThe NAD<sup>+</sup>/sirtuin pathway modulates longevity through activation of mitochondrial UPR and FOXO signaling. *Cell* 154, 430. <https://doi.org/10.1016/j.cell.2013.06.016>.
- Müller, S., Möller, P., Bick, M.J., Wurr, S., Becker, S., Günther, S., and Kümmerer, B.M. (2007). Inhibition of filovirus replication by the zinc finger antiviral protein. *J. Virol.* 81, 2391–2400. <https://doi.org/10.1128/jvi.01601-06>.
- Murat, F., Van De Peer, Y., and Salse, J. (2012). Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biol. Evol.* 4, 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., and 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. *J. Biol. Chem.* 279, 31455–31461. <https://doi.org/10.1074/jbc.M403561200>.
- Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26, 331–348. <https://doi.org/10.1111/j.1463-6409.1997.tb00423.x>.
- Pereira-Lorenzo, S., Ramos-Cabrera, A.M., and Fischer, M. (2009). Breeding apple (*Malus x domestica* Borkh.). In *Breeding Plantation Tree Crops: Temperate Species*, P.M. Priyadarshan and S. Mohan Jain, eds. (Springer), pp. 33–81.
- Piovesan, G., and Biondi, F. (2020). On tree longevity. *New Phytol.* <https://doi.org/10.1111/nph.17148>.
- Plomion, C., Aury, J.M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillon, N., Labadie, K., Le Provost, G., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. <https://doi.org/10.1038/s41477-018-0172-3>.
- Proost, S., van Bel, M., Sterck, L., Billiau, K., van Parys, T., van de Peer, Y., and Vandepoele, K. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* 21, 3718–3731. <https://doi.org/10.1105/tpc.109.071506>.
- Rissel, D., and Peiter, E. (2019). Poly(ADP-ribose) polymerases in plants and their human counterparts: parallels and peculiarities. *Int. J. Mol. Sci.* 20, 1–29. <https://doi.org/10.3390/ijms20071638>.
- 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 Biol.* 16, 1058–1064. <https://doi.org/10.1111/plb.12167>.
- Sancar, A., Lindsey-Boltz, L.A., Ünsal-Kaçmaz, K., and Linn, S. (2004). Molecular mechanisms of

- mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85. <https://doi.org/10.1146/annurev.biochem.73.011303.073723>.
- Schreiber, V., Dantzer, F., Ame, J.C., and De Murcia, G. (2006). Poly (ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* 7, 517–528.
- Schulz, P., Neukermans, J., van der Kelen, K., Mühlenbock, P., van Breusegem, F., Noctor, G., Teige, M., Metzloff, M., and 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, e37287. <https://doi.org/10.1371/journal.pone.0037287>.
- Singh, S.K., Roy, S., Choudhury, S.R., and Sengupta, D.N. (2010). DNA repair and recombination in higher plants: insights from comparative genomics of arabidopsis and rice. *BMC Genomics* 11, 443. <https://doi.org/10.1186/1471-2164-11-443>.
- De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C.E., Maere, S., and Van De Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. U S A* 110, 2898–2903. <https://doi.org/10.1073/pnas.1300127110>.
- Song, J., and Bent, A.F. (2014). Microbial pathogens trigger host DNA double-strand breaks whose abundance is reduced by plant defense responses. *PLoS Pathog.* 10, e1004030. <https://doi.org/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 Genet.* 11, 1–24. <https://doi.org/10.1371/journal.pgen.1005200>.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Storey, J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. B Stat. Methodol.* 64, 479–498. <https://doi.org/10.1111/1467-9868.00346>.
- Storey, J., Bass, A.J., Dabney, A., and Robinson, D. (2015). Qvalue: Q-Value Estimation for False Discovery Rate Control (R Package). <http://github.com/jdstorey/qvalue>.
- Storozhenko, S., Inzé, D., Van Montagu, M., and 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. *J. Exp. Bot.* 52, 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., and 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, 1–30. <https://doi.org/10.7554/eLife.11994>.
- Takasaki, Y., Seki, Y., Nojima, H., and Isoda, A. (1994). Growth of an annual strain of oryza glaberrima steud. And a perennial cultivar of oryza sativa L. After heading. *Jpn. J. Crop Sci.* 63, 632–637. <https://doi.org/10.1626/jcs.63.632>.
- Tian, X., Seluanov, A., and Gorbunova, V. (2017). Molecular mechanisms determining lifespan in short- and long-lived species. *Trends Endocrinol. Metab.* 28, 722–734. <https://doi.org/10.1016/j.tem.2017.07.004>.
- Tivoli, B., Baranger, A., Sivasithamparam, K., and 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. *Ann. Bot.* 98, 1117–1128. <https://doi.org/10.1093/aob/mcl132>.
- Tobias, P.A., and Guest, D.I. (2014). Tree immunity: growing old without antibodies. *Trends Plant Sci.* 19, 367–370. <https://doi.org/10.1016/j.tplants.2014.01.011>.
- Tung Ho, L.S., and Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* 63, 397–408. <https://doi.org/10.1093/sysbio/syu005>.
- Umeda, M., Ikeuchi, M., Ishikawa, M., Ito, T., Nishihama, R., Kyozuka, J., Torii, K.U., Satake, A., Goshima, G., and Sakakibara, H. (2021). Plant stem cell research is uncovering the secrets of longevity and persistent growth. *Plant J.* 106, 326–335. <https://doi.org/10.1111/tj.15184>.
- Vainonen, J.P., Shapiguzov, A., Vaattovaara, A., and Kangasjärvi, J. (2016). Plant PARPs, PARGs and PARP-like proteins. *Curr. Protein Pept. Sci.* 17, 713–723. <https://doi.org/10.2174/1389203717666160419144721>.
- Vanderauwera, S., De Block, M., Van De Steene, N., Van De Cotte, B., Metzloff, M., and Van Breusegem, F. (2007). Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. *Proc. Natl. Acad. Sci. U S A* 104, 15150–15155. <https://doi.org/10.1073/pnas.0706668104>.
- Watson, J.M., and Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett.* 584, 3752–3759. <https://doi.org/10.1016/j.febslet.2010.06.017>.
- White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., et al. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* 286, 1571–1577. <https://doi.org/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, 1338–1356. <https://doi.org/10.3390/biology2041338>.
- Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., Wynne, J.W., Xiong, Z., Baker, M.L., Zhao, W., et al. (2013a). Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity. *Science* 339, 456 LP–460. <https://doi.org/10.1126/science.1230835>.
- Zhang, X.C., Nooteboom, H.P., and Kato, M. (2013b). Selaginellaceae. In *Flora of China*, Science Press, Beijing and Missouri Botanical Garden Press, St. Louis, 2, Z.Y. Wu, P.H. Raven, and D.Y. Hong, eds., pp. 37–66.
- Zhang, H., Gu, Z., Wu, Q., Yang, L., Liu, C., Ma, H., Xia, Y., and Ge, X. (2015). Arabidopsis PARG1 is the key factor promoting cell survival among the enzymes regulating post-translational poly(ADP-ribosylation). *Sci. Rep.* 5, 1–13. <https://doi.org/10.1038/srep15892>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Software and algorithms</b>		
R Version 3.6.3		<a href="http://www.r-project.org/">http://www.r-project.org/</a>
MEGA X version 10.1.5	<a href="#">Kumar et al. (2018)</a>	<a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a>
RaXML version 8.0.0	<a href="#">Stamatakis (2014)</a>	<a href="https://cme.h-its.org/exelixis/web/software/raxml/">https://cme.h-its.org/exelixis/web/software/raxml/</a>
NCBI Taxonomy Browser		<a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi</a>
Tree explore tool in Dicots PLAZA 4.0		<a href="https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/">https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/</a>
MaxAlign	<a href="#">Gouveia-Oliveira et al. (2007)</a>	<a href="http://www.cbs.dtu.dk/services/MaxAlign/">http://www.cbs.dtu.dk/services/MaxAlign/</a>
MAFFT version 7	<a href="#">Katoh et al. (2018)</a>	<a href="https://mafft.cbrc.jp/alignment/software/">https://mafft.cbrc.jp/alignment/software/</a>
BLAST+	<a href="#">Camacho et al. (2009)</a>	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download">https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download</a>
qvalue package in R version 2.16.0	<a href="#">Storey et al. (2015)</a>	<a href="http://github.com/jdstorey/qvalue">http://github.com/jdstorey/qvalue</a>
phylolm package in R version 2.6	<a href="#">Tung Ho and Ané (2014)</a>	<a href="https://CRAN.R-project.org/package=phylolm">https://CRAN.R-project.org/package=phylolm</a>
<b>Other</b>		
Copy number data of 121 gene families in 61 plant species from Gymno PLAZA 1.0 and Dicots PLAZA 4.0	Gymno PLAZA 1.0; <a href="#">Proost et al. (2009)</a>	<a href="https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/">https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/</a>
	Dicots PLAZA 4.0; <a href="#">Van Bel et al. (2018)</a>	<a href="https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/">https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/</a>
Sequence data of <i>rbcl</i> and <i>matK</i> to construction of phylogenetic tree from NCBI	National Center for Biotechnology Information (NCBI)	<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>
Data regarding the individual ages and heights in 11 tree species to calculate the growth rates	<a href="#">Köstler (1956)</a> ; <a href="#">Burns and Honkala (1990a)</a> and <a href="#">1990b</a> ; <a href="#">Liebhard et al. (2003)</a> ; <a href="#">Bravo-Oviedo et al. (2004)</a>	N/A
The database used to categorize each species included in the database into five groups according to life form	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  The Gymnosperm Database	<a href="http://plants.usda.gov/">http://plants.usda.gov/</a> <a href="http://www.plantsoftheworldonline.org/">http://www.plantsoftheworldonline.org/</a> <a href="https://pfaf.org/user/Default.aspx">https://pfaf.org/user/Default.aspx</a> <a href="https://ucjeps.berkeley.edu/">https://ucjeps.berkeley.edu/</a> <a href="http://extension.umass.edu/landscape/weed-herbarium">http://extension.umass.edu/landscape/weed-herbarium</a> <a href="http://www.mobot.org/MOBOT/research/APweb/">http://www.mobot.org/MOBOT/research/APweb/</a>  <a href="https://www.conifers.org/">https://www.conifers.org/</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Akiko Satake ([akiko.satake@kyudai.jp](mailto:akiko.satake@kyudai.jp)).

#### Materials availability

This study did not generate new unique reagents.

### Data and code availability

This paper analyses existing, publicly available data. The source of data for analyses is listed in the [key resources table](#). This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### 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 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) (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:

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

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 (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 (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 (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 (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 (Table S9 and 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) (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 (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/>).