Kingdom-wide evolutionary characterization of RNA editing factors across Archaeplastida

Ming Chen[1, 2, 3], Qingming Qu[4], Zhang Zhang[1, 2, 3, *]

[1]National Genomics Data Center, China National Center for Bioinformation, Beijing, China

[2]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

[3]University of Chinese Academy of Sciences, Beijing, China

[4]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, China

*Corresponding author: Zhang Zhang (zhangzhang@big.ac.cn)

Abstract

RNA editing, a post-transcriptional modification in plant mitochondria and plastids, is essential for environmental adaptation and diverse physiological processes. Despite extensive identification of RNA editing factors, primarily including pentatricopeptide repeat (PPR), multiple organelle RNA editing factor (MORF), organelle RNA recognition motif-containing (ORRM), and organelle zinc finger (OZ) proteins, their evolutionary history remains poorly understood. Here, we perform kingdom-wide evolutionary analyses across 364 high-quality Archaeplastida genomes and find massive PPR gene expansions in early-diverging land plants, predominantly driven by dispersed duplication associated with retroposition. Furthermore, integrative analyses indicate that DYW subgroup PPR genes have been horizontally transferred from plants to bdelloid rotifers, potentially enhancing the recipients' adaptation to ultraviolet radiation. MORF proteins, accessory partners of PPRs, possess MORF hallmark domains structurally similar to protein-folding peptidase S8 propeptide/proteinase inhibitor I9 domains, suggesting a role in protein folding during RNA editing. Considering diverse domain compositions, we reclassify MORF, ORRM, and OZ proteins and uncover prevalent hallmark domain fusions. Together, these findings illuminate the kingdom-wide evolution of plant RNA editing machinery.

Key words: RNA editing; PPR; MORF; ORRM; OZ; evolution

38 ## Introduction

39 RNA editing, a post-transcriptional modification primarily involving cytidine-to-
40 uridine (C-to-U) and uridine-to-cytidine (U-to-C) alterations in mitochondria and
41 plastids, serves as a pivotal layer of genetic regulation in plants[1,2]. This process
42 is essential for adaptation to diverse environments[3-6] and various physiological
43 processes[7], including development[8], growth[9], flowering[10], fruit ripening[11], and
44 resistance to stresses[12,13]. Editing factors, a group of nuclear-encoded proteins,
45 are assembled into an editing complex to mediate RNA editing[6]. Given the
46 remarkable diversity of editing factors in plants, elucidating deciphering their
47 evolution is essential for understanding how the RNA editing machinery
48 emerged and diversified[3,14-16].
49
50 Decades of work have been made to identify RNA editing factors, primarily
51 including Pentatricopeptide Repeat (PPR), Multiple Organelle RNA Editing
52 Factor (MORF) (also known as RNA Editing Factor Interacting Protein, RIP),
53 Organelle RNA Recognition Motif-containing (ORRM), and Organelle Zinc (OZ)
54 finger proteins[14], each characterized by hallmark domains or motifs that
55 contribute distinctly to regulating RNA editing[17]. Among them, PPR proteins are
56 key editing factors, specifically recognizing RNA sequences through P/PLS
57 motif tracts and catalyzing C-to-U and U-to-C editing via carboxyl-terminal (C-
58 terminal) DYW and DW:KP domains, respectively[18,19]. Furthermore, PPR
59 proteins are classified into two subfamilies, P and PLS, with the latter further
60 divided into PLS, E, E+, and DYW subgroups[17]. PPRs' accessory proteins,
61 including MORF, ORRM, and OZ, which are named after their MORF, RRM,
62 Ran Binding Protein 2 (RanBP2) type Zinc finger (Znf) hallmark domains,
63 respectively, coordinate RNA editing through interactions both with PPR
64 proteins and among themselves[6,10,20-27]. Beyond their hallmark domains, these

proteins also incorporate diverse additional domains, such as Glycine-Rich (GR) domain in MORF proteins[28-30] and PPR/MORF domains in ORRM proteins[31].

Building on these advances, studies have increasingly focused on the evolutionary histories of PPR, MORF, ORRM, and OZ genes[6,14,15,32,33]. The PPR gene family is one of the largest gene families in plants[19], ranging from tens of members in early-diverging chlorophytes to hundreds or thousands in land plants[4,6] and thereby reflecting massive expansions[4,34]. PPR genes also occur beyond plants, with P subfamily members widespread across eukaryotes and occasionally detected in prokaryotes[35,36]. Notably, DYW subgroup PPR genes are observed in multiple non-plant eukaryote lineages[37-40], including Amoebozoa, Discoba, Fungi, Metazoa, and Malawimonadidae, presumably resulting from horizontal gene transfer (HGT)[41]. In contrast, MORF genes are present exclusively in angiosperms and gymnosperms[34]; ORRM genes are broadly distributed across Viridiplantae lineages including chlorophytes, mosses, liverworts, and angiosperms[31]; and OZ genes are sporadically distributed in gymnosperms, mosses, and lycophytes, but are absent from chlorophytes[25]. Despite these insights, a comprehensive evolutionary characterization of these RNA editing factors across Archaeplastida remains unexplored.

Toward this end, we collect a comprehensive list of genomes across Archaeplastida, systematically identify PPR, MORF, ORRM, and OZ genes, and reconstruct an evolutionary landscape of plant RNA editing factors. Building on the landscape, we perform phylogenomic analyses to elucidate the mechanisms underlying PPR expansions and investigate the extent and evolutionary significance of HGT in DYW subgroup PPR genes. Through sequence and structural alignments, we explore the conservation and diversity

of hallmark domains of MORF proteins. Finally, we reclassify MORF, ORRM, and OZ proteins based on their domain composition.

Results

Evolutionary landscape of PPR, MORF, ORRM, and OZ genes across Archaeplastida

To systematically investigate the evolution of PPR, MORF, ORRM, and OZ genes, we obtain a total of 364 high-quality genomes across Archaeplastida that span early-diverging aquatic algae to highly derived land plants (for details see Materials and Methods), including glaucophytes (*n*=1), rhodophytes (*n*=7), prasinodermophytes (*n*=1), chlorophytes (*n*=33), charophytes (*n*=7), hornworts (*n*=22), mosses (*n*=13), liverworts (*n*=4), lycophytes (*n*=6), ferns (*n*=5), gymnosperms (*n*=13), and angiosperms (*n*=262) (Supplementary Table 1). Based on these genomes, we first identify 138 single-copy orthologous gene families (Supplementary Table 2) and then build a plant phylogenomic tree (Fig. 1a and Supplementary Fig. 1) that is topologically congruent with previous studies[42,43]. Next, we systematically identify 212,779 PPR, 2,759 MORF, 7,760 ORRM, and 829 OZ genes (Supplementary Tables 3 and 4), and evolutionarily characterize these genes in the phylogenomic tree as built above. Noticeably, we find that PPR genes are scarcely present in glaucophytes, rhodophytes, prasinodermophytes, chlorophytes, and charophytes, ranging only from 8 to 182 (Fig. 1b and Supplementary Table 4). In contrast, hundreds to thousands of PPR genes are detected in land plants, particularly some hornworts, mosses, and lycophytes, followed by gymnosperms, ferns, and angiosperms. Remarkably, two lycophytes, namely, *Selaginella moellendorffii* and *Isoetes sinensis*, have the highest number of PPR genes (*n*=3,861 and 3,256, respectively) among the 364 Archaeplastida species in our study (Fig. 1b and Supplementary Table 4). On the contrary, liverworts are short of PPR genes,

less than 100 in count (Supplementary Table 4).

Compared to PPR genes, the other three types of genes present distinct patterns. Specifically, MORF genes are found only in seed plants, including angiosperms and gymnosperms (Fig. 1c), in agreement with a previous study[34]. Additionally, seed plants generally harbor about 10 MORF genes, whereas some species, such as *Brassica napus*, *Pogostemon cablin*, and *Cissus rotundifolia*, have more MORF genes (*n*=40, 39, and 37, respectively; Supplementary Table 4). Conversely, ORRM genes are ubiquitous across Archaeplastida and exhibit a gradual increase from glaucophytes to angiosperms, indicating their ancient origin (Fig. 1d). Albeit much fewer in number than the rest three types of genes, OZ genes are enriched in angiosperms, with a minor presence in chlorophytes, hornworts, mosses, liverworts, lycophytes, ferns, and gymnosperms (Fig. 1e). Besides, many angiosperms that have undergone diploidization or polyploidization, such as *P. cablin*[44], *Crocus sativus*[45], and *Triticum aestivum*[46], exhibit concurrent expansions of PPR, MORF, ORRM, and OZ genes (Fig. 1 and Supplementary Table 4). In contrast, among these four types of genes, MORF genes are the only ones markedly expanded in drought-resistant *C. rotundifolia* relative to other angiosperms, whereas OZ genes are the only ones significantly expanded in *Crataegus pinnatifida* and *Malus fusca*. Overall, PPR genes exhibit dramatic expansions in land plants, in contrast to MORF, ORRM, and OZ genes.

PPR gene expansions in early-diverging land plants

Gene expansion can be achieved by gene duplication with four different modes, viz., whole genome duplication (WGD), tandem duplication (TD), proximal duplication (PD), and dispersed duplication (DD), which are classified according to genomic loci of duplicated gene pairs[47-49]. To explore the

148 underlying mechanisms of PPR gene expansions in early-diverging land plants,
149 we select high-quality chromosome-level genomes, including four hornworts
150 (*Anthoceros agrestis*, 2,308 PPRs; *Notothylas orbicularis*, 1,772 PPRs;
151 *Paraphymatoceros pearsonii*, 1,583 PPRs; and *Phymatoceros phymatodes*,
152 1,627 PPRs), one moss (*Takakia lepidozioides*, 2,016 PPRs), and two
153 lycophytes (*I. sinensis*, 3,256 PPRs; and *Huperzia asiatica*, 1,583 PPRs)
154 (Supplementary Table 4), with the charophyte *Zygnema circumcarinatum* (67
155 PPRs) serving as the phylogenetic outgroup. Furthermore, we select well-
156 annotated PPR genes, accounting for 93.40~100% of our identified PPRs, and
157 detect that 90.04~95.89% of them form duplicated gene pairs based on
158 sequence similarity analysis (Supplementary Table 5). Meanwhile, we infer
159 WGDs in these selected plants through comprehensive analyses of gene
160 synteny and synonymous substitution rates (Ks) of homologous gene pairs (for
161 details see Materials and Methods). Accordingly, we find that *Z.*
162 *circumcarinatum* and four hornworts lack evidence of WGD, whereas *T.*
163 *lepidozioides* and *I. sinensis* each experience one WGD, and *H. asiatica*
164 undergoes two WGDs (Supplementary Fig. 2-9). These results are consistent
165 with previous studies[5,50-54], except that *I. sinensis* has been reported to
166 experience two WGDs based solely on Ks distributions[55]. Consequently,
167 integration of inferred WGD events with comparison of genomic loci of
168 duplicated PPR gene pairs (for details see Materials and Methods) indicates
169 that DD is primarily responsible for PPR gene expansions in the selected
170 species, followed by WGD in *I. sinensis* and *H. asiatica* (Fig. 2a and
171 Supplementary Table 5).
172
173 Furthermore, it has been reported that DD is related to retroposition[47,49], a
174 process that the complementary DNA transcript of a gene is inserted back into
175 the genome, thus producing a retrogene[56]. Generally, the retrogene of a

176 duplicated pair features intron loss and a polyadenylation (polyA) tail, and the
177 pair is flanked by short direct repeats (e.g., transposable element, TE)[56,57]. Thus,
178 we examine these features of DD PPR gene pairs as indicators of retroposition.
179 Specifically, we compare intron counts of PPR genes in each dispersed
180 duplicate pair, and find that 74.78~93.21% of DD-derived PPR genes in
181 hornworts and lycophytes have undergone intron loss, followed by the moss *T.*
182 *lepidozioides* with 51.98% (Supplementary Table 6 and Fig. 2b; for details see
183 Materials and Methods). Furthermore, we investigate the presence of polyA
184 tails downstream of the 3' untranslated regions (3'UTRs) of intronless DD-
185 derived PPR genes and find that 97.99~99.26% possess polyA tails in
186 lycophytes, 70.37~75.61% in hornworts, and 81.55% in the moss *T.*
187 *lepidozioides* (Supplementary Fig. 10). Meanwhile, we identify TEs within 20
188 kilobases (kb) upstream and downstream of intronless DD PPR gene pairs,
189 showing that 70.52~99.93% of them are flanked by TEs in lycophytes,
190 26.56~68.31% in hornworts, and 80.11% in *T. lepidozioides* (Supplementary
191 Fig. 11). Notably, most of these TEs are long terminal repeats (LTRs), including
192 LTR/Copia, LTR/Gypsy, and LTR/unknown (Supplementary Table 7).
193 Additionally, we present two DD PPR gene pairs with intron loss that are flanked
194 by the same type of LTR with high sequence similarity, viz., E2 subgroup PPR
195 genes, *HaA01G1373* and *HaA43G0090* in *H. asiatica* flanked by LTR/Copia
196 (96.43% upstream and 100% downstream) (Fig. 2c), and DYW subgroup PPR
197 genes, *Tle3c03864.1* and *Tle3c03878.1* in *T. lepidozioides* by LTR/Gypsy
198 (99.86% upstream and 100% downstream) (Fig. 2d). Furthermore, co-
199 occurrence of these three retroposition features is detected in 69.93~97.92%
200 of DD PPR gene pairs in lycophytes, 17.75~51.86% in hornworts, and 64.97%
201 in *T. lepidozioides* (Fig. 2e), suggesting lineage-specific degeneration of
202 retroposition signals. Collectively, these results indicate that PPR gene
203 expansions in early-diverging land plants are primarily driven by DD associated

with retroposition.

Horizontal transfer of DYW subgroup PPR genes from plants to bdelloid rotifers

As PPR genes consist of the P and PLS subfamilies, we explore their phylogenetic patterns by reconstructing an evolutionary landscape across Archaeplastida (Supplementary Fig. 12). We find that the P subfamily PPR genes are ubiquitously distributed in Archaeplastida, with their numbers progressively increasing from early-diverging glaucophytes to highly derived angiosperms. Conversely, the PLS subfamily, including PLS, E, E+, DYW, and DYW:KP subgroups, is confined to land plants, with rare occurrence in liverworts and some mosses. In addition, the PLS and E subgroups are numerous in lycophytes, while the E+ subgroup is exclusively enriched in hornworts. The DYW subgroup is broadly distributed in land plants, with a maximum of 750 members observed in the moss *T. lepidozioides*. In contrast, its variant, the DYW:KP subgroup, is restricted to hornworts, lycophytes, and ferns (Supplementary Table 3 and Supplementary Fig. 12), conforming with a previous study[34].

Inspired by the previous findings that DYW subgroup PPR genes have also been detected in non-plants[37,41], we next investigate the evolutionary relationships of the DYW and DYW:KP subgroups between plants and non-plants. Given that repetitive and variable PLS motif triplets may hinder reliable phylogenetic inference and DYW domain catalyzes C-to-U editing via its PG box (covering the first two β-strands), the zinc ion binding signature HxE$(x)_n$CxxC, and C-terminal DYW tripeptide[58-60], we extract 6,646 DYW and 3,761 DYW:KP domains of 68 land plants as representatives of their respective subgroup PPR genes (Supplementary Table 8). Based on these plant domains,

232 we further survey non-plant DYW and DYW:KP domains across 40 lineages
233 spanning eukaryotes, prokaryotes, and viruses (Supplementary Table 9; for
234 details see Materials and Methods). In total, we detect 1,666 DYW and 6
235 DYW:KP domains in Metazoa, 31 DYW in Rhizaria, 22 DYW and 1 DYW:KP in
236 Discoba, 15 DYW in Fungi, 9 DYW in Alveolata, 3 DYW in Haptista, and 1 DYW
237 in Amoebozoa (Supplementary Table 10), exhibiting 21.21~80% sequence
238 similarity to their plant counterparts (Supplementary Table 11). To reduce
239 domain redundancy, we further select 1,151 DYW and 356 DYW:KP domains
240 from six representative species spanning major land plant lineages, including
241 *P. pearsonii* (hornwort), *T. lepidozioides* (moss), *I. sinensis* (lycophyte),
242 *Alsophila spinulosa* (fern), *Gnetum montanum* (gymnosperm), and *Amborella*
243 *trichopoda* (angiosperm), together with all non-plant DYW and DYW:KP
244 domains for phylogenetic reconstruction (Supplementary Table 10). Notably,
245 three major clusters of non-plant DYW domains are embedded within plant
246 lineages, suggesting multiple waves of HGT from plants to non-plants (Fig. 3).
247
248 To explore the potential role of DYW subgroup PPR genes in non-plants, we
249 select a representative DYW domain from each lineage and compare its
250 sequence and structure with those of plant donors. On the one hand, most of
251 these non-plant DYW domains contain the conserved HxE(x)nCxxC motif,
252 except for the HxE(x)nGxxN in Fungi, whereas the PG-box (approximately 15
253 amino acid residues (aa) beginning with proline and glycine[61]) and DYW
254 tripeptide are variable (Fig. 4a and 4b; Supplementary Fig. 13). On the other
255 hand, these non-plant DYW domains exhibit significant structural similarity to
256 those of their plant donors (Root Mean Square Deviation, RMSD ≤ 2.5Å[62]), such
257 as the DYW domain in the bdelloid rotifer *Adineta vaga* (Metazoa) and its
258 counterpart in the moss *T. lepidozioides* (Fig. 4b and Supplementary Fig. 13).
259 Surprisingly, Metazoa acquire the largest number of DYW domains from plants

260 (Fig. 3), all of which occur in bdelloid rotifers that are resistant to ultraviolet (UV)
261 radiation and inhabit diverse environments, such as mosses, lichens, soils, and
262 deserts[63,64]. Notably, these bdelloid rotifers comprise *A. vaga*, *A. ricciae*, *A.*
263 *steineri*, *Didymodactylos carnosus*, *Philodina roseola*, *Rotaria magnacalcarata*,
264 *R. socialis*, *R. sordida*, *R.* sp. 'Silwood1', and *R.* sp. 'Silwood2'. Among them,
265 we find that *A. vaga* harbors lots of PPR genes (*n*=191), nearly one third of
266 which (*n*=60) are members of the DYW subgroup (Supplementary Table 12).
267 The majority of these DYW subgroup PPR genes (58/60) are expressed (Fig.
268 4c), suggesting their potential roles in biological processes. Intriguingly, the
269 moss *T. lepidozioides*, a potential donor of DYW subgroup PPR genes to *A.*
270 *vaga*, is also resistant to UV radiation[5]. Consistent with this resistance, its DYW
271 genes are upregulated during recovery from UV irradiation (Fig. 4d and
272 Supplementary Table 13), highlighting their potential roles in response to UV
273 radiation. Taken together, these results suggest that bdelloid rotifers living in
274 mosses may have acquired the DYW subgroup PPR genes as an adaptation
275 to high UV radiation (Fig. 4e).
276

277 **Evolutionary conservation and diversity of MORF domains**

278 MORF proteins, exclusive to seed plants (Fig. 1c) and comprising nine
279 members (MORF1~MORF9) in *A. thaliana*, are characterized by their MORF
280 hallmark domains and are essential for PPR-mediated RNA editing[28-30]. To
281 explore the evolutionary conservation of MORF domains, we extract 334
282 instances from 33 representative seed plants (Supplementary Table 14) and
283 identify their homologs across 45 lineages spanning eukaryotes, prokaryotes,
284 and viruses (Supplementary Table 9). We find that MORF domains are
285 homologous to peptidase S8 propeptide/proteinase inhibitor I9 domains (also
286 known as PA domain[28]) in metazoa, fungi, bryophytes, lycophytes, and ferns
287 (Fig. 5a and Supplementary Table 15). Furthermore, we compare the structures

288 of the MORF domain and its homologous PA domain with the highest sequence
289 similarity in each of these five lineages. We find that although they have
290 relatively low sequence similarity (33.33~41.94%), their structures are
291 significantly similar (RMSD ≤ 2.5Å), with RMSD values varying from
292 0.705~1.039Å (Fig. 5b and 5c; Supplementary Fig. 14). Notably, MORF
293 domains from MORF3 and MORF8 proteins share the same α-helix and β-
294 strand architecture as PA domains (Fig. 5c and Supplementary Fig. 14). As PA
295 domain is important in assisting fold zymogens and mature peptidases[65,66],
296 these results imply that MORF domains from MORF3 and MORF8 proteins may
297 help fold PPR proteins to regulate RNA editing.
298
299 To investigate the evolutionary diversity of MORF domains, we classify them
300 into MORF1 to MORF9 as identified in *A. thaliana* and survey PA domains
301 across our collected 364 high-quality Archaeplastida genomes. Accordingly, PA
302 domains are ubiquitous across major lineages, including glaucophytes,
303 charophytes, mosses, liverworts, hornworts, lycophytes, ferns, gymnosperms,
304 and angiosperms (Supplementary Fig. 15). Furthermore, we reconstruct a
305 phylogenetic tree of the 334 MORF domains, using five non-seed plant PA
306 domains with the highest sequence similarity as the outgroup. Consequently,
307 MORF2, MORF5, and MORF6 domains cluster into early-diverging clades in
308 the phylogenetic tree, indicating their earlier origins relative to other MORF
309 domains in seed plants (Fig. 5d). Moreover, most MORF domains (MORF3~6,
310 8, and 9) are broadly distributed across angiosperms and gymnosperms,
311 suggesting that MORF diversification predates the major divergences among
312 seed plant lineages. At the gene level, we analyze the expression of
313 MORF1~MORF9 genes, each harboring a single MORF hallmark domain,
314 across multiple tissues of the gymnosperm *Torreya grandis* and the angiosperm
315 *A. thaliana*. Notably, among these MORF genes, MORF8 and MORF9 display

316 higher expression across tissues in both species, consistent with observations
317 in *Apium graveolens* and *Populus trichocarpa*[67,68], indicating their conservation
318 in expression (Supplementary Fig. 16).
319

320 Hallmark domain fusions among MORF, ORRM, OZ, and PPR 321 proteins

322 While MORF proteins contain the MORF hallmark domain, the presence of
323 additional domains, such as the GR domain in MORF1, MORF4, and MORF8
324 proteins[29], suggests the novel domain composition and underscores the need
325 for an updated classification of MORF proteins. Thus, in terms of domain
326 composition, we classify our identified 2,759 MORF proteins into four groups:
327 (1) Single-Pure MORF (SP-MORF, *n*=1,916), containing a single MORF domain;
328 (2) Single-Mixed MORF (SM-MORF, *n*=544), containing a single MORF domain
329 plus one or more non-MORF domains; (3) Multiple-Pure MORF (MP-MORF,
330 *n*=42), containing multiple MORF domains; and (4) Multiple-Mixed MORF (MM-
331 MORF, *n*=257), containing multiple MORF domains plus one or more non-
332 MORF domains (Fig. 6a and Supplementary Table 16).
333
334 Furthermore, we reconstruct the evolutionary landscape of these reclassified
335 MORF proteins across 262 angiosperms and 13 gymnosperms (Fig. 6b).
336 Specifically, SP-MORF proteins are widespread in gymnosperms and
337 angiosperms with larger numbers, whereas SM-MORF proteins are abundant
338 in angiosperms but rare in gymnosperms. Conversely, MP-MORF proteins
339 occur sparsely in both lineages and MM-MORF proteins are only observed in
340 angiosperms. Importantly, these reclassified MORF proteins exhibit a wide
341 range of domain compositions. Among them, 492 SM-MORF and 9 MM-MORF
342 proteins harbor a GR domain alongside MORF1~6, 8, and 9 domains in
343 angiosperms, whereas in the gymnosperm *T. grandis*, a single SM-MORF

344 protein contains a GR domain co-occurring with MORF9 domain, extending the
345 previously observed MORF-GR domain architectures in MORF1, MORF4, and
346 MORF8 proteins[29]. Surprisingly, we find that MORF domains and P motifs are
347 fused in two SM-MORF proteins, viz., AY_017782_RA in the angiosperm
348 *Aquilaria yunnanensis* and KAH9320946.1 in the gymnosperm *Taxus chinensis*
349 (Supplementary Table 16). Moreover, MORF domain and the RRM hallmark
350 domain of ORRM are fused in SM- and MM-MORF proteins in angiosperms.
351 Specifically, these domain fusions fall into four types: 1-MORF & 1-RRM (*n*=23),
352 1-MORF & 3-RRM (*n*=1), 2-MORF & 1-RRM (*n*=243), 3-MORF & 1-RRM (*n*=3)
353 (Supplementary Fig. 17 and Supplementary Table 16), where the leading
354 number indicates the copy number of each corresponding domain. Notably, the
355 most abundant fused type, 2-MORF & 1-RRM, is represented by the RNA
356 editing factor ORRM1[69].
357
358 ORRM proteins, specifically, ORRM1 to ORRM6, have been identified in
359 several angiosperms as mediators of RNA editing[23,70]. Likewise, we reclassify
360 our identified 7,760 ORRM proteins into SP-ORRM (*n*=2,689), SM-ORRM
361 (*n*=1,751), MP-ORRM (*n*=2,948), and MM-ORRM (*n*=372) (Supplementary
362 Table 17). Furthermore, we reconstruct their evolutionary landscape across
363 Archaeplastida, and find that SP-, SM-, and MP-ORRM proteins exhibit a
364 gradual increase from glaucophytes to angiosperms with larger number, while
365 MM-ORRM proteins show a sparse distribution (Supplementary Fig. 18). In
366 addition to previously identified ORRM1 to ORRM6 with a single RRM domain,
367 MP-ORRM proteins, such as CP29A, CP31A, and CP31B, each containing two
368 RRM domains, also affect RNA editing[23]. Intriguingly, numerous RRM and PPR
369 domains are fused in SM- and MM-ORRM proteins across Archaeplastida
370 (Supplementary Fig. 19). Specifically, SM-ORRM proteins contain the highest
371 number of RRM-P fusions in angiosperms (*n*=171), followed by hornworts (*n*=6),

372 mosses (*n*=4), liverworts (*n*=4), charophytes (*n*=3), gymnosperms (*n*=3),
373 lycophytes (*n*=1), and ferns (*n*=1). Moreover, SM-ORRM proteins exhibit one
374 RRM-DYW fusion in hornworts and angiosperms, and three RRM-E fusions in
375 angiosperms (Supplementary Table 17). Conversely, MM-ORRM proteins have
376 only five RRM-P and one RRM-E fusions in angiosperms. In addition, fusions
377 of RRM and the RanBP2 Znf hallmark domain of OZ protein are detected in
378 proteins that are not targeted to mitochondria or plastids (Supplementary Table
379 18).
380
381 Similarly, we reclassify OZ proteins and explore their evolution using the same
382 approach applied to MORF and ORRM proteins. Consequently, we identify 761
383 MP-OZ, 61 MM-OZ, 6 SP-OZ, and 1 SM-OZ proteins (Supplementary Table 19).
384 Based on their evolutionary landscape across Archaeplastida, we find that MP-
385 OZ proteins are enriched in angiosperms (*n*=725), followed by mosses (*n*=14),
386 lycophytes (*n*=6), ferns (*n*=6), gymnosperms (*n*=4), chlorophytes (*n*=3) and
387 liverworts (*n*=3) (Supplementary Fig. 20). Additionally, 60 MM-OZ, 6 SP-OZ,
388 and 1 SM-OZ proteins are found in angiosperms, while a single MM-OZ occurs
389 in hornworts. Notably, we detect that RanBP2 Znf and PPR domains (P and E)
390 are fused in four angiosperm proteins, despite their absence of targeting to
391 mitochondria or plastids (Supplementary Table 20). Collectively, our results
392 indicate that hallmark domains of PPR, MORF, ORRM, and OZ proteins can be
393 fused with more diverse patterns than previously thought (Fig. 6c).

394 Discussion

395 Our kingdom-wide characterization of RNA editing factors across
396 Archaeplastida reveals a complex evolutionary trajectory defined by massive
397 gene expansions, HGT, and the diversity and fusion of hallmark domains.
398

399 PPR genes, the primary mediators of plant RNA editing, exhibit dramatic
400 expansions in hornworts, mosses, and lycophytes. We demonstrate that these
401 expansions are predominantly driven by dispersed duplication associated with
402 retroposition, evidenced by signatures such as intron loss, polyA tails, and
403 flanking TEs. In contrast, WGD plays a subsidiary role in PPR expansion in
404 lycophytes and contributes little to that in the moss *T. lepidozioides* (Fig. 2a).
405 Notably, *T. lepidozioides* exhibits predominant dispersed duplication but limited
406 retroposition features, suggesting large-scale gene fractionation after an
407 ancient WGD event[71]. These results imply that retroposition, as well as WGD,
408 provided the genomic plasticity necessary for early land plants to diversify their
409 RNA editing machinery during the colonization of harsh terrestrial
410 environments[4].
411
412 The evolutionary reach of plant RNA editing extends far beyond Archaeplastida.
413 The discovery of DYW subgroup PPR genes in diverse non-plant lineages—
414 including Metazoa, Amoebozoa, Fungi, Haptista, and Rhizaria—indicates that
415 DYW-mediated C-to-U editing is a broadly distributed eukaryotic mechanism,
416 which is more common than previously recognized[41,72,73]. Noticeably, the high
417 expression of these genes in both recipient bdelloid rotifers (Metazoa) and their
418 putative plant donors (e.g., *T. lepidozioides*), coupled with shared resistance to
419 UV radiation, suggests a convergent adaptation where RNA editing may
420 facilitate the repair of UV-induced damage[74,75]. This highlights the functional
421 redeployment of editing factors following cross-kingdom transfer to meet
422 specific environmental pressures.
423
424 MORF proteins, restricted to seed plants, display significant structural similarity
425 between their hallmark MORF domains and PA domains of other proteins.
426 Specifically, MORF3 and MORF8 domains more closely resemble PA domains

427 associated with protein folding, whereas MORF1 domain has been reported to
428 be more structurally similar to an N-terminal ferredoxin-like domain that confers
429 RNA substrate positioning in bacterial 4-thio-uracil tRNA synthetases[28]. Our
430 results suggest that the structural diversity of MORF domains underpins their
431 functional versatility in RNA editing regulation.
432
433 Prompted by the structural diversity observed among MORF domains, we
434 extended our analysis to MORF, ORRM, and OZ proteins, systematically
435 reclassifying them based on domain composition. This reclassification revealed
436 not only a broader repertoire of potential RNA editing factors, but also extensive
437 hallmark domain fusions among PPR, MORF, ORRM, and OZ proteins (except
438 between MORF and OZ), mirroring their reported interactions[17,24,26,27,29] and
439 exemplifying the general link between domain fusion and protein-protein
440 networking[76-78]. Based on these findings, we propose an evolutionary model
441 where the composition of the editing machinery has progressively diversified
442 from hornworts to angiosperms, likely contributing to the maintenance of
443 efficient RNA editing (Extended Data Fig. 1). Collectively, these insights provide
444 a refined view of the genomic innovations that have shaped the plant RNA
445 machinery.

446 **Materials and Methods**

447 **Collection and quality control of Archaeplastida genomes**

448 To estimate the numbers of PPR, MORF, ORRM, and OZ genes and
449 reconstruct their evolutionary landscapes, 364 Archaeplastida high-quality
450 genomes were retrieved from public databases (Supplementary Table 1). To
451 ensure accurate gene counts, only the longest protein isoform of each gene
452 was retained. The completeness of annotated gene sets was estimated by
453 Benchmarking Universal Single-Copy Orthologue (BUSCO) (version 5.8.2) with

454 three datasets chlorophyta_odb10, viridiplantae_odb10, and
455 eukaryota_odb10[79]. As these datasets contain abundant sequences of
456 chlorophytes and angiosperms, genomes in these two lineages with BUSCO
457 scores ("complete") greater than 90% were retained. For other lineages,
458 chromosome-level genomes with BUSCO scores ("complete") greater than
459 85%; otherwise, the top 20% genomes with the highest BUSCO scores were
460 retained within each lineage.

461 Phylogenomic tree reconstruction of Archaeplastida

462 To reconstruct the phylogenomic tree of 364 Archaeplastida species, single-
463 copy orthologous gene families were inferred by OrthoFinder (version
464 3.0.1b1)[80], and 138 single-copy orthologous gene families covering at least 75%
465 species were retained. These single-copy orthologous gene families were
466 aligned by MAFFT (version 7.453)[81] with the parameter "--auto", and poorly
467 aligned regions were identified and trimmed by BMGE (version 1.12) with
468 BLOSUM30[82]. Gene trees were constructed by IQ-TREE (version 2.1.2)[83], with
469 the best-fit substitution model for each gene automatically selected according
470 to Bayesian Information Criterion, and using parameters "-B 1000 -bnni -m MFP
471 -T AUTO -ntmax 20". The resulting trees were subsequently merged, and the
472 species tree was inferred by ASTRAL (version 5.7.8)[84] using *Cyanophora*
473 *paradoxa* as the outgroup.

474 Identification of PPR, MORF, ORRM, and OZ proteins

475 The hallmark domains and motifs of PPR (P, P1, P2, S1, S2, SS, L1, L2, E1,
476 E2, E+, DYW, and DYW:KP), MORF (MORF), ORRM (RRM), and OZ (RanBP2
477 Znf) proteins were identified using hmmsearch in HMMER (version 3.2.1)[85] with
478 corresponding domain and motif profiles. PPR hallmark domain and motif
479 profiles were obtained from a previous study[34] and searched using hmmsearch
480 with parameters "--domtblout -noali -E 0.1". All hmmsearch scores for P, P1, P2,

481 S1, S2, L1, L2, E1, E2, and E+ domains or motifs were greater than 0, with SS
482 motif greater than 10, and DYW/DYW:KP domain greater than 30. As most PPR
483 hallmark domain-containing proteins are localized to mitochondria and/or
484 plastids[19,86], they are defined as PPR proteins. In addition, PPR proteins were
485 required to contain at least three of P, P1, P2, L1, L2, or SS motifs, unless they
486 contained a DYW or DYW:KP domain.
487
488 The MORF domain profile, constructed using hmmbuild in HMMER (version
489 3.2.1) with nine MORF domains from *A. thaliana*, was searched using
490 hmmsearch with parameters "--domtblout -noali -E 1e-10". As MORF domain
491 was homologous to PA domain, only hits with hmmsearch score greater than
492 36 were kept. The RRM domain profile was obtained from Pfam (PF00076.26)[87]
493 and searched using hmmsearch with parameters "--domtblout -noali -E 1e-10".
494 The hmmsearch score was greater than 0. GR domains, present in some
495 MORF and ORRM proteins, were defined as regions containing at least eight
496 glycines within a 20-amino-acid sliding window. The RanBP2 Znf profile was
497 obtained from Pfam (PF00641.22) and searched using hmmsearch with the
498 same parameters as for the RRM domain. As MORF, ORRM, and OZ proteins
499 were localized to mitochondrion and/or plastids, their subcellular localization
500 was predicted using DeepLoc (version 2.1)[88] with "Accurate" mode, TargetP
501 (version 2)[89], Plant-mSub[90] (https://bioinfo.usu.edu/Plant-mSubP), and
502 LOCALIZER[91] (https://localizer.csiro.au). Only localization results supported by
503 at least two tools were retained. The domain composition was identified using
504 InterPro (https://www.ebi.ac.uk/interpro)[92].

505 **WGD Inference**

506 To infer WGD, intra-genome alignments were conducted using BLASTP in the
507 Basic Local Alignment Search Tool (BLAST, version 2.12.0+)[93] with an E-value
508 cutoff of 1e-5. Based on the alignments, homologous dotplots were generated

509 using WGDI[94], and synteny blocks were inferred using MCScanX[95]. Ks values
510 of homologs in synteny blocks were calculated using KaKs_Calulator (version
511 3.0)[96] and ParaAT (version 2.0)[97] with parameters "-p proc -m mafft -f axt -g -k".
512 The median Ks value for each synteny block was calculated. Collectively, Ks
513 peak counts and synteny blocks in the homologous dotplot were used to infer
514 the number of WGD events.

515 Identification and classification of duplicated gene pairs

516 To identify duplicated gene pairs, all protein sequences within a genome were
517 aligned using BLASTP (BLAST, version 2.12.0+) with an E-value cutoff of 1e-5.
518 The highest-scoring alignment for each gene was considered a duplicated pair.
519 Alignments covering more than 60% of the longer gene (query or subject) were
520 required to have a sequence similarity exceeding 20%. Alternatively, for
521 alignments covering less than 60% of the longer gene, the alignment was
522 required to span at least 105 amino acid residues with a sequence similarity
523 exceeded 40%. Otherwise, the query and subject genes were considered
524 singletons. Based on genomic loci, duplicated gene pairs were classified into
525 four categories: tandem pairs (adjacent genes), proximal pairs (within 10 loci
526 on the same chromosome), dispersed pairs (more than 10 loci apart on the
527 same chromosome or on different chromosomes), and WGD pairs (located in
528 synteny blocks).

529 TE annotation

530 TEs were identified by a combination of *ab initio* and homology-based methods.
531 For *ab initio* prediction, a consensus sequence library was built using
532 RepeatModeler (version 2.0.6)[98] with the parameter "-LTRStruct". An LTR
533 library was then constructed using LTR_Finder (version
534 1.07)[99]/LTR_FINDER_parallel (version 1.1)[100], LTRharvest (version 1.6.6)[101]
535 with parameters "-minlenltr 100 -maxlenltr 7000", and LTR_retriever (version

536 2.9.0)[102]. Both libraries were used to annotate TEs using RepeatMasker
537 (version 4.1.8)[103]. For homology-based detection, TEs were identified using
538 RepeatMasker (version 4.1.8) with Dfam (version 3.9)[104] database. TE
539 annotations obtained from both methods were then combined.

540 Retroposition inference

541 Retroposition, underlying dispersed duplication gene pairs, was characterized
542 by intron loss, polyA enrichment downstream of the 3'UTR, and flanking TEs.
543 First, intron loss in one dispersed duplication gene pair was required to meet
544 one of following criteria: (1) both genes had no intron; (2) a gene had no intron,
545 while the other had at least one intron; (3) considering intron retention, a gene
546 contained one intron, while the other had at least one intron; (4) both genes had
547 more than one intron, differing by more than 10 introns. Second, polyA
548 enrichment downstream of the 3'UTR of the intronless gene (or polyT
549 enrichment upstream of the 5'UTR) was defined as at least eight
550 adenines/thymines within a 10-base window. Genomic annotations were used
551 to detect polyA/T enrichment in regions extending 1 kb downstream of the
552 3'UTR or upstream of the 5'UTR. If the 3'UTR or 5'UTR of a gene was not
553 annotated, the shortest corresponding UTR in the genome was used. If no
554 3'UTR or 5'UTR was annotated in the genome, they were defined as the 100-
555 bp regions downstream of the stop codon and upstream of the start codon,
556 respectively. Third, dispersed duplication gene pairs with TEs located within 20
557 kb upstream and downstream were defined as TE-flanked.

558 Identification of DYW and DYW:KP domain homologs and HGT 559 inference

560 To explore HGT of plant DYW and DYW:KP domains, 10,407 protein
561 sequences (hmmsearch score ≥ 30, length ≥ 130 aa) of 68 representative
562 species were collected (Supplementary Table 8). DYW and DYW:KP domain

563 homologs were searched in the NCBI Non-redundant Protein Sequence
564 Database (NR) (acquisition date: 2025.01.19) using BLASTP (BLAST, version
565 2.12.0+) with an E-value cutoff of 1e-5. Protein sequences retrieved from NR
566 originated from 40 lineages (Supplementary Table 9). Homologs exhibiting
567 more than 20% sequence similarity to plant DYW and DYW:KP domains were
568 further confirmed via hmmsearch in HMMER (version 3.2.1) using a score cutoff
569 of 30. To avoid contamination, non-plant DYW and DYW:KP domains were
570 retained only if present in at least two species within the corresponding lineage.
571 Otherwise, non-plant DYW and DYW:KP domains detected in a single species
572 were retained only if present in at least one additional genome assembly.
573
574 To reconstruct the phylogenetic tree of DYW and DYW:KP domains from plants
575 and non-plants, domains from six representative plants (*P. pearsonii*, *T.*
576 *lepidozioides*, *I. sinensis*, *A. spinulosa*, *G. montanum*, and *A. trichopoda*) and
577 all non-plant counterparts were aligned using MAFFT (version 7.525), with
578 poorly aligned regions trimmed by trimAl (version 1.5.rev0)[105] using parameters
579 "-gt 0.7 -cons 60 -st 0.1". The phylogenetic tree was inferred using IQ-TREE
580 (version 2.1.2) (Best-fit model: JTT + R10) with parameters "-B 1000 -bnni -m
581 MFP -T AUTO --safe" and visualized with iTOL (version 7)[106]. Bootstraps were
582 from 1,000 replicates. Multiple sequence alignment of DYW domains from
583 plants and non-plants was performed using MEGA (version 11)[107] and
584 visualized with Jalview (version 2.11.4.0)[108].

585 **Identification of MORF domain homologs**

586 To identify homologs of plant MORF domains, 334 protein sequences
587 (hmmsearch score ≥ 0, length ≥ 86 aa) from 33 representative species were
588 collected (Supplementary Table 14). These MORF domain homologs searched
589 in the NCBI NR using BLASTP (BLAST, version 2.12.0+) with an E-value cutoff
590 of 1e-5. The protein sequences retrieved from NR originated from 45 lineages

591 (Supplementary Table 9). Homologs with more than 20% similarity to plant
592 MORF domains were retained. The MORF domains and their homologs were
593 aligned using MAFFT (version 7.525), and the phylogenetic tree was
594 constructed using FastTree[109] (Model: Jones-Taylor-Thorton, CAT
595 approximation with 20 rate categories) with parameters "-pseudo -gamma" and
596 visualized by iTOL (version 7).

597 Phylogenetic analysis of MORF domains

598 To reconstruct the phylogenetic tree of MORF domains, all MORF domains from
599 gymnosperms and angiosperms were selected. These domains were aligned
600 to MORF1~MORF9 domains from *A. thaliana* to assign their types. Five PA
601 domains from non-seed plants (Dicom.22G070000.1.p, HaB36G042,
602 Phcar.S3G189600.t1, Phsp.C2G192800.t1, XP_024533311.1), exhibiting the
603 highest similarity to MORF domains, were used as the outgroup. All MORF
604 domains, along with the outgroup sequences, were aligned using MAFFT
605 (version 7.525). The phylogenetic tree was inferred using IQ-TREE (version
606 2.1.2) (Best-fit model: JTT + R8) with parameters "-B 1000 -bnni -m MFP -T
607 AUTO" and visualized by iTOL (version 7)[106]. Bootstraps were from 1,000
608 replicates.

609 Analysis of PPR and MORF gene expression

610 To calculate gene expression, reference genome indexes were built using
611 hisat2-build in HISAT2 (version 2.0.5)[110] and rsem-prepare-reference in RSEM
612 (version 1.3.1)[111]. Raw RNA sequencing data were quality-filtered using fastp
613 (version 0.20.1)[112] with parameters "-q 20 -u 40 -l 15 -g -x -r -W 4 -M 20 -w 12".
614 Strand specificity was inferred using infer_experiment.py in RSeQC (version
615 2.6.4)[113]. Reads mapping was conducted using STAR (version 2.7.1a)[114]. Gene
616 expression matrix was generated using rsem-calculate-expression in RSEM
617 (version 1.3.1) with parameters "--keep-intermediate-files --time --star --

618 append-names --output-genome-bam --sort-bam-by-coordinate".

619 Protein structure prediction and alignment

620 Protein structures were predicted using ColabFold
621 (https://colab.research.google.com/github/sokrypton/ColabFold)[115], and
622 different protein structures were aligned using "Matchmaker" in ChimeraX
623 (version 1.9)[116].

624 Data availability

625 All data used and generated in this study are available at the Open Archive for
626 Miscellaneous Data (OMIX: OMIX012585)[117] in the National Genomics Data
627 Center (NGDC), China National Center for Bioinformation / Beijing Institute of
628 Genomics, Chinese Academy of Sciences[118].

629 Code availability

630 The computational codes used in this study have been deposited in GitHub
631 (https://github.com/Bioinfo-Ming/PlantEditosomeEvolution).

632 ACKNOWLEDGEMENTS

633 We thank Shuhui Song, Lina Ma, Chao Zhang, Tianyi Xu, Tongtong Zhu,
634 Dechang Yang, Shuangyang Wu, Lin Liu, Zhao Li, Yang Zhang, Xing Zheng,
635 Yue Qi, Xinyu Zhou, Wenzhuo Cheng, Yuxin Qin, Miaomiao Wang, Shiting
636 Wang, Lingjie Wang, Zihan Wang, Kehua Ma, Pan Li, Yiran Zhan, and Zheng
637 Luo for their valuable discussions and advice. This work was supported by
638 grants from National Natural Science Foundation of China [T2425005 and
639 32030021] and International Partnership Program of the Chinese Academy of
640 Sciences [153F11KYSB20160008].

CONTRIBUTOR INFORMATION

M.C. collected and analyzed the data, and wrote the manuscript. Z.Z. conceived, initiated and supervised this study. Q.M.Q. and Z.Z. revised the manuscript.

FUNDING

The authors declare no competing interests.

REFERENCES

1 Castandet, B. & Araya, A. RNA editing in plant organelles. Why make it easy? *Biochemistry (Moscow)* **76**, 924-931 (2011).

2 Hao, W. *et al.* RNA editing and its roles in plant organelles. *Frontiers in Genetics* **12**, 757109 (2021).

3 Takenaka, M., Zehrmann, A., Verbitskiy, D., Haertel, B. & Brennicke, A. RNA editing in plants and its evolution. *Annual Review of Genetics* **47**, 335-352 (2013).

4 Fujii, S. & Small, I. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytologist* **191**, 37-47 (2011).

5 Hu, R. *et al.* Adaptive evolution of the enigmatic Takakia now facing climate change in Tibet. *Cell* **186**, 3558-3576. e3517 (2023).

6 Small, I. D., Schallenberg‐Rüdinger, M., Takenaka, M., Mireau, H. & Ostersetzer‐Biran, O. Plant organellar RNA editing: what 30 years of research has revealed. *The Plant Journal* **101**, 1040-1056 (2020).

7 Wang, Y. *et al.* Empty Pericarp21 encodes a novel PPR-DYW protein that is required for mitochondrial RNA editing at multiple sites, complexes I and V biogenesis, and seed development in maize. *PLoS Genetics* **15**, e1008305 (2019).

665 8 Sun, Y. K., Gutmann, B., Yap, A., Kindgren, P. & Small, I. Editing of
666 chloroplast rps14 by PPR editing factor EMB2261 is essential for
667 Arabidopsis development. *Frontiers in Plant Science* **9**, 841 (2018).
668 9 Li, J. *et al.* The Pentatricopeptide Repeat Protein OsPPR674 Regulates
669 Rice Growth and Drought Sensitivity by Modulating RNA Editing of the
670 Mitochondrial Transcript ccmC. *International Journal of Molecular Sciences*
671 **26**, 2646 (2025).
672 10 Shi, X., Germain, A., Hanson, M. R. & Bentolila, S. RNA recognition motif-
673 containing protein ORRM4 broadly affects mitochondrial RNA editing and
674 impacts plant development and flowering. *Plant Physiology* **170**, 294-309
675 (2016).
676 11 Yang, Y. *et al.* The RNA editing factor SlORRM4 is required for normal fruit
677 ripening in tomato. *Plant Physiology* **175**, 1690-1702 (2017).
678 12 Luo, Z. *et al.* Pentatricopeptide repeat gene-mediated mitochondrial RNA
679 editing impacts on rice drought tolerance. *Frontiers in Plant Science* **13**,
680 926285 (2022).
681 13 Yuan, H. & Liu, D. Functional disruption of the pentatricopeptide protein
682 SLG1 affects mitochondrial RNA editing, plant development, and
683 responses to abiotic stresses in Arabidopsis. *The Plant Journal* **70**, 432-
684 444 (2012).
685 14 Sun, T., Bentolila, S. & Hanson, M. R. The unexpected diversity of plant
686 organelle RNA editosomes. *Trends in Plant Science* **21**, 962-973 (2016).
687 15 Wang, T. & Takenaka, M. The molecular basis and evolution of the
688 organellar RNA editosome by complementary DYW deaminases in seed
689 plants. *Plant Physiology* **197**, kiaf142 (2025).
690 16 Knoop, V. C-to-U and U-to-C: RNA editing in plant organelles and beyond.
691 *Journal of Experimental Botany* **74**, 2273-2294 (2023).
692 17 Yan, J., Zhang, Q. & Yin, P. RNA editing machinery in plant organelles.

693 *Science China Life Sciences* **61**, 162-169 (2018).
694 18 Ichinose, M. & Sugita, M. RNA editing and its molecular mechanism in plant
695 organelles. *Genes* **8**, 5 (2016).
696 19 Wang, Y. & Tan, B.-C. Pentatricopeptide repeat proteins in plants: Cellular
697 functions, action mechanisms, and potential applications. *Plant*
698 *Communications* **6** (2025).
699 20 Yan, J. *et al.* MORF9 increases the RNA-binding activity of PLS-type
700 pentatricopeptide repeat protein in plastid RNA editing. *Nature Plants* **3**, 1-
701 8 (2017).
702 21 Wang, Y. *et al.* Maize PPR-E proteins mediate RNA C-to-U editing in
703 mitochondria by recruiting the trans deaminase PCW1. *The Plant Cell* **35**,
704 529-551 (2023).
705 22 Yang, Y.-Z. *et al.* GRP23 plays a core role in E-type editosomes via
706 interacting with MORFs and atypical PPR-DYWs in Arabidopsis
707 mitochondria. *Proceedings of the National Academy of Sciences* **119**,
708 e2210978119 (2022).
709 23 Shi, X., Bentolila, S. & Hanson, M. R. Organelle RNA recognition motif-
710 containing (ORRM) proteins are plastid and mitochondrial editing factors in
711 Arabidopsis. *Plant Signaling & Behavior* **11**, 294-309 (2016).
712 24 Sun, T. *et al.* An RNA recognition motif-containing protein is required for
713 plastid RNA editing in Arabidopsis and maize. *Proceedings of the National*
714 *Academy of Sciences* **110**, E1169-E1178 (2013).
715 25 Sun, T. *et al.* A zinc finger motif-containing protein is essential for
716 chloroplast RNA editing. *PLoS Genetics* **11**, e1005028 (2015).
717 26 Gipson, A. B., Hanson, M. R. & Bentolila, S. The RanBP2 zinc finger
718 domains of chloroplast RNA editing factor OZ1 are required for protein–
719 protein interactions and conversion of C to U. *The Plant Journal* **109**, 215-
720 226 (2022).

721 27 Wang, Z. *et al.* OsOZ1, a zinc finger motif-containing protein, is essential
722 for chloroplast development through RNA editing and intron splicing in rice.
723 *Plant Growth Regulation* **105**, 449-461 (2025).
724 28 Haag, S. *et al.* Crystal structures of the Arabidopsis thaliana organellar RNA
725 editing factors MORF1 and MORF9. *Nucleic Acids Research* **45**, 4915-
726 4928 (2017).
727 29 Li, J., Yuan, J., Jing, Y. & Lin, R. MORF proteins: A small family regulating
728 organellar RNA editing and beyond. *Journal of Integrative Plant Biology*
729 (2025).
730 30 Takenaka, M. *et al.* Multiple organellar RNA editing factor (MORF) family
731 proteins are required for RNA editing in mitochondria and plastids of plants.
732 *Proceedings of the National Academy of Sciences* **109**, 5104-5109 (2012).
733 31 Sugita, M., Uchiyama, H. & Ichinose, M. RRM and PPR proteins network
734 as a global regulator for plastid RNA metabolism. *Endocytobiosis Cell Res.*
735 **27**, 41-44 (2016).
736 32 Baudry, K., Monachello, D., Castandet, B., Majeran, W. & Lurin, C.
737 Dissecting the molecular puzzle of the editosome core in Arabidopsis
738 organelles. *Plant Science* **344**, 112101 (2024).
739 33 Shikanai, T. RNA editing in plant organelles: machinery, physiological
740 function and evolution. *Cellular and Molecular Life Sciences CMLS* **63**, 698-
741 708 (2006).
742 34 Gutmann, B. *et al.* The expansion and diversification of pentatricopeptide
743 repeat RNA-editing factors in plants. *Molecular plant* **13**, 215-230 (2020).
744 35 Manna, S. An overview of pentatricopeptide repeat proteins and their
745 applications. *Biochimie* **113**, 93-99 (2015).
746 36 Manna, S. & Barth, C. Identification of a novel pentatricopeptide repeat
747 subfamily with a C-terminal domain of bacterial origin acquired via ancient
748 horizontal gene transfer. *BMC Research Notes* **6**, 525 (2013).

749 37 Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J. M. & Knoop,
750 V. A survey of PPR proteins identifies DYW domains like those of land plant
751 RNA editing factors in diverse eukaryotes. *RNA Biology* **10**, 1549-1556
752 (2013).
753 38 Knoop, V. & Rüdinger, M. DYW-type PPR proteins in a heterolobosean
754 protist: plant RNA editing factors involved in an ancient horizontal gene
755 transfer? *FEBS Letters* **584**, 4287-4291 (2010).
756 39 Schaap, P. *et al.* The Physarum polycephalum genome reveals extensive
757 use of prokaryotic two-component and metazoan-type tyrosine kinase
758 signaling. *Genome Biology and Evolution* **8**, 109-125 (2016).
759 40 Manna, S., Brewster, J. & Barth, C. Identification of pentatricopeptide
760 repeat proteins in the model organism Dictyostelium discoideum.
761 *International Journal of Genomics* **2013**, 586498 (2013).
762 41 Fu, C.-J., Sheikh, S., Miao, W., Andersson, S. G. & Baldauf, S. L. Missing
763 genes, multiple ORFs, and C-to-U type RNA editing in Acrasis kona
764 (Heterolobosea, Excavata) mitochondrial DNA. *Genome Biology and*
765 *Evolution* **6**, 2240-2257 (2014).
766 42 Puttick, M. N. *et al.* The interrelationships of land plants and the nature of
767 the ancestral embryophyte. *Current Biology* **28**, 733-745. e732 (2018).
768 43 Wu, W. *et al.* Genomic convergence in terrestrial root plants through
769 tandem duplication in response to soil microbial pressures. *Cell Reports* **43**
770 (2024).
771 44 Shen, Y. *et al.* Chromosome-level and haplotype-resolved genome
772 provides insight into the tetraploid hybrid origin of patchouli. *Nature*
773 *Communications* **13**, 3511 (2022).
774 45 Xu, Z. *et al.* Crocus genome reveals the evolutionary origin of crocin
775 biosynthesis. *Acta Pharmaceutica Sinica B* **14**, 1878-1891 (2024).
776 46 Wang, Z. *et al.* Near-complete assembly and comprehensive annotation of

777 the wheat Chinese Spring genome. *Molecular Plant* **18**, 892-907 (2025).
778 47 Panchy, N., Lehti-Shiu, M. & Shiu, S.-H. Evolution of gene duplication in
779 plants. *Plant Physiology* **171**, 2294-2316 (2016).
780 48 Freeling, M. Bias in plant gene content following different sorts of
781 duplication: tandem, whole-genome, segmental, or by transposition.
782 *Annual Review of Plant Biology* **60**, 433-453 (2009).
783 49 Wang, X. *et al.* A recent burst of gene duplications in Triticeae. *Plant*
784 *Communications* **3** (2022).
785 50 Feng, X. *et al.* Genomes of multicellular algal sisters to land plants
786 illuminate signaling network evolution. *Nature Genetics* **56**, 1018-1031
787 (2024).
788 51 Zhang, J. *et al.* (2020).
789 52 Li, F.-W. *et al.* Anthoceros genomes illuminate the origin of land plants and
790 the unique biology of hornworts. *Nature Plants* **6**, 259-272 (2020).
791 53 Schafran, P. *et al.* Pan-phylum genomes of hornworts reveal conserved
792 autosomes but dynamic accessory and sex chromosomes. *Nature Plants*
793 **11**, 49-62 (2025).
794 54 Li, C. *et al.* Extraordinary preservation of gene collinearity over three
795 hundred million years revealed in homosporous lycophytes. *Proceedings*
796 *of the National Academy of Sciences* **121**, e2312607121 (2024).
797 55 Cui, J. *et al.* Chromosome-level reference genome of tetraploid Isoetes
798 sinensis provides insights into evolution and adaption of lycophytes.
799 *GigaScience* **12**, giad079 (2023).
800 56 Zhang, C., Gschwend, A. R., Ouyang, Y. & Long, M. Evolution of gene
801 structural complexity: an alternative-splicing-based model accounts for
802 intron-containing retrogenes. *Plant Physiology* **165**, 412-423 (2014).
803 57 Schrader, L. & Schmitz, J. The impact of transposable elements in adaptive
804 evolution. *Molecular Ecology* **28**, 1537-1549 (2019).

805 58 Takenaka, M. *et al.* DYW domain structures imply an unusual regulation
806 principle in plant organellar RNA editing catalysis. *Nature catalysis* **4**, 510-
807 522 (2021).
808 59 Boussardon, C. *et al.* The cytidine deaminase signature HxE(x)nCxxC of
809 DYW1 binds zinc and is necessary for RNA editing of ndhD‐1. *New*
810 *Phytologist* **203**, 1090-1095 (2014).
811 60 Yang, Y. *et al.* DYW cytidine deaminase domains have a long‐range impact
812 on RNA recognition by the PPR array of chimeric plant C‐to‐U RNA editing
813 factors and strongly affect target selection. *The Plant Journal* **116**, 840-854
814 (2023).
815 61 Hayes, M. L., Giang, K., Berhane, B. & Mulligan, R. M. Identification of two
816 pentatricopeptide repeat genes required for RNA editing and zinc binding
817 by C-terminal cytidine deaminase-like domains. *Journal of Biological*
818 *Chemistry* **288**, 36519-36529 (2013).
819 62 Bolhuis, P. Sampling kinetic protein folding pathways using all-atom models.
820 *Lecture Notes in Physics*, 703, 393-433 (2006).
821 63 Hespeels, B. *et al.* Back to the roots, desiccation and radiation resistances
822 are ancestral characters in bdelloid rotifers. *BMC Biology* **21**, 72 (2023).
823 64 Wang, W., Yang, Y., Örstan, A., He, Z. & Wang, Q. Species diversity of
824 bdelloid rotifer (Rotifera, Bdelloidea) in different areas in China, with a
825 description of two new species. *ZooKeys* **1229**, 1 (2025).
826 65 Jain, S. C., Shinde, U., Li, Y., Inouye, M. & Berman, H. M. The crystal
827 structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex
828 at 2.0 Å resolution. *Journal of Molecular Biology* **284**, 137-144 (1998).
829 66 Li, Y., Hu, Z., Jordan, F. & Inouye, M. Functional Analysis of the Propeptide
830 of Subtilisin E as an Intramolecular Chaperone for Protein Folding:
831 Refolding And Inhibitory Abilities of Propeptide Mutants (∗). *Journal of*

832 *Biological Chemistry* **270**, 25127-25132 (1995).
833 67 Wang, D. *et al.* Genome-wide analysis of multiple organellar RNA editing
834 factor family in poplar reveals evolution and roles in drought stress.
835 *International Journal of Molecular Sciences* **20**, 1425 (2019).
836 68 Liu, P.-Z. *et al.* Genome-wide identification and expression analysis of the
837 MORF gene family in celery reveals their potential role in chloroplast
838 development. *Journal of Genetic Engineering and Biotechnology* **22**,
839 100443 (2024).
840 69 Searing, A. M., Satyanarayan, M. B., O′ Donnell, J. P. & Lu, Y. Two
841 organelle RNA recognition motif proteins affect distinct sets of RNA editing
842 sites in the Arabidopsis thaliana plastid. *Plant Direct* **4**, e00213 (2020).
843 70 Shi, X., Hanson, M. R. & Bentolila, S. Two RNA recognition motif-containing
844 proteins are plant mitochondrial editing factors. *Nucleic Acids Research* **43**,
845 3814-3825 (2015).
846 71 Liang, Z. & Schnable, J. C. Functional divergence between subgenomes
847 and gene pairs after whole genome duplications. *Molecular Plant* **11**, 388-
848 397 (2018).
849 72 Rüdinger, M., Fritz-Laylin, L., Polsakiewicz, M. & Knoop, V. Plant-type
850 mitochondrial RNA editing in the protist Naegleria gruberi. *RNA* **17**, 2058-
851 2062 (2011).
852 73 Yang, J., Harding, T., Kamikawa, R., Simpson, A. G. & Roger, A. J.
853 Mitochondrial genome evolution and a novel RNA editing system in deep-
854 branching heteroloboseids. *Genome Biology and Evolution* **9**, 1161-1174
855 (2017).
856 74 He, P., Huang, S., Xiao, G., Zhang, Y. & Yu, J. Abundant RNA editing sites
857 of chloroplast protein-coding genes in Ginkgo biloba and an evolutionary
858 pattern analysis. *BMC Plant Biology* **16**, 257 (2016).
859 75 Xhemalçe, B., Miller, K. M. & Gromak, N. Epitranscriptome in action: RNA

860 modifications in the DNA damage response. *Molecular Cell* **84**, 3610-3626
861 (2024).
862 76 Enright, A. J., Iliopoulos, I., Kyrpides, N. C. & Ouzounis, C. A. Protein
863 interaction maps for complete genomes based on gene fusion events.
864 *Nature* **402**, 86-90 (1999).
865 77 Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg,
866 D. A combined algorithm for genome-wide prediction of protein function.
867 *Nature* **402**, 83-86 (1999).
868 78 Marcotte, E. M. *et al.* Detecting protein function and protein-protein
869 interactions from genome sequences. *Science* **285**, 751-753 (1999).
870 79 Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M.
871 BUSCO update: novel and streamlined workflows along with broader and
872 deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and
873 viral genomes. *Molecular Biology and Evolution* **38**, 4647-4654 (2021).
874 80 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for
875 comparative genomics. *Genome Biology* **20**, 1-14 (2019).
876 81 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software
877 version 7: improvements in performance and usability. *Molecular Biology*
878 *and Evolution* **30**, 772-780 (2013).
879 82 Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with
880 Entropy): a new software for selection of phylogenetic informative regions
881 from multiple sequence alignments. *BMC Evolutionary Biology* **10**, 1-21
882 (2010).
883 83 Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for
884 phylogenetic inference in the genomic era. *Molecular Biology and Evolution*
885 **37**, 1530-1534 (2020).
886 84 Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree
887 estimation. *Bioinformatics* **30**, i541-i548 (2014).

85 Eddy, S. Sequence analysis using profile hidden Markov models. http://hmmer. *wustl. edu* (2004).

86 Meng, L. *et al.* PPR proteins in plants: Roles, mechanisms, and prospects for rice research. *Frontiers in Plant Science* **15**, 1416742 (2024).

87 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419 (2021).

88 Ødum, M. T. *et al.* DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Research* **52**, W215-W220 (2024).

89 Armenteros, J. J. A. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* **2** (2019).

90 Sahu, S. S., Loaiza, C. D. & Kaundal, R. Plant-mSubP: a computational framework for the prediction of single-and multi-target protein subcellular localization using integrated machine-learning approaches. *AoB Plants* **12**, plz068 (2020).

91 Sperschneider, J. *et al.* LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports* **7**, 44598 (2017).

92 Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research* **53**, D444-D456 (2025).

93 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 1-9 (2009).

94 Sun, P. *et al.* WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant* **15**, 1841-1851 (2022).

95 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**, e49-e49 (2012).

916 96 Zhang, Z. KaKs_Calculator 3.0: calculating selective pressure on coding
917 and non-coding sequences. *Genomics, Proteomics and Bioinformatics* **20**,
918 536-540 (2022).
919 97 Zhang, Z. *et al.* ParaAT: a parallel tool for constructing multiple protein-
920 coding DNA alignments. *Biochemical and Biophysical Research*
921 *Communications* **419**, 779-781 (2012).
922 98 Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of
923 transposable element families. *Proceedings of the National Academy of*
924 *Sciences* **117**, 9451-9457 (2020).
925 99 Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-
926 length LTR retrotransposons. *Nucleic Acids Research* **35**, W265-W268
927 (2007).
928 100 Ou, S. & Jiang, N. LTR_FINDER_parallel: parallelization of LTR_FINDER
929 enabling rapid identification of long terminal repeat retrotransposons.
930 *Mobile DNA* **10**, 48 (2019).
931 101 Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible
932 software for de novo detection of LTR retrotransposons. *BMC*
933 *Bioinformatics* **9**, 1-14 (2008).
934 102 Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program
935 for identification of long terminal repeat retrotransposons. *Plant Physiology*
936 **176**, 1410-1422 (2018).
937 103 Chen, N. Using RepeatMasker to identify repetitive elements in genomic
938 sequences. *Current Protocols in Bioinformatics* **5**, 4.10. 11-14.10. 14 (2004).
939 104 Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam
940 community resource of transposable element families, sequence models,
941 and genome annotations. *Mobile DNA* **12**, 2 (2021).
942 105 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for
943 automated alignment trimming in large-scale phylogenetic analyses.

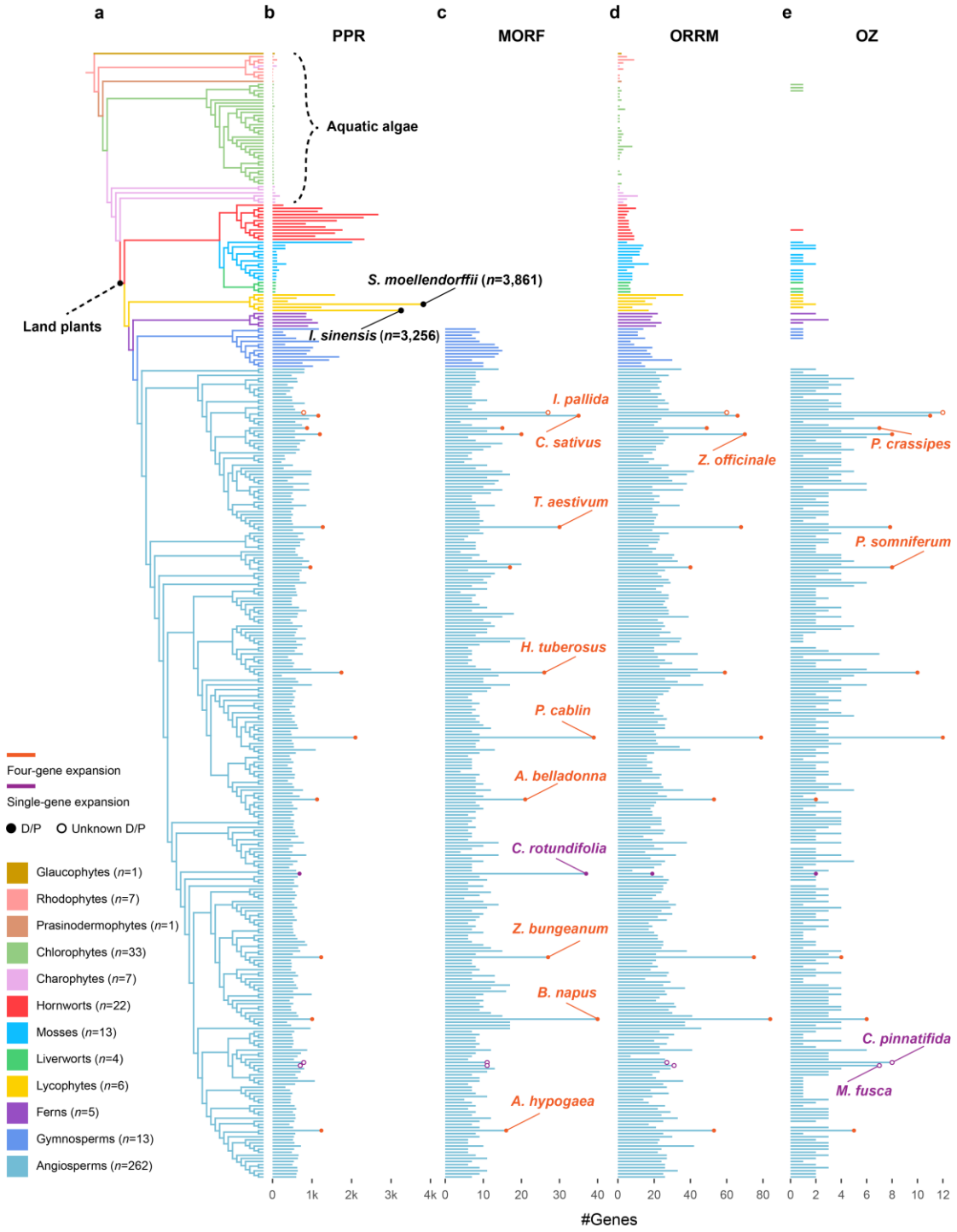
944 *Bioinformatics* **25**, 1972-1973 (2009).
945 106 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to
946 the phylogenetic tree display and annotation tool. *Nucleic Acids Research*
947 **52**, W78-W82 (2024).
948 107 Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary
949 genetics analysis version 11. *Molecular Biology and Evolution* **38**, 3022-
950 3027 (2021).
951 108 Procter, J. B. *et al.* Alignment of biological sequences with Jalview. *Methods*
952 *in Molecular Biology* **2231**, 203-224 (2021).
953 109 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately
954 maximum-likelihood trees for large alignments. *PloS One* **5**, e9490 (2010).
955 110 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based
956 genome alignment and genotyping with HISAT2 and HISAT-genotype.
957 *Nature Biotechnology* **37**, 907-915 (2019).
958 111 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-
959 Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1-
960 16 (2011).
961 112 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ
962 preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
963 113 Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq
964 experiments. *Bioinformatics* **28**, 2184-2185 (2012).
965 114 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*
966 **29**, 15-21 (2013).
967 115 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature*
968 *Methods* **19**, 679-682 (2022).
969 116Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis.
970 *Protein Science* **32**, e4792 (2023).
971 117 Zhang, S. *et al.* The GSA Family in 2025: A Broadened Sharing Platform

for Multi-Omics and Multimodal Data. *Genomics, Proteomics & Bioinformatics*, qzaf072 (2025).

118 CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Research* **53**, D30-D44 (2025).

977 # Figures

978 ## Fig. 1



979

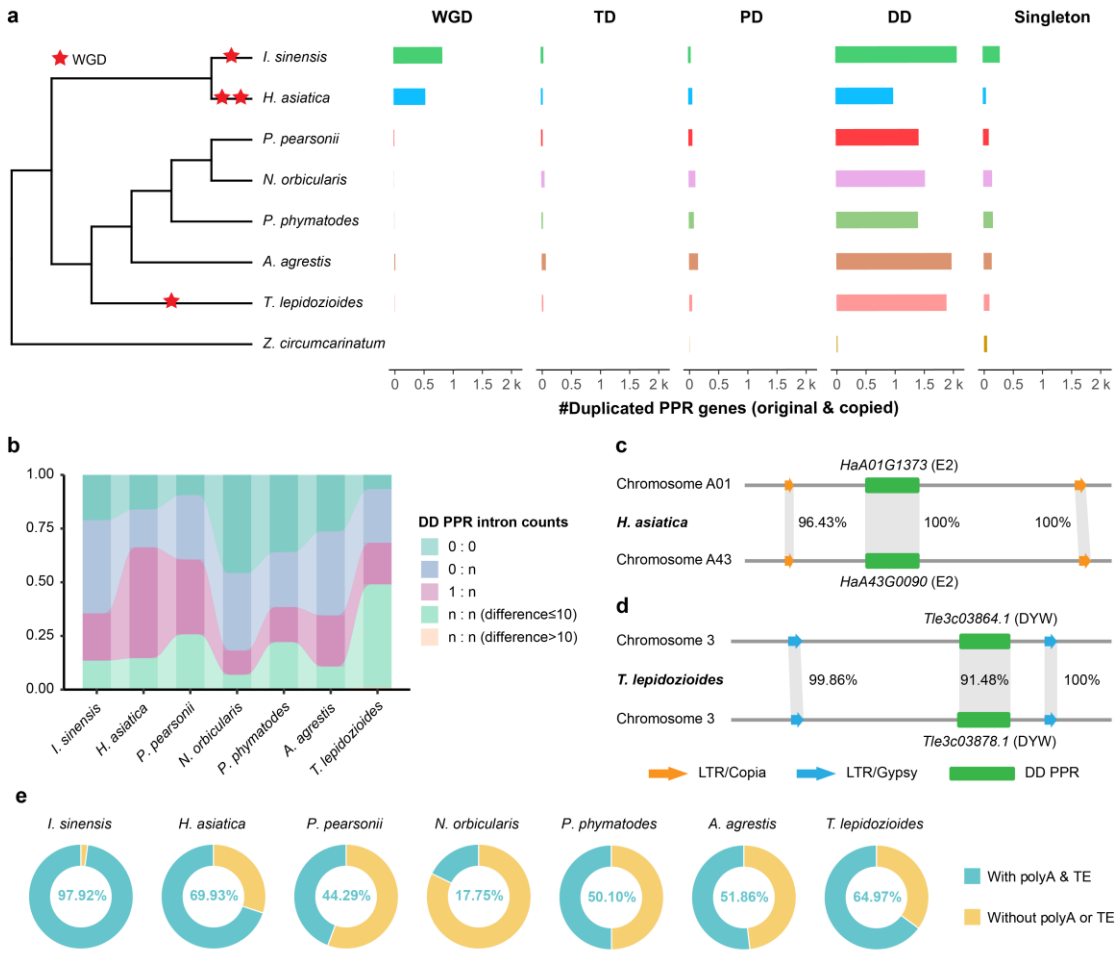
980 **Fig. 1: The evolutionary landscape of PPR, MORF, ORRM, and OZ genes**

981 **across Archaeplastida. (a)** The maximum likelihood phylogenomic tree of 364

982 Archaeplastida species, including glaucophytes (*n*=1), rhodophytes (*n*=7),

983 prasinodermophytes (*n*=1), chlorophytes (*n*=33), charophytes (*n*=7), hornworts
984 (*n*=22), mosses (*n*=13), liverworts (*n*=4), lycophytes (*n*=6), ferns (*n*=5),
985 gymnosperms (*n*=13), and angiosperms (*n*=262). The best-fit model of each
986 single-copy orthologous gene tree is automatically selected according to
987 Bayesian Information Criterion in IQ-TREE. All gene trees (*n*=138) are merged
988 and the final species tree is inferred by ASTRAL. Bootstraps are from 1,000
989 replicates. The distribution of **(b)** PPR, **(c)** MORF, **(d)** ORRM, and **(e)** OZ gene
990 across Archaeplastida lineages. Two lycophytes, *S. moellendorffii* and *I.*
991 *sinensis*, have the highest number of PPR genes (*n*=3,861 and 3,256,
992 respectively). Orange and purple bars indicate gene expansions occurring
993 simultaneously across all four gene families, and within individual families,
994 respectively. Filled and open black circles denote species that have undergone
995 genome diploidization/polyploidization (D/P), and those with unknown genome
996 expansion, respectively. D/P: *Arachis hypogaea*, *Atropa bell adonna*, *B. napus*,
997 *C. rotundifolia*, *C. sativus*, *Helianthus tuberosus*, *Papaver somniferum*, *P.*
998 *cablin*, *Pontederia crassipes*, *T. aestivum*, *Zanthoxylum bungeanum*, and
999 *Zingiber officinale*. Unknown D/P: *C. pinnatifida*, *Iris pallida*, and *M. fusca*.

1000 **Fig. 2**



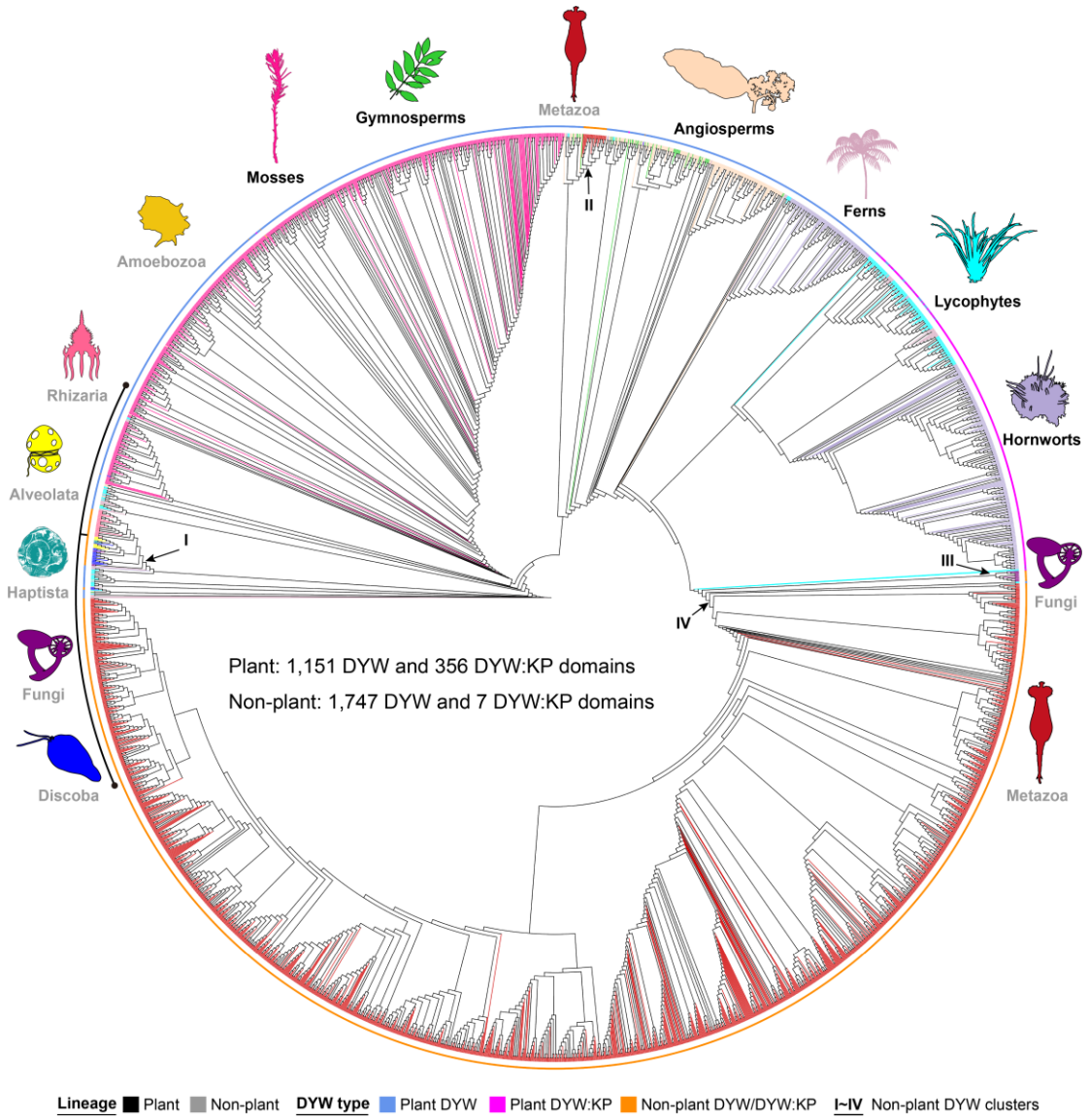
1001

1002 **Fig. 2: Mechanisms of PPR gene expansions. (a)** Numbers of duplicated
1003 PPR genes (original and copied) derived from WGD, TD, PD, and DD in seven
1004 representative species (*A. agrestis*, *N. orbicularis*, *P. pearsonii*, *P. phymatodes*,
1005 *T. lepidozioides*, *I. sinensis*, and *H. asiatica*) and the outgroup (*Z.*
1006 *circumcarinatum*). A red five-pointed star denotes a whole genome duplication
1007 event. WGD: Whole Genome Duplication, TD: Tandem Duplication, PD:
1008 Proximal Duplication, DD: Dispersed Duplication. **(b)** Percentage of intronless
1009 DD PPR gene pairs. Numbers to the left and right of the colon indicate intron
1010 counts in each DD PPR gene pair. Categories “0:0”, “0:n”, “1:n”, and “n:n
1011 (difference > 10)” denote intron loss in DD PPR gene pairs. **(c)** E2 and **(d)** DYW
1012 subgroup PPR gene pairs (DD) flanked by similar LTR/Copia and LTR/Gypsy,
1013 respectively. **(e)** Percentages of intronless DD PPR gene pairs with polyA tails

1014 and flanked by LTRs.

1015

1016 **Fig. 3**



1017

1018 **Fig. 3: The maximum likelihood phylogenetic tree of DYW and DYW:KP**
1019 **domains in plants and non-plants.** Three predominant clusters (I, II, and III)
1020 of non-plant DYW domains are embedded within plants. Among non-plant
1021 lineages, Metazoa contain the largest number of DYW domains. Best-fit model
1022 (IQ-TREE): JTT + R10. Bootstraps are from 1,000 replicates.

Fig. 4

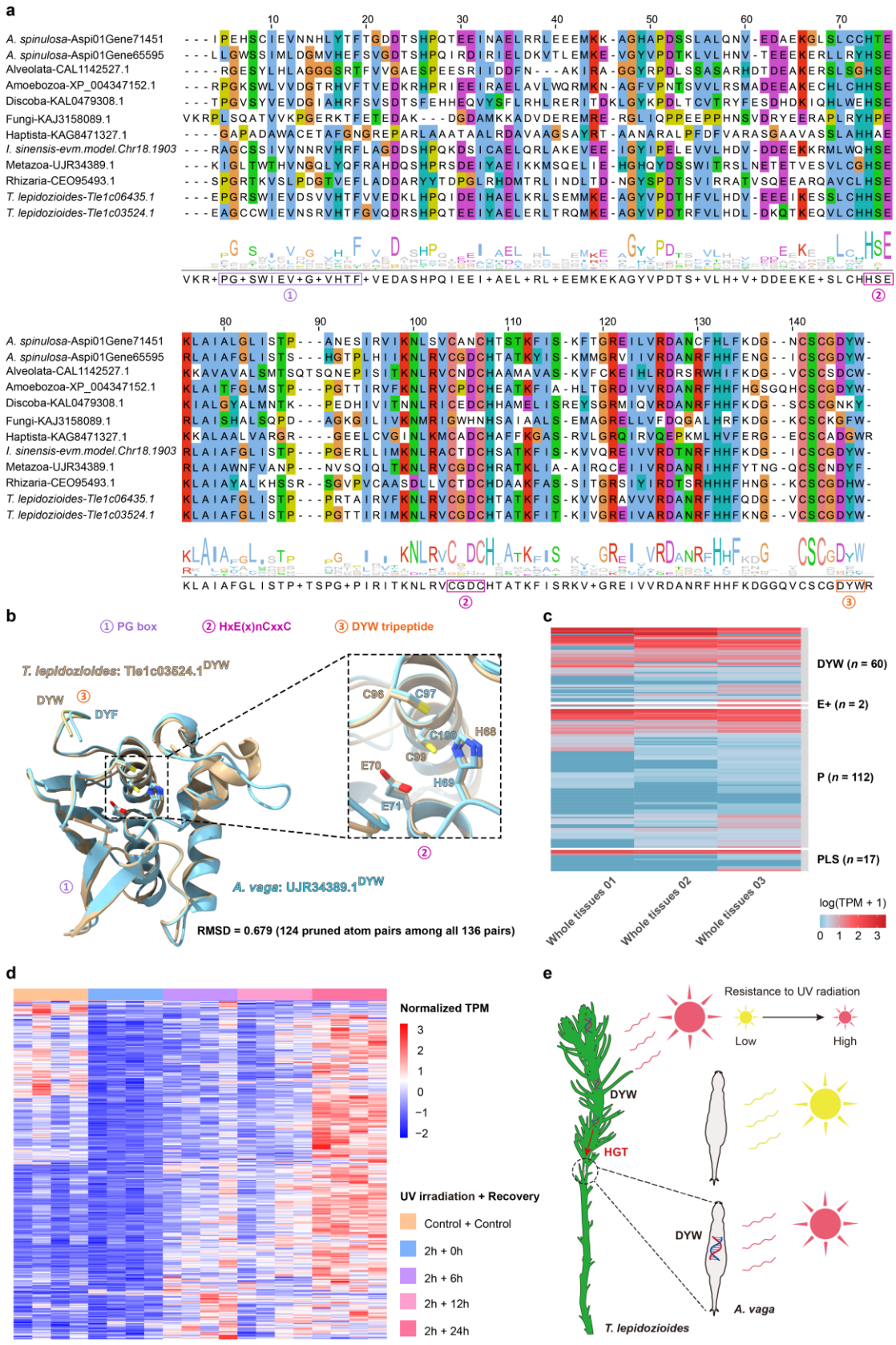
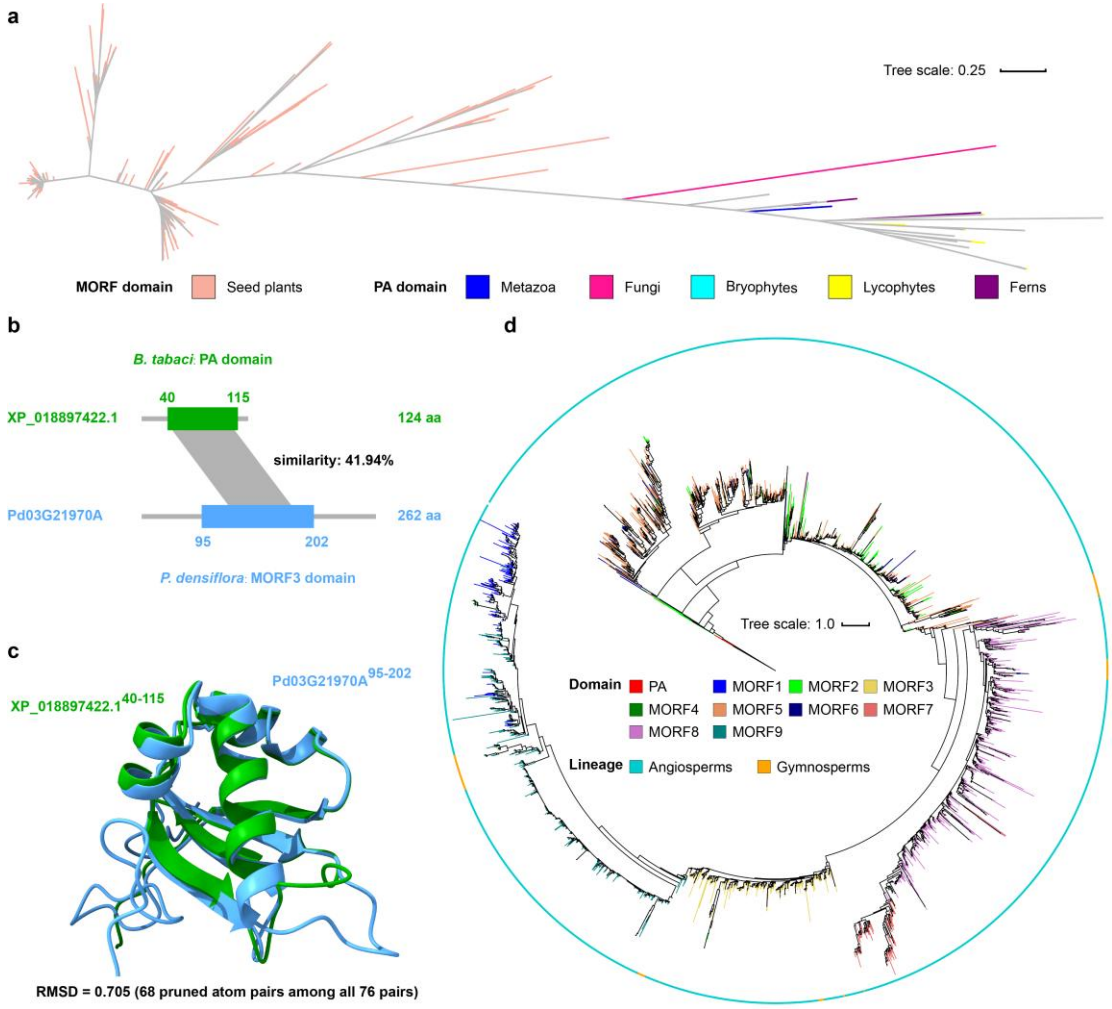


Fig. 4: HGT of DYW domains from plants to bdelloid rotifers. (a) Sequence

1026 alignments between plant and non-plant DYW domains. Amino acid residues in
1027 rectangles denote PG box, HxE(x)nCxxC, and DYW tripeptide. **(b)** Protein
1028 structure alignment between DYW domain (protein ID: Tle1c03524.1) in the
1029 moss *T. lepidozioides* and its counterpart (protein ID: UJR34389.1) in the
1030 bdelloid rotifer *A. vage*. **(c)** Expression of PPR genes (DYW, E+, PLS, and P
1031 subgroups) in whole tissues of *A. vage*. TPM: Transcripts Per Million. **(d)**
1032 Expression of DYW subgroup PPR genes in *T. lepidozioides* under control
1033 condition and after 2-hour UV irradiation, followed by 0, 6, 12, and 24-hour
1034 recovery. **(e)** A proposed model of horizontal transfer of DYW subgroup PPR
1035 genes from mosses (e.g., *T. lepidozioides*) to bdelloid rotifers (e.g., *A. vage*).
1036

1037 **Fig. 5**



1038

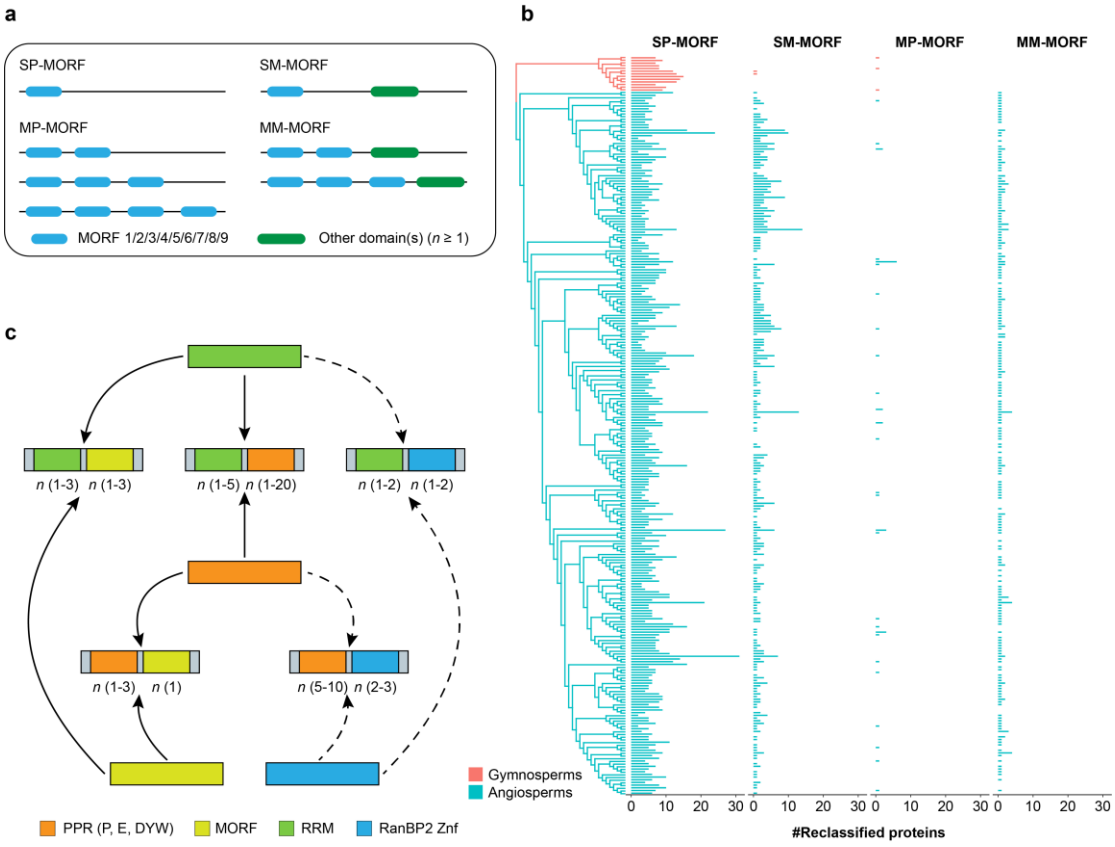
1039 **Fig. 5: Conservation and diversity of MORF domains. (a)** The maximum-
1040 likelihood phylogenetic tree of MORF and PA domains. The tree scale bar
1041 represents the number of substitutions per site. Model (FastTree): Jones-
1042 Taylor-Thorton, CAT approximation with 20 rate categories. **(b)** Sequence and
1043 **(c)** structure alignments between MORF3 domain (protein ID: Pd03G21970A)
1044 from *Pinus densiflora* and its homologous PA domain (protein ID:
1045 XP_018897422.1) from *Bemisia tabaci*. **(d)** The maximum-likelihood
1046 phylogenetic tree of MORF1~MORF9 domains from angiosperms and
1047 gymnosperms. Five non-seed plant PA domains with highest sequence
1048 similarity to MORF domains are used as the outgroup (protein IDs:
1049 Dicom.22G070000.1.p, HaB36G042, Phcar.S3G189600.t1,

1050 Phsp.C2G192800.t1, and XP_024533311.1). Best-fit model (IQ-TREE): JTT +

1051 R8. Bootstraps are from 1,000 replicates.

1052

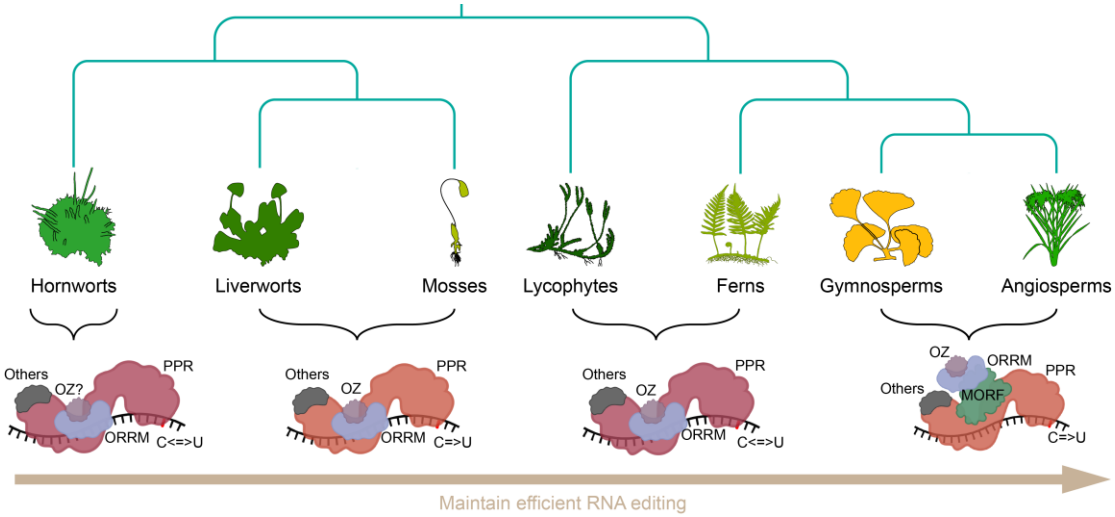
1053 **Fig. 6**



1054

1055 **Fig. 6: Reclassification and diversity of PPR accessory proteins. (a)**
1056 Reclassification of MORF proteins based on domain composition. Blue and
1057 green bars represent MORF1~MORF9 and other domains, respectively. SP-
1058 MORF: Single-Pure MORF, SM-MORF: Single-Mixed MORF; MP-MORF:
1059 Multiple-Pure MORF; MM-MORF: Multiple-Mixed MORF. **(b)** The phylogenetic
1060 tree of reclassified MORF proteins from angiosperms and gymnosperms. **(c)**
1061 Fusions among PPR, MORF, RRM, and RanBP2 Znf domains. Numbers in
1062 parentheses indicate domain counts. Proteins marked with solid and dashed
1063 arrows are localized to mitochondria/plastids and other subcellular
1064 compartments, respectively.
1065

1066 **Extended Data Fig. 1**



1067

1068 **Extended Data Fig. 1: A proposed model for the evolutionary**
1069 **diversification of plant RNA editing factors.** Editing factor composition varies
1070 from hornworts to angiosperms, potentially maintaining efficient RNA editing.
1071 The question mark indicates that OZ protein is rare in hornworts. Deep-rose
1072 PPR: DYW or DYW:KP subgroup PPR protein; red-orange PPR: DYW
1073 subgroup PPR protein. C=>U: C-to-U RNA editing; C<=>U: C-to-U or U-to-C
1074 RNA editing.