Functional characterization of RNase H1 proteins in *Arabidopsis thaliana*

Jan Kuciński[1], Aleksandra Kmera[2,3], M. Jordan Rowley[1,4], Pragya Khurana[1], Marcin Nowotny[2], Andrzej T. Wierzbicki[1,2*]

[1]University of Michigan, Department of Molecular, Cellular and Developmental Biology, Ann Arbor, MI, USA
[2]Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, Warsaw, Poland

[3]Current address: Department of Protein Biosynthesis, Institute of Biochemistry and Biophysics, Warsaw, Poland
[4]Current address: Department of Biology, Emory University, Atlanta, GA, USA

*Corrersponding author, wierzbic@umich.edu

ABSTRACT

RNase H1 is an endonuclease specific towards RNA:DNA hybrids. Members of this protein family are present in most living organisms and are essential for removing RNA that base pairs with DNA. It prevents detrimental effects of RNA:DNA hybrids and is involved in several biological processes. We show that *Arabidopsis thaliana* contains four RNase H1-like proteins originating from two gene duplication events and alternative splicing. These proteins have the canonical RNase H1 activity, which requires at least four ribonucleotides for activity. Two of those proteins are nuclear, one is localized to mitochondria and one to plastids. While the nuclear RNases H1 are dispensable, the presence of at least one organellar RNase H1 is required for embryonic development. The plastid protein RNH1C affects plastid DNA copy number and sensitivity to hydroxyurea. This indicates that three genomes present in each plant cell are served by at least one specialized RNase H1 protein.

INTRODUCTION

Double stranded nucleic acids, which contain deoxyribonucleotides on one strand and one or more ribonucleotides on the other strand are known as RNA:DNA hybrids. They are common byproducts of replication, transcription and other processes. Ribonucleotides within RNA:DNA hybrids are specifically removed by a class of endonucleases known as RNases H[1]. RNases H2 are multi subunit complexes capable of removing even individual ribonucleotides incorporated in double stranded DNA and have been studied in various eukaryotes, including plants[2,3]. RNases H1 act as monomers[4] and require at least four ribonucleotide incorporated into double stranded DNA to bind the substrate[5].

Among the substrates of RNases H1 are R-loops[6] (RNA:DNA hybrid and a displaced ssDNA strand) which are often formed during transcription and replication[7–9]. These structures have also been implicated in DNA repair[10–12], telomere maintenance[13], IgG class-switch recombination[14] and regulation of gene expression[9,15]. RNase H1 digests ribonucleotides within its substrate in a metal ion-dependent manner[16], leading to single stranded DNA formation. Main feature of the proteins encoded by RNase H1 genes is the presence of the catalytic domain[16,17]. Additionally, some bacterial and most eukaryotic RNase H1 proteins contain a RNA:DNA hybrid binding domain (HBD)[18].

Genes encoding RNase H1 proteins are present in the vast majority of living organisms, including Archea[19,20], Bacteria[21] and all kingdoms of Eukarya[1]. They are not essential in prokaryotes and lower eukaryotes but are required for survival in higher eukaryotes[1]. Eukaryotic genomes usually contain unique RNase H1 genes[1,21], which may

however be subject to alternative splicing[22]. *Arabidopsis thaliana* RNases H1 are encoded by three different genes with different predicted subcellular localizations[23].

Among the three RNase H1 proteins in *Arabidopsis thaliana,* only the chloroplast-localized paralog has been studied so far[23]. It is essential for proper plastid development by maintaining the integrity of chloroplast DNA. It works with its interacting partner, DNA gyrase, to resolve transcription-replication conflicts and prevent DNA damage[23]. The role of the remaining two proteins remains unknown beyond the presumption that they resolve R-loops, which are relatively common in the *Arabidopsis* genome[24].

Here, we characterize all RNase H1 proteins detectable in the *Arabidopsis* genome. We identify two ancient gene duplication events, which led to the formation of RNase H1 proteins targeted to various cellular compartments in monocots and dicots. A more recent duplication and alternative splicing produced four RNase H1 proteins in *Arabidopsis thaliana*, two of which are targeted to the nucleus, one to the chloroplast and one to the mitochondria. The proteins localized to endosymbiotic organelles are required for proper embryonic development. These proteins have the canonical RNase H1 activity and are involved in nucleic acid metabolism.

RESULTS

Origin of angiosperm RNases H1

Plant genomes have been shown to encode multiple RNase H1-proteins with different subcellular localizations[23]. To determine their evolutionary origins, we performed an in-depth phylogenetic analysis of plant RNase H1-like proteins. We first used BLAST to identify all plant proteins, which display sequence similarity to *Arabidopsis thaliana* AtRNH1C[23] and contain both RNase H1 and RNA:DNA hybrid binding (HBD) domains. The list of identified proteins is in the Supplemental File 1. The identified proteins were subject to a simultaneous Bayesian alignment and phylogenetic analysis implemented using BAli-Phy package in order to integrate over the uncertainty in both the phylogeny and alignment[25]. In parallel, we predicted the subcellular localization of each identified protein using the TargetP prediction tool[26]. RNase H1-like proteins from angiosperms grouped into four distinct clades (Fig. 1, Fig. S1). The first group included proteins from monocots with mostly nuclear predicted localization. The second group included proteins from monocots with mostly chloroplast or mitochondrial predicted localization (referred to as organellar). The third group included proteins from dicots with mostly organellar localization. Finally, the fourth group included

proteins from dicots with nuclear localization (Fig 1, Fig. S1, Fig. 2A). Phylogenetic relationships between these proteins indicate that the common ancestor of monocots and dicots had one RNase H1-like protein. Two independent gene duplication events then led to both monocots and dicots acquiring at least two proteins with distinct subcellular localizations.

Origin of four RNases H1 in *Arabidopsis*

Arabidopsis thaliana has been shown to contain three genes encoding RNase H1-like proteins: RNH1A (AT3G01410), RNH1B (AT5G51080) and RNH1C (AT1G24090). Products of these genes localize to the nucleus, mitochondria and chloroplasts, respectively[23]. While the split into nuclear and organellar proteins occurred early in dicot evolution (Fig. 1), the origin of two organellar proteins remains unknown. To determine the relationship between RNH1B and RNH1C we analyzed the phylogeny of RNase H1-like proteins in *Brasicaceae* (Fig. 2B). The phylogenetic tree identified one clade of nuclear and two distinct clades of organellar RNase H1-like proteins (Fig. 2B). This indicates that a second diversification event occurred in the common ancestor of *Brasicaceae*, which led to the formation of two organellar proteins. While in *Arabidopsis thaliana* these proteins are localized to mitochondria and chloroplasts, these specific localizations cannot be conclusively predicted for other *Brasicaceae*.

Multiple subcellular localizations of proteins produced from unique RNase H1 genes in animals are commonly determined by alternative splicing[27]. Araport11 genome annotation[28] suggest a similar mechanism in *Arabidopsis thaliana*, where three splice variants of RNH1B (AT5G51080) have been identified (Fig. 2C). We confirmed these annotations using 5'RACE (Fig. 2D). While one splice variant (AT5G51080.1 and AT5G51080.2) encodes a full-length protein, the second variant (AT5G51080.3) has a truncated mitochondrial presequence region and is predicted to localize to the nucleus (Fig. 2E). We conclude that *Arabidopsis thaliana* contains four RNase H1-like proteins originating from two independent gene duplication events and alternative splicing.

RNase H1-like proteins from *Arabidopsis* have canonical RNase H1 activity

Although *Arabidopsis thaliana* RNase H1-like proteins have extensive sequence similarity with RNase H1 proteins, their exact enzymatic activity remains unknown. RNH1C has been shown to remove binding sites of S9.6 antibody from chloroplast DNA[23]. This however, does not conclusively show RNase H1 activity of this protein. To determine if

RNH1A and RNH1B are indeed RNases H1, we incubated the recombinant proteins with oligonucleotide substrates. Double stranded DNA and RNA was not digested by the recombinant proteins (Fig. 3AB). The same was true for oligonucleotides containing one or two ribonucleotides, which differentiates RNase H1 from RNase H2. Oligonucleotides containing four ribonucleotides were however digested by both RNH1A and RNH1B (Fig. 3AB). We did not test RNH1C because we were unable to produce a soluble recombinant protein. These results indicate that RNH1A and RNH1B have the canonical RNase H1 activity. Because of the extensive sequence similarity, especially within the catalytic domain, RNH1C is likely to have the same enzymatic activity.

RNH1A and RNH1B do not affect development

RNH1C has been shown to be required for proper chloroplast development[23]. To determine the roles of all three RNase H1 encoding genes, we obtained T-DNA mutants in RNH1A (SALK_150285C) and RNH1B (SAIL_1174_C11) as well as the previously published RNH1C (SAIL_97_E11). Single mutants *atrnh1a* and *atrnh1b* did not exhibit any obvious developmental phenotypes (Fig. 4A-C). *atrnh1c* mutant had the expected pale leaf and dwarf phenotype (Fig. 4D). The pale green phenotype of *atrnh1c* single mutant was confirmed by chlorophyll content quantification (Fig. 4I).

To determine if the studied T-DNA mutants may express truncated proteins, we performed RT-PCR with primers upstream or downstream of the T-DNA insertion sites (Fig. 4J). All mutants had strongly and significantly reduced RNA accumulation downstream of T-DNA insertions (Fig. 4KL). This indicates that these mutants are unlikely to produce truncated proteins containing the C-terminal RNase H1 domain. RNA accumulation upstream of T-DNA was reduced in *atrnh1a*, unchanged in *atrnh1c* and strongly increased in at*rnh1b* (Fig. 4KL). This indicates that a truncated N-terminal fragment is unlikely to be produced in *atrnh1a*. Truncated N-terminal fragments may be produced in *atrnh1b* and *atrnh1c*, however *atrnh1b* is not expected to contain full length RNase H1 domain. On the other hand, *atrnh1c* may produce a truncated protein including the RNase H domain (Fig. 4J), which indicates that the C-terminal part of the protein is important for its function. The strong increase of upstream RNA accumulation in *atrnh1b* may indicate the presence of an autoregulatory mechanism within *RNH1B*.

Presence of both RNH1B and RNH1C is required for viability

Because RNases H1 are required for viability in animals[27], we tested the phenotypes of all combinations of RNH1 double mutants. The *atrnh1a*, *atrnh1b* double mutant did not show any visible developmental phenotypes (Fig. 4E) and the phenotype of the *atrnh1a*, *atrnh1c* double mutant was similar to the *atrnh1c* single mutant. Phenotypic similarity of *atrnh1c* single mutant and *atrnh1a*, *atrnh1c* double mutant was confirmed by chlorophyll content quantification (Fig. 4I). Despite several attempts we were unable to obtain the *atrnh1b*, *atrnh1c* double mutant. Among 211 F1 plants obtained from crossing *atrnh1b*$^{-/-}$ with *atrnh1c*$^{+/-}$ we found no viable double homozygous mutant. However, the cross resulted in about 25% of seeds not developing properly (Fig. 5A-C), which likely are the double mutants. These improperly developing seeds contain aborted embryos which stop developing at approximately heart-shape stage of embryonic development (Fig. 5DE). These results suggest that while the nuclear RNH1A is not required for viability, the presence of at least one organellar protein RNH1B or RNH1C is essential.

RNH1C is involved in chloroplast nucleic acid metabolism

RNH1C has been shown to be involved in chloroplast DNA maintenance[23]. To test if RNases H1 expressed from the three *Arabidopsis* genes affect DNA copy number in various cellular compartments we performed real time PCR with primers specific to nuclear, mitochondrial and chloroplast genomes. *atrnh1c* mutant and *atrnh1a, atrnh1c* double mutant contained approximately three times more chloroplast DNA than wild type mutant plants (Fig. 6A-C). This result indicates that RNH1C protein is involved in controlling DNA copy number in chloroplasts, which is consistent with its postulated role in chloroplast genome maintenance[23]. Surprisingly, *atrnh1a*, *atrnh1b* mutants and the *atrnh1a*, *atrnh1b* double mutant did not affect the relative content of the three genomes (Fig. 6A-C). This indicates that nuclear and mitochondrial RNases H1 have different roles than the chloroplast RNase H1. Alternatively, some or all RNase H1 proteins may be targeted to more than one compartment.

The increased amount of chloroplast DNA in *atrnh1c* mutants might be a result of impaired DNA replication. To test this possibility, we treated wild type plants and RNase H1 mutants with replication stress and measured root length. To induce replication stress, we applied hydroxyurea[2,3,29], which is known to inhibit replication by inhibiting rNTP reductase and therefore decreasing the amount of available dNTPs[30]. Treatment with 2mM hydroxyurea caused a decrease of root length in wild type plants to approximately 40% of root length in untreated plants (Fig. 6 E). *rnh1a* and *rnh1b* mutants had no effect on hydroxyurea sensitivity (Fig. 6 E). However, roots of *atrnh1c* mutant and *atrnh1a*, *atrnh1c* double mutant grew to less

than 30% of root length in untreated plants (Fig. 6 E). This indicates that RNH1C affects sensitivity to hydroxyurea and is likely to be involved in DNA replication.

DISCUSSION

RNase H1 proteins are known to digest the RNA component of RNA:DNA hybrids and this substrate must contain at least four sequential ribonucleotides[16]. In contrast, RNase H2 complexes only require a single ribonucleotide incorporated into double-stranded DNA for activity[1,31]. Therefore, RNase H2 has a broader substrate range and is partially redundant with RNase H1[32]. We demonstrate for the first time that *Arabidopsis thaliana* RNH1A and RNH1B have the canonical RNase H1 activity. Because the presence of RNase H2 complexes has been previously shown[2,3], this indicates that plants, like other eukaryotes, have both RNase H1 and H2. Our results suggest that all three RNase H1 genes identified in the *Arabidopsis thaliana* genome encode canonical *bona fide* RNase H1 proteins.

Arabidopsis thaliana contains four RNase H1 proteins. Their evolutionary origin may be traced down to three events. The first event occurred in the common ancestor of dicots, where nuclear and organellar paralogs have been formed. An independent event in the common ancestor of monocots (or possibly the common ancestor of all angiosperms) led to the formation of similar, yet evolutionarily independent nuclear and organellar paralogs in monocots. Nuclear and organellar RNases H1 do not display any obvious structural differences beyond the presence or absence of a transit peptide/presequence. Therefore, it is unknown if they are functionally distinct beyond having different subcellular localizations and expression patterns.

The second event leading to the formation of four RNase H1 proteins in *Arabidopsis thaliana* occurred in the common ancestor of *Brassicaceae*. The ancestral organellar protein diversified into two proteins, which are represented by RNH1B and RNH1C in *Arabidopsis*. Although these proteins are localized to mitochondria and chloroplasts[23], their ortologs in *Brassicaceae* do not have consistent predicted localizations and the functional impact of this event remains unknown. The third event is alternative splicing of RNH1B, which results in translation of proteins predicted to localize to mitochondria or the nucleus. This mechanism is reminiscent of dual localization of metazoan RHases H1[22]. Overall, our results are consistent with recent evidence that each genome within *Arabidopsis thaliana* cells has at least one RNase H1[23]. However, the nuclear genome is served by two RNase H1 proteins. Additionally, most eukaryotes, including *Arabidopsis thaliana* contain RNase H2 complexes, which are nuclear localized and at least partially redundant with RNases H1[2,31,32]. This may suggests

that the nuclear genome requires the presence of several activities resolving RNA:DNA hybrids.

Our results indicate that at least one organellar RNase H1 is needed for proper embryonic development. This is consistent with data from metazoans, where RNase H1 is crucial for proper mitochondrial DNA replication during embryo development[27,33,34]. Viability of single *atrnh1b* and *atrnh1c* mutants may speculatively be explained by dual (mitochondrial and plastid) localization of those proteins, which is not an uncommon phenomenon[35,36]. Alternatively, defects in both mitochondria and chloroplasts may have a synergistic effect in embryonic development.

RNase H1 has been shown to be required for genome maintenance in chloroplasts[23]. Our observation that plastid DNA copy number is substantially increased in *atrnh1c* may indicate that genome instability in *atrnh1c* leads to overamplification of the entire genome. Because the plastid genome is copied by a combination of replication and recombination[37], this overamplification may rely on homologous recombination. Interestingly, loss of mitochondrial RNH1B does not affect the level of mitochondrial DNA. This may suggest that, contrary to animals, plant mitochondria contain enzymes at least partially redundant with RNH1s.

The reported role of RNH1C in release of replication-transcription conflicts and chloroplast DNA integrity[23] is consistent with our observations that *atrnh1c* mutation increases sensitivity to hydroxyurea. This may be interpreted as evidence of disrupted DNA replication in *atrnh1c*. This result should however be interpreted carefully because hydroxyurea inhibits synthesis of deoxyribonucleotides and increases misincorporation of ribonucleotides[29,30], which may have a replication-independent effect on RNase H1-dependent processes. Additionally, hydroxyurea may cause oxidative stress [38–40], which could also have a replication-independent effect.

Our results raise questions about the roles of nuclear and mitochondrial RNases H1 in *Arabidopsis thaliana*. Single mutations in genes encoding these proteins do not result in visible phenotypes, which may likely be attributed to protein redundancy and localization to multiple cellular compartments. Resolving the roles of those proteins remain an important goal for future studies.

MATERIALS AND METHODS

Plant material and oligonucleotides

Plant lines used in this study: Col-0 (CS 70000), *atrnh1a* (SALK_150285C), *atrnh1b* (SAIL_1174_C11), *atrnh1c* (SAIL_97_E11) and crosses between abovementioned. Oligonucleotides used for expression analysis: P1: 5'-TTTAGTTTGGGTTGATGGGTTCC-3' + 5'-CACACTCATTGCAGGATGTGATAC-3'; P2: 5'-ATCTCTTAGACGGGGAAGATTTGT-3' + 5'-AATGCAAGTCATGTCAAAGATGAT-3'; P3: 5'-TCAATAATGCAAGTTTCATATGAGGT-3' + 5'-GTTTGGAGCTCTTACACCTTGTCT-3'; P4: 5'-ATCCCTTATAAACGCTAACTGGAG-3' + 5'-TCAAGTTGGATCTTCGGTTTATG-3'; P5: 5'-TACACCATGTCTTTTCCAGGAG-3' + 5'-GATATAGAAGCTGAAGGAAGTTGATCT-3'; P6: 5'-AAAGATCCGGAGTTACACACTAGC-3' + 5'-CATTCTGGTTCTTCACCAGTTTCT-3'. Sequences of oligonucleotides used for organellar DNA content quantification were previously described by Kim et al[41].

Phylogenetic analysis

Sequences of putative plant RNH1 proteins were retrieved from Phytozome[42], CoGE[43], Ensembl Plants[44], PLAZA[45] and 1KP[46] through BLAST search and aligned in MAFFT[47]. Only sequences containing full length hybrid binding domain (HBD), RNASEH1-like domain and four amino acids (DEDD) crucial for catalytic activity[17], were kept. BAli-Phy[48] with default settings on CIPRES[49] platform was used for phylogenetic analysis. Consensus tree was generated with 25% burn-in and visualized using FigTree. Subcellular localization was predicted by TargetP[26].

***In vitro* activity of RNH1s**

6xHis-tagged HBDs and catalytic domains of RNH1A and B were expressed in *E. coli* from pET28 plasmid. Tagged proteins were bound on HisTrap (HP)(GE Healthcare) and eluted with increasing imidazole concentration. Enzymatic reaction was performed in Reaction Buffer (60mM NaCl, 16mM HEPES-KOH pH=7.0, 4% glycerol, 0.8mM DTT, 80µg/ml BSA, 4mM $MgCl_2$) with 20nM fluor-labelled substrate and 2nM-2µM purified RNH1 in 20µl reaction. Reaction was conducted for 30min at 37^0C and stopped by addition of EDTA to final concentration of 40mM. Products were analyzed on 15% denaturing TBE-urea polyacrylamide gels and visualized by fluorescence readout.

Chlorophyll content measurement

Chlorophyll content was determined as described by Lichtenthaler[50]. Briefly, 1g of frozen 2-3 week old seedlings was ground in liquid nitrogen and chlorophyll was extracted with 100% acetone. Samples were centrifuged at 10000 x g for 10min at 4^0C and absorbance was measured at 645 and 662nm. Chlorophyll concentration was calculated using following formula: C_{a+b}(µg/ml)= 18.09 x A645 + 7.05 x A662.

Microscopic observations

Mature siliques of Col-0 and *rnh1b*$^{-/-}$ / *rnh1c*$^{+/-}$ plants were destained in 70% ethanol and photographed under preparative microscope. For embryos observations siliques of Col-0 and *rnh1b*$^{-/-}$ / *rnh1c*$^{+/-}$ were dissected, seeds were split based on morphology and cleared for 2h at room temperature with Visikol[51]. Embryos inside seeds were visualized using Nomarski optics.

Organellar DNA content quantification and expression analysis

Total DNA was isolated from 2-3 weeks old seedlings with DNeasy Plant Mini Kit (QIAGEN). 1ng of DNA per qPCR reaction was used as template. For expression analysis total RNA was extracted with RNeasy Plant Mini Kit (QIAGEN) from 2-3 weeks old seedlings and 500ng was treated with DNaseI (Ambion) and used for reverse transcription. 0.5ng of cDNA was used as template in qPCR. Significance was determined using confidence intervals.

Replicative stress assay

Seeds were germinated on ½ MS plates supplemented with 1% sucrose and 0.75% agar. After 1 week seedlings were transferred to liquid ½ MS + 1% sucrose with 0 or 2mM hydroxyurea (HU) and grown in continuous light. After 2 weeks root length was measured.

ACKNOWLEDGEMENTS

We want to thank Steven Smith for his help with phylogenetic analysis as well as Karolina Górecka and Małgorzata Figiel for their help with protein purification and biochemical activity assays. This work was supported by the National Science Foundation grant MCB 1120271 and the National Science Centre, Poland grant Polonez 2015/19/P/NZ1/03619 to A.T.W. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665778.

FIGURE LEGENDS

Figure 1. Evolutionary origin of plant RNase H1 proteins. Phylogenetic tree of all full length predicted RNase H1 proteins. Support values at tree branches are posterior probability scores, which integrate over the uncertainty in both the alignment and the phylogeny. Predicted protein localizations obtained using TargetP are marked with colors. Mitochondria – red, chloroplast – green, other – blue. A version of this phylogenetic tree with all species names and support values is shown in Fig. S1.

Figure 2. Origin of four RNase H1 proteins in *Arabidopsis thaliana*. **(A)** Predicted localization of RNase H1 proteins from distinct clades identified in Fig 1. **(B)** Detailed phylogenetic tree of all RNase H1 proteins in *Brassicaceae*. Support values and color-coded predicted localization like in Fig 1. **(C)** Splice variants of *RNH1B* predicted in Araport11. **(D)** 5' RACE of *RNH1B*. Asterisk indicates a non-specific PCR product. **(E)** Domain composition of four RNase H1 proteins in *Arabidopsis thaliana*.

Figure 3. RNH1A and RNH1B proteins from *Arabidopsis thaliana* exhibit canonical RNase H1 activity *in vitro*. Recombinant RNH1A **(A)** and RNH1B **(B)** were incubated with labeled oligonucleotide substrates containing different combinations of deoxyribonucleotides (black) and ribonucleotides (red).

Figure 4. Effects of mutations in genes encoding RNase H1 proteins. Approximately 3-week-old plants of Col-0 wild type **(A)**, *atrnh1a* **(B)**, *atrnh1b* **(C)**, *atrnh1c* **(D)**, *atrnh1a, atrnh1b* double mutant **(E)**, *atrnh1a, atrnh1c* double mutant **(F)**, *atrnh1a, atrnh1b* double mutant heterozygous for *atrnh1c* **(G)** and *atrnh1a, atrnh1b, atrnh1c* triple mutant expressing RNH1C::RNH1C-GFP **(H)**. **(I)** Relative chlorophyll content of Col-0 wild type and mutants in genes encoding RNase H1 proteins. Error bars indicate standard deviation from three biological replicates. **(J)** Schematic representation of RNH1 genes in *A. thaliana*. Arrows indicate START codons, triangles indicate T-DNA insertion positions. Red boxes indicate positions of the conserved RNase domain. **(K, L)** Expression of genes encoding RNases H1 measured **(K)** upstream of T-DNA insertion and **(L)** downstream of T-DNA insertion. Primer pairs are marked in (J). Error bars indicate standard deviation from three biological replicates. Asterisks indicate significant change in expression determined by 95% confidence intervals.

Figure 5. At least one organellar RNase H1 is required for embryo development**. (A)** Mature silique of Col-0 wild type. **(B)** *atrnh1a, atrnh1b* double mutant heterozygous for *atrnh1c* (*atrnh1a/b/c*$^{+/-}$). **(C)** Percentage of properly developing seeds in siliques. Error bars indicate standard deviation from three biological replicates. **(D)** Development of a Col-0 wild-type embryo. **(E)** Development of an aborted embryo.

Figure 6. Loss of RNH1C leads to increased chloroplast DNA content and hypersensitivity to hydroxyurea. Relative levels of nuclear **(A)**, mitochondrial **(B)** and chloroplast **(C)** DNA in Col-0 wild type *atrnh1* mutants. Asterisks indicate significant changes with $p<0.05$. **(D)** Phenotype of plants subjected to HU stress. **(E)** Relative root length of Col-0 wild type and *atrnh1* mutants upon 2mM hydroxyurea treatment. Asterisks indicate $p<0.0001$determined using 2-way ANOVA.

SUPPLEMENTAL DATA

Figure S1. Evolutionary origin of plant RNase H1 proteins. Phylogenetic tree of all full length predicted RNase H1 proteins as shown in Fig. 1 with all species names and support values included.

File S1. Aminoacid sequences of all RNase H1 proteins shown in Fig. 1 and Fig. S1.

REFERENCES

1. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
2. Kalhorzadeh, P. *et al.* Arabidopsis thaliana RNase H2 deficiency counteracts the needs for the WEE1 checkpoint kinase but triggers genome instability. *Plant Cell* **26**, 3680–3692 (2014).
3. Eekhout, T., Kalhorzadeh, P. & De Veylder, L. Lack of RNase H2 activity rescues HU-sensitivity of WEE1 deficient plants. *Plant Signal. Behav.* **10**, (2015).
4. Stein, H. & Hausen, P. Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science* **166**, 393–395 (1969).
5. Lima, W. F. *et al.* Human RNase H1 discriminates between subtle variations in the structure of the heteroduplex substrate. *Mol. Pharmacol.* **71**, 83–91 (2007).
6. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).

7. Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5785–5790 (2015).
8. Stuckey, R., García-Rodríguez, N., Aguilera, A. & Wellinger, R. E. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5779–5784 (2015).
9. Cloutier, S. C. *et al.* Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Mol. Cell* **61**, 393–404 (2016).
10. Stirling, P. C. & Hieter, P. Canonical DNA Repair Pathways Influence R-Loop-Driven Genome Instability. *J. Mol. Biol.* **429**, 3132–3138 (2017).
11. Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair* **19**, 84–94 (2014).
12. Sollier, J. *et al.* Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
13. Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5**, 5220 (2014).
14. Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **4**, 442–451 (2003).
15. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* **340**, 619–621 (2013).
16. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016 (2005).
17. Nowotny, M. *et al.* Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. *Mol. Cell* **28**, 264–276 (2007).
18. Evans, S. P. & Bycroft, M. NMR structure of the N-terminal domain of Saccharomyces cerevisiae RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9. *J. Mol. Biol.* **291**, 661–669 (1999).
19. Ohtani, N., Yanagawa, H., Tomita, M. & Itaya, M. Identification of the first archaeal Type 1 RNase H gene from Halobacterium sp. NRC-1: archaeal RNase HI can cleave an RNA-DNA junction. *Biochem. J.* **381**, 795–802 (2004).

20. Ohtani, N., Yanagawa, H., Tomita, M. & Itaya, M. Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, Sulfolobus tokodaii 7. *Nucleic Acids Res.* **32**, 5809–5819 (2004).
21. Tadokoro, T. & Kanaya, S. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.* **276**, 1482–1493 (2009).
22. Suzuki, Y. *et al.* An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Mol. Cell. Biol.* **30**, 5123–5134 (2010).
23. Yang, Z. *et al.* RNase H1 Cooperates with DNA Gyrases to Restrict R-Loops and Maintain Genome Integrity in Arabidopsis Chloroplasts. *Plant Cell* **29**, 2478–2497 (2017).
24. Xu, W. *et al.* The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat. Plants* **3**, 704–714 (2017).
25. Suchard, M. A. & Redelings, B. D. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinforma. Oxf. Engl.* **22**, 2047–2048 (2006).
26. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
27. Cerritelli, S. M. *et al.* Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol. Cell* **11**, 807–815 (2003).
28. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J. Cell Mol. Biol.* **89**, 789–804 (2017).
29. Arudchandran, A. *et al.* The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes Cells Devoted Mol. Cell. Mech.* **5**, 789–802 (2000).
30. Yarbro, J. W. Mechanism of action of hydroxyurea. *Semin. Oncol.* **19**, 1–10 (1992).
31. Chon, H. *et al.* RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* **41**, 3130–3143 (2013).
32. Zimmer, A. D. & Koshland, D. Differential roles of the RNases H in preventing chromosome instability. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12220–12225 (2016).
33. Holmes, J. B. *et al.* Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9334–9339 (2015).

34. Lima, W. F. *et al.* Viable RNaseH1 knockout mice show RNaseH1 is essential for R loop processing, mitochondrial and liver function. *Nucleic Acids Res.* **44**, 5299–5312 (2016).
35. Carrie, C., Giraud, E. & Whelan, J. Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *FEBS J.* **276**, 1187–1195 (2009).
36. Silva-Filho, M. C. One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.* **6**, 589–595 (2003).
37. Oldenburg, D. J. & Bendich, A. J. DNA maintenance in plastids and mitochondria of plants. *Front. Plant Sci.* **6**, 883 (2015).
38. Wood, K. C. & Granger, D. N. Sickle Cell Disease: Role of Reactive Oxygen and Nitrogen Metabolites. *Clin. Exp. Pharmacol. Physiol.* **34**, 926–932
39. Burkitt, M. J. Nitric oxide generation from hydroxyurea: significance and implications for leukemogenesis in the management of myeloproliferative disorders. *Blood* **107**, 2219–2222 (2006).
40. Sakano, K., Oikawa, S., Hasegawa, K. & Kawanishi, S. Hydroxyurea induces site-specific DNA damage via formation of hydrogen peroxide and nitric oxide. *Jpn. J. Cancer Res. Gann* **92**, 1166–1174 (2001).
41. Kim, M., Lee, U., Small, I., des Francs-Small, C. C. & Vierling, E. Mutations in an Arabidopsis Mitochondrial Transcription Termination Factor–Related Protein Enhance Thermotolerance in the Absence of the Major Molecular Chaperone HSP101[W]. *Plant Cell* **24**, 3349–3365 (2012).
42. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
43. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J. Cell Mol. Biol.* **53**, 661–673 (2008).
44. Kersey, P. J. *et al.* Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **46**, D802–D808 (2018).
45. Van Bel, M. *et al.* PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190–D1196 (2018).
46. Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**, 17 (2014).
47. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).

48. Suchard, M. A. & Redelings, B. D. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinforma. Oxf. Engl.* **22**, 2047–2048 (2006).
49. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *2010 Gateway Computing Environments Workshop (GCE)* 1–8 (2010). doi:10.1109/GCE.2010.5676129
50. Lichtenthaler, H. K. [34] Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. in *Methods in Enzymology* **148**, 350–382 (Academic Press, 1987).
51. Villani, T. S., Koroch, A. R. & Simon, J. E. An improved clearing and mounting solution to replace chloral hydrate in microscopic applications1. *Appl. Plant Sci.* **1**, (2013).

Figure 1

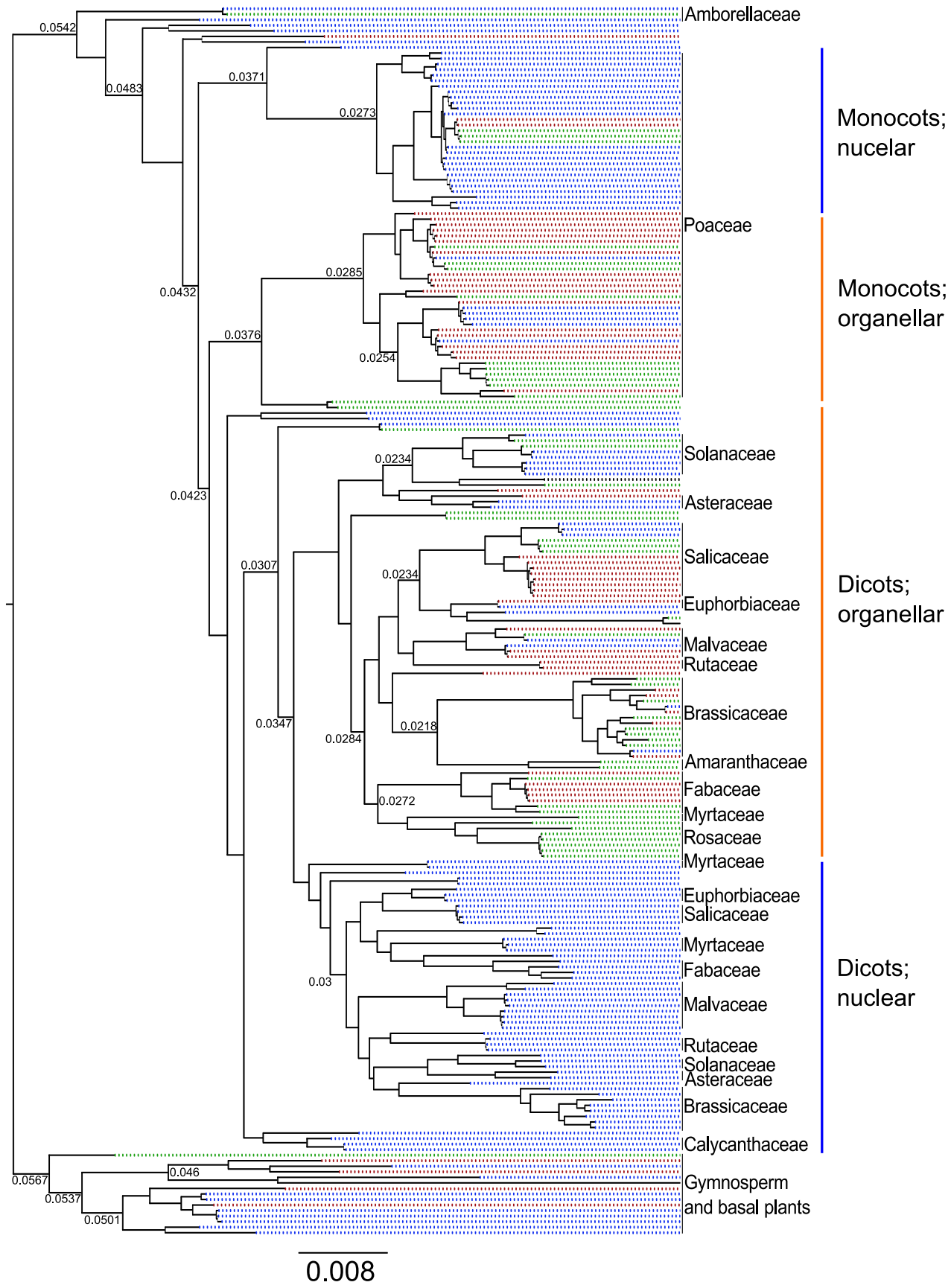

Amborellaceae
Poaceae
Solanaceae
Asteraceae
Salicaceae
Euphorbiaceae
Malvaceae
Rutaceae
Brassicaceae
Amaranthaceae
Fabaceae
Myrtaceae
Rosaceae
Myrtaceae
Euphorbiaceae
Salicaceae
Myrtaceae
Fabaceae
Malvaceae
Rutaceae
Solanaceae
Asteraceae
Brassicaceae
Calycanthaceae
Gymnosperm
and basal plants
Monocots;
nucelar
Monocots;
organellar
Dicots;
organellar
Dicots;
nuclear
0.0542
0.0483
0.0371
0.0273
0.0432
0.0285
0.0376
0.0254
0.0423
0.0234
0.0307
0.0234
0.0347
0.0284
0.0218
0.0272
0.03
0.0567
0.046
0.0537
0.0501
0.008

Figure 2

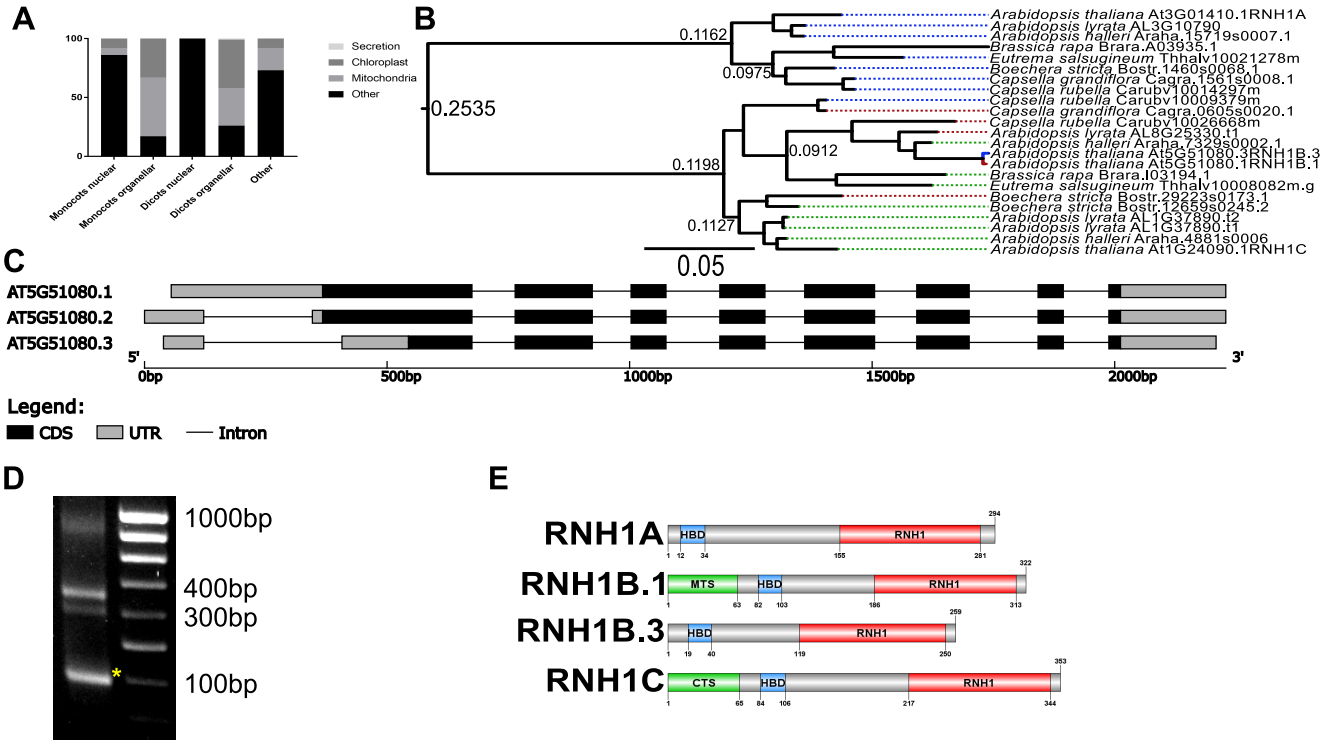

A
100
50
0
Secretion
Chloroplast
Mitochondria
Other
Monocots nuclear
Monocots organellar
Dicots nuclear
Dicots organellar
Other
B
0.1162
0.0975
0.2535
0.1198
0.0912
0.1127
0.05
Arabidopsis thaliana At3G01410.1RNH1A
Arabidopsis lyrata AL3G10790
Arabidopsis halleri Araha.15719s0007.1
Brassica rapa Brara.A03935.1
Eutrema salsugineum Thhalv10021278m
Boechera stricta Bostr.1460s0068.1
Capsella grandiflora Cagra.1561s0008.1
Capsella rubella Carubv10014297m
Capsella rubella Carubv10009379m
Capsella grandiflora Cagra.0605s0020.1
Capsella rubella Carubv10026668m
Arabidopsis lyrata AL8G25330.t1
Arabidopsis halleri Araha.7329s0002.1
Arabidopsis thaliana At5G51080.3RNH1B.3
Arabidopsis thaliana At5G51080.1RNH1B.1
Brassica rapa Brara.I03194.1
Eutrema salsugineum Thhalv10008082m.g
Boechera stricta Bostr.29223s0173.1
Boechera stricta Bostr.12659s0245.2
Arabidopsis lyrata AL1G37890.t2
Arabidopsis lyrata AL1G37890.t1
Arabidopsis halleri Araha.4881s0006
Arabidopsis thaliana At1G24090.1RNH1C
C
AT5G51080.1
AT5G51080.2
AT5G51080.3
5'
3'
0bp
500bp
1000bp
1500bp
2000bp
Legend:
CDS
UTR
Intron
D
1000bp
400bp
300bp
100bp
*
E
RNH1A
HBD
RNH1
RNH1B.1
MTS
HBD
RNH1
RNH1B.3
HBD
RNH1
RNH1C
CTS
HBD
RNH1

Figure 3

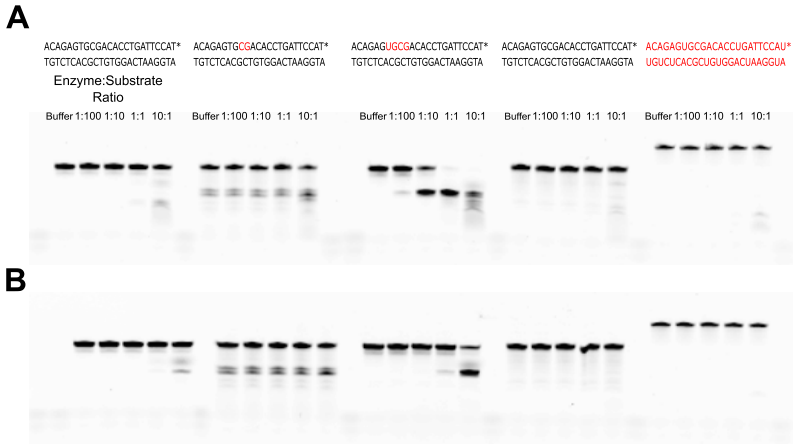

A
ACAGAGTGCGACACCTGATTCCAT*
TGTCTCACGCTGTGGACTAAGGTA
ACAGAGTGCGACACCTGATTCCAT*
TGTCTCACGCTGTGGACTAAGGTA
ACAGAGUGCGACACCTGATTCCAT*
TGTCTCACGCTGTGGACTAAGGTA
ACAGAGTGCGACACCTGATTCCAT*
TGTCTCACGCTGTGGACTAAGGTA
ACAGAGUGCGACACCUGATTCCAU*
UGUCUCACGCUGUGGACUAAGGUA
Enzyme:Substrate Ratio
Buffer 1:100 1:10 1:1 10:1
Buffer 1:100 1:10 1:1 10:1
Buffer 1:100 1:10 1:1 10:1
Buffer 1:100 1:10 1:1 10:1
Buffer 1:100 1:10 1:1 10:1
B

Figure 4

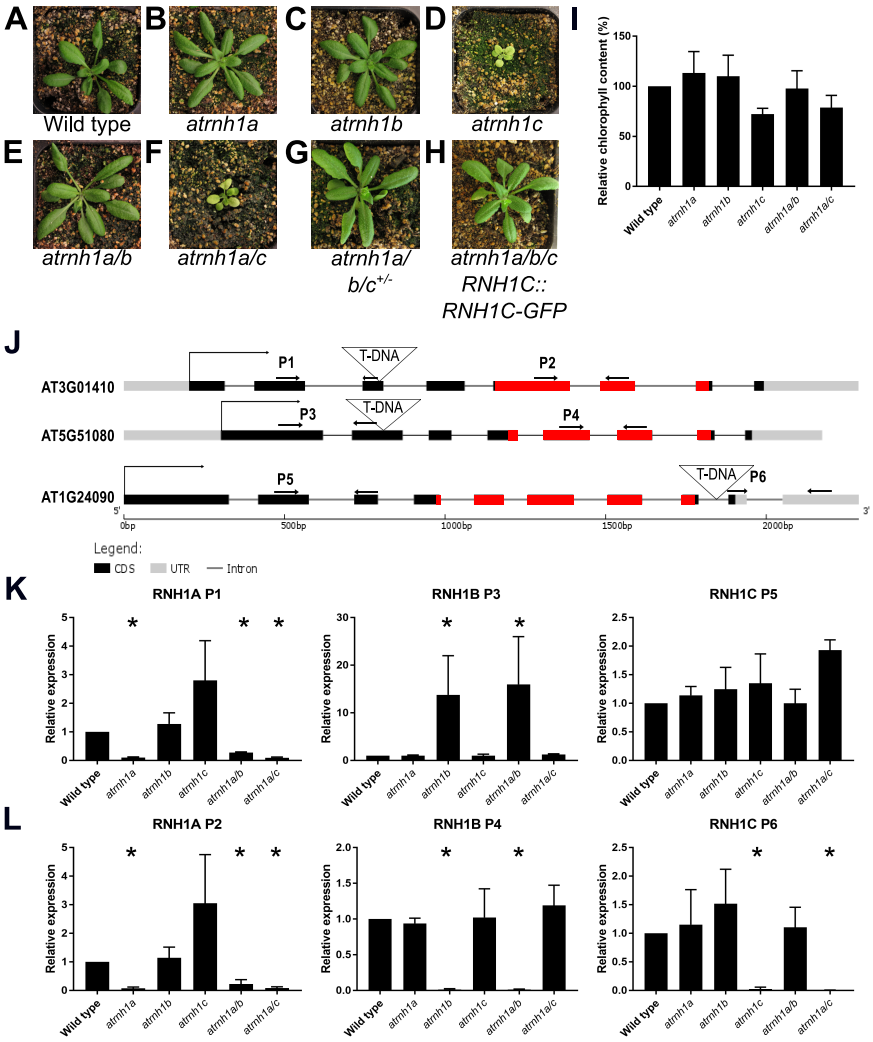

A
Wild type
B
atrnh1a
C
atrnh1b
D
atrnh1c
E
atrnh1a/b
F
atrnh1a/c
G
atrnh1a/ b/c+/-
H
atrnh1a/b/c RNH1C:: RNH1C-GFP
I
Relative chlorophyll content (%)
J
AT3G01410
AT5G51080
AT1G24090
P1
P2
P3
P4
P5
P6
T-DNA
Legend:
CDS
UTR
Intron
K
RNH1A P1
RNH1B P3
RNH1C P5
L
RNH1A P2
RNH1B P4
RNH1C P6
Relative expression
Wild type
atrnh1a
atrnh1b
atrnh1c
atrnh1a/b
atrnh1a/c

Figure 5

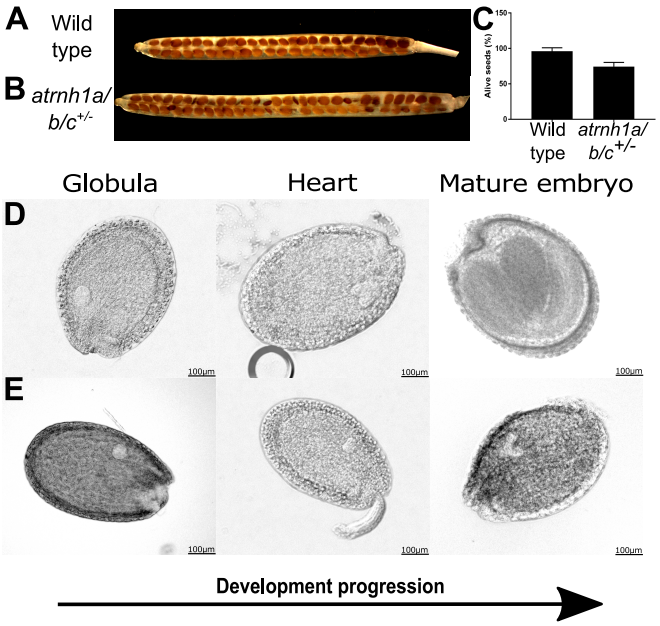
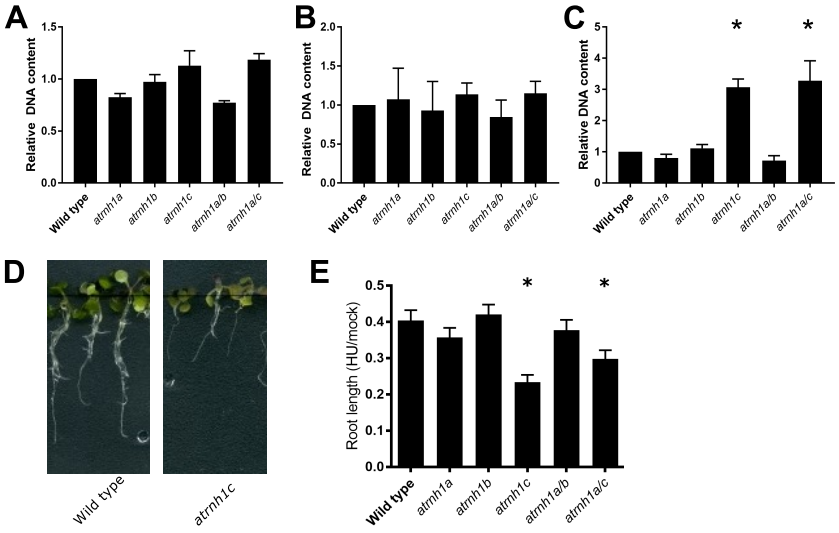
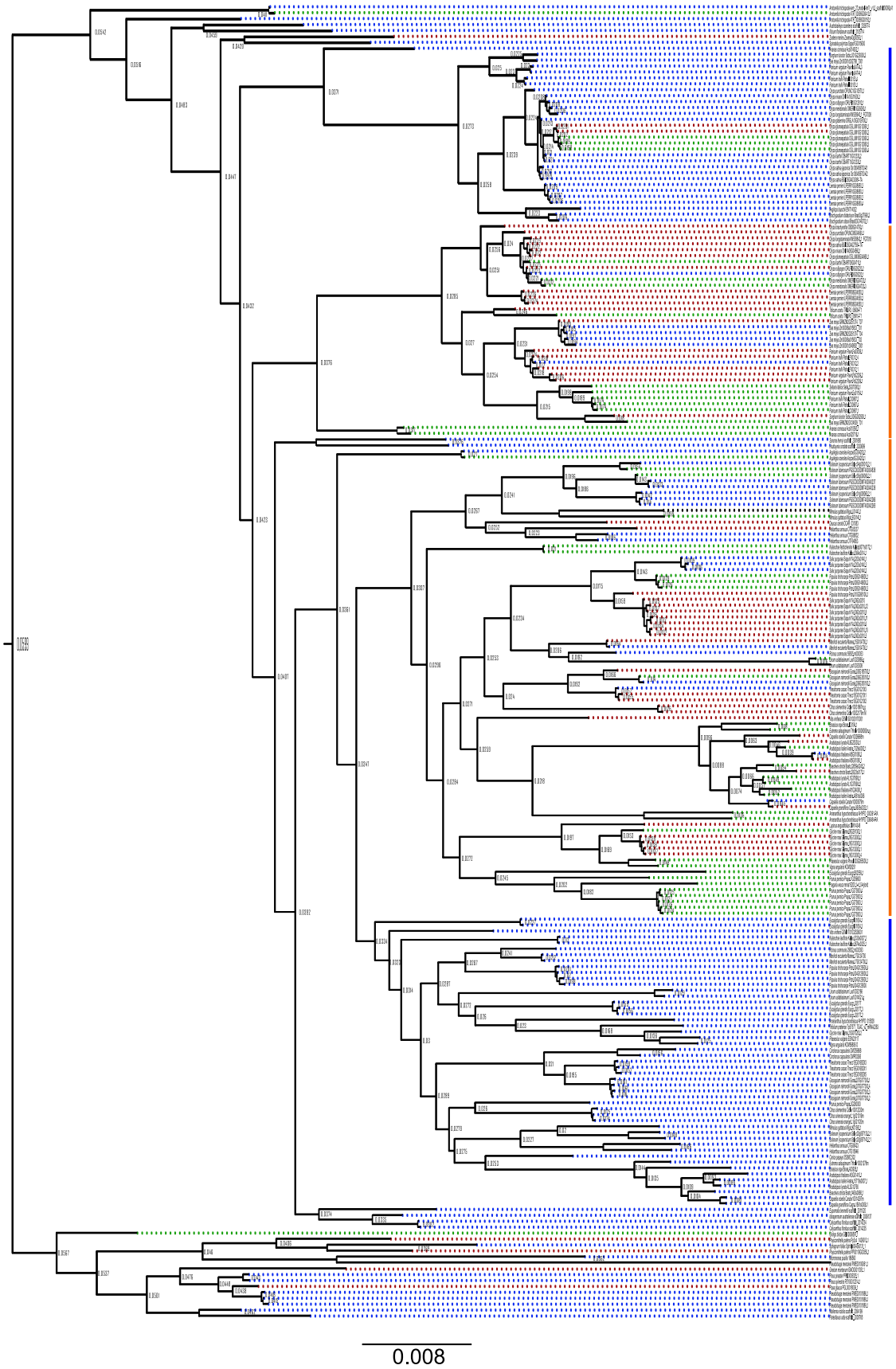

A
Wild type
B
atrnh1a/ b/c+/-
C
Alive seeds (%)
150
100
50
0
Wild type
atrnh1a/ b/c+/-
Globula
Heart
Mature embryo
D
E
100µm
Development progression

Figure 6


A
Relative DNA content
1.5
1.0
0.5
0.0
Wild type
atrnh1a
atrnh1b
atrnh1c
atrnh1a/b
atrnh1a/c
B
Relative DNA content
2.0
1.5
1.0
0.5
0.0
Wild type
atrnh1a
atrnh1b
atrnh1c
atrnh1a/b
atrnh1a/c
C
Relative DNA content
5
4
3
2
1
0
*
*
Wild type
atrnh1a
atrnh1b
atrnh1c
atrnh1a/b
atrnh1a/c
D
Wild type
atrnh1c
E
Root length (HU/mock)
0.5
0.4
0.3
0.2
0.1
0.0
*
*
Wild type
atrnh1a
atrnh1b
atrnh1c
atrnh1a/b
atrnh1a/c

Supplemental Figure 1