

Title:

The Chaperone NASP Contributes to *De Novo* Deposition of the Centromeric Histone Variant CENH3 in *Arabidopsis* Early Embryogenesis

Running head:

CENH3 Deposition by NASP during Embryogenesis

Authors:

Hidenori Takeuchi ^{1,2*}, Shiori Nagahara ¹, Tetsuya Higashiyama ^{1,3,4} and Frédéric Berger ⁵

¹ Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

² Institute for Advanced Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

³ Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

⁴ Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁵ Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Vienna, Austria

***Correspondence:**

Hidenori Takeuchi (hidenori.takeuchi@itbm.nagoya-u.ac.jp)

26 **Abstract**

27 The centromere is an essential chromosome region where the kinetochore is formed to
 28 control equal chromosome distribution during cell division. The centromere-specific
 29 histone H3 variant CENH3 (also called CENP-A) is a prerequisite for the kinetochore
 30 formation. Since CENH3 evolves rapidly, associated factors, including histone
 31 chaperones mediating the deposition of CENH3 on the centromere, are thought to act
 32 through species-specific amino-acid sequences. The functions and interaction networks
 33 of CENH3 and histone chaperons have been well-characterized in animals and yeasts.
 34 However, molecular mechanisms involved in recognition and deposition of CENH3 are
 35 still unclear in plants. Here, we used a swapping strategy between domains of CENH3
 36 of *Arabidopsis thaliana* and the liverwort *Marchantia polymorpha* to identify specific
 37 regions of CENH3 involved in targeting the centromeres and interacting with the
 38 general histone H3 chaperone, NASP (nuclear autoantigenic sperm protein). CENH3's
 39 LoopN- α 1 region was necessary and sufficient for the centromere targeting in
 40 cooperation with the α 2 region and was involved in interaction with NASP in
 41 cooperation with α N, suggesting a species-specific CENH3 recognition. In addition, by
 42 generating an *Arabidopsis nasp* knockout mutant in the background of a fully fertile
 43 *GFP-CENH3/cenh3-1* line, we found that NASP was implicated for *de novo* CENH3
 44 deposition after fertilization and thus for early embryo development. Our results imply
 45 that the NASP mediates the supply of CENH3 in the context of the rapidly evolving
 46 centromere identity in land plants.

47 **Keywords:**

48 *Arabidopsis thaliana*, centromere-specific histone H3 (CENH3), embryogenesis, histone

50 chaperone, nuclear autoantigenic sperm protein (NASP)

51

52 **Introduction**

53 In nuclei of eukaryotic cells, DNA is organized into chromatin that consists of an array of
 54 nucleosomes and associated proteins. The nucleosome, a basic unit of chromatin, is
 55 formed with ~150 bp DNA and four core histones, H2A, H2B, H3, and H4 (Luger et al.
 56 1997). Most eukaryote species possess a set of variants for each core histone, so-called
 57 histone variants, and use them to confer specific properties on each genomic region
 58 (Pusarla and Bhargava 2005, Talbert and Henikoff 2010, Kawashima et al. 2015, Jiang
 59 and Berger 2017). The histone monomer consists of the histone-fold domain, which
 60 contains α -helices connected by loops and mediates histone-histone and histone-DNA
 61 interaction, and the unstructured N- and C-terminal tails, which protrude from the
 62 nucleosome core particle and generally undergo covalent modifications, including
 63 methylation, acetylation, phosphorylation, and ubiquitination (Luger et al. 1997,
 64 Khorasanizadeh 2004).

65 The histone variants H3.1 and H3.3 are incorporated into nucleosomes of
 66 non-centromeric chromatin during DNA replication and throughout cell cycle,
 67 respectively (Filipescu et al. 2013, Talbert and Henikoff 2017, Borg et al. 2021). The
 68 histone variant CENH3 specifically marks the centromere. Due to the importance of
 69 CENH3 targeting to the centromere in eukaryotes, regulatory factors and DNA sequences
 70 related to CENH3 deposition as well as their evolution have been extensively studied in
 71 yeast, animals, and plants (Talbert et al. 2002, Malik et al. 2002, Henikoff and Dalal
 72 2005, Morris and Moazed 2007, Zhang et al. 2008, Malik and Henikoff 2009, Hirsch et
 73 al. 2009, Nagaki et al. 2010, Yuan et al. 2015, Maheshwari et al. 2017, Rosin and
 74 Mellone 2017, Naish et al. 2021). In eukaryotes, the deposition of CENH3 proteins, also
 75 called CENP-A in mammals, CID in flies, Cse4 in budding yeast *Saccharomyces*

76 *cerevisiae*, and Cnp1 in fission yeast *Schizosaccharomyces pombe*, is required for proper
77 point kinetochore formation and thus equal chromosome distribution to daughter cells
78 after cell division. In contrast with other H3 variants, the amino-acid sequences of
79 CENH3 proteins are variable among species, even close relatives (Malik et al. 2002,
80 Nagaki et al. 2010, Sanei et al. 2011, Rosin and Mellone 2016). Similarly, both
81 centromeric DNA repeats and centromeric/heterochromatic proteins evolve rapidly.

82 In contrast with relatively-conserved canonical H3s, rapidly evolving CENH3
83 proteins are recognized by distantly-related or non-conserved histone chaperones,
84 HJURP in mammals, Scm3 in yeasts, and CAL1 in flies (Camahort et al. 2007, Stoler et al.
85 2007, Dunleavy et al. 2009, Foltz et al. 2009, Zhou et al. 2011, Phansalkar et al. 2012,
86 Chen et al. 2014). In humans and yeasts, the chaperones HJURP and Scm3 recognize the
87 loop 1 and $\alpha 2$ helix (designated as the CENP-A targeting domain, CATD) within the
88 histone-fold domain of CENH3 (CENP-A/Cse4) (Cho and Harrison 2011, Bassett et al.
89 2012). In *Drosophila* species, CAL1 distinguishes *Drosophila* CENH3 homologs via the
90 loop 1 in a species-specific manner (Rosin and Mellone 2016). HJURP, Scm3, and CAL1
91 chaperones specifically interact with CENH3 and mediate *de novo* CENH3 deposition via
92 self-sustaining epigenetic mechanism, which maintains centromere positioning and
93 occasionally creates novel centromeres (Barnhart et al. 2011, Roure et al. 2019,
94 Medina-Pritchard et al. 2020, Palladino et al. 2020). Similar to these features of animals
95 and yeasts, it is suggested in plants that CENH3 recognition depends on its
96 species-specific sequence and that CENH3 deposition is mediated by a yet
97 uncharacterized self-sustaining mechanism (Ravi et al. 2010, Maheshwari et al. 2015,
98 Marimuthu et al. 2021). In flowering plants (angiosperms), failure of CENH3 deposition
99 and function leads to uniparental genome elimination after fertilization and

100 consequently to generation of haploid individuals, which is utilized as doubled haploid
101 technology to rapidly obtain homozygous lines (Ravi and Chan 2010, Ravi et al. 2014).
102 Therefore, molecular mechanisms involved in CENH3 recognition and deposition have
103 been intensively studied in crop species as well as *A. thaliana* (Lermontova et al. 2006,
104 Ingouff et al. 2010, Sanei et al. 2011, Karimi-Ashtiyania et al. 2015, Kuppu et al. 2020).
105 However, our knowledge about the requirements for recognition and deposition of plant
106 CENH3 is still limited.

107 A homolog of NASP (nuclear autoantigenic sperm protein) in *A. thaliana* is also
108 characterized as a common histone H3 chaperone (Maksimov et al. 2016, Le Goff et al.
109 2020). NASP is a tetratricopeptide repeat (TPR)-containing protein conserved in
110 wide-range of eukaryotes and is involved in various cellular events, such as DNA
111 replication, cell proliferation, cell growth, and embryonic development in mammals
112 (Richardson et al. 2000, 2006, Dunleavy et al. 2007, Finn et al. 2012). NASP binds to
113 H3-H4 dimers to reserve them in the soluble/nonnucleosomal fraction and hand off
114 them to chromatin assembly factors, suggesting the function of histone supply chain for
115 the demand in mammalian cells (Campos et al. 2010, Cook et al. 2011). Human NASP
116 protein possess the direct nucleosome assembly activity toward CENH3 (CENP-A) as
117 well as the canonical H3.1 and H3.3 *in vitro* (Osakabe et al. 2010). In *S. pombe*, NASP
118 (called as Sim3) binds to CENH3 (Cnp1) and is required for its deposition on
119 centromere (Dunleavy et al. 2007). Recent studies in *A. thaliana* have identified NASP
120 as an interactor of both H3.3 and CENH3 through co-immunoprecipitation and mass
121 spectrometry analysis using *A. thaliana* tissues or cultured cell line (Maksimov et al.
122 2016, Le Goff et al. 2020). Similar to yeast and human NASPs, *A. thaliana* NASP is
123 broadly expressed in actively dividing tissues and is a soluble protein localized at the

nucleoplasm. *In vitro* experiments using purified histones and NASP have demonstrated that *A. thaliana* NASP binds monomeric H3.1 and H3.3 as well as H3.1-H4 and H3.3-H4 dimers, but not monomeric H4, stabilizes H3-H4 tetramers from H3-H4 dimers, and promotes tetrasome formation with DNA (Maksimov et al. 2016). Interaction assay using a bimolecular fluorescence complementation (BiFC) have suggested that *A. thaliana* NASP interacts with both N-terminal tail and histone-fold domain of CENH3 *in planta* (Le Goff et al. 2020). Knockdown of *NASP* expression causes moderate reduction of nuclear CENH3 levels and ~30% reduction of CENH3 signal at the centromere (Le Goff et al. 2020).

Here we show large-scale and detailed structure-function analysis of plant CENH3 in terms of centromere targeting in *A. thaliana* cells and investigate NASP function in *A. thaliana* embryogenesis by generating a *nasp* knockout mutant. By a domain swapping strategy using CENH3 proteins of *A. thaliana* and the liverwort *Marchantia polymorpha*, we reveal the requirement of specific CENH3 regions, such as LoopN- α 1 region, for centromere targeting and interaction with NASP in a lineage-specific manner. Additionally, we characterize the role of NASP in *de novo* CENH3 deposition after fertilization

Results and Discussion

Targeting *A. thaliana* centromere by CENH3 protein involves its angiosperm-specific amino-acid sequence context

To understand a lineage-specific recognition mechanism of plant CENH3 proteins for the centromere targeting, we compared CENH3 amino-acid sequences of the dicot *A.*

148 *thaliana* (HTR12; AtCENH3), the monocot *Zea mays* (ZmCENH3) that has previously
149 shown to target *A. thaliana* centromere (Ravi et al. 2010), the liverwort *M. polymorpha*
150 (MpCENH3), and the human (CENP-A) (Fig. 1A and Supplementary Fig. S1). In
151 contrast to the canonical H3.3 histones, which share the same amino-acid sequences
152 between *A. thaliana* and *M. polymorpha* and have only six amino-acid differences
153 between human and land plants (Supplementary Fig. S1A), CENH3 proteins are highly
154 variable (Fig. 1A) as illustrated previously (Ravi et al. 2010). The N-terminal tail of
155 CENH3 proteins is especially hypervariable with different length, in contrast with the
156 strong degree of conservation of the histone-fold domain (Fig. 1A and Supplementary Fig.
157 S1B).

158 Previous studies have shown that CENH3 proteins from angiosperms, including
159 several Brassicaceae species, the grapevine *Vitis vinifera*, and the monocot *Z. mays*, are
160 able to target *A. thaliana* centromeres and to complement the embryo lethality of *cenh3-1*
161 null mutation when expressed without GFP tag (Ravi et al. 2010, Maheshwari et al. 2015).
162 The histone-fold domain of CENH3 is sufficient to target the centromere (Ravi and Chan
163 2010). To obtain further insight into the sequence specificity for the CENH3 function, we
164 used the MpCENH3 sequence of the distant land plant *M. polymorpha*, which is
165 substantially different from *A. thaliana* CENH3 in the histone-fold domain [58% (56/97),
166 including gaps], even though ZmCENH3 is similarly different (58%, 57/98) (Fig. 1A). To
167 assess the protein localization of CENH3 variants after virtually comparable translation
168 levels in plant cells, we constructed an expression vector that consists of the *cauliflower*
169 *mosaic virus 35S* promoter, *mClover* encoding the green fluorescent protein, and
170 *NLS-mCherry* for the nuclear-localized red fluorescent protein, connected by *P2A*
171 sequence for the 2A peptide from porcine teschovirus-1 (Fig. 1B). The P2A sequence

causes co-translational “self-cleavage” during translation as a result of ribosome skipping (Kawashima et al. 2014; Liu et al. 2017), which generates mClover-fused CENH3 variant and NLS-mCherry proteins from a single mRNA.

To obtain steady and reliable results, we generated transgenic *A. thaliana* individuals harboring series of CENH3 variants and observed mClover and mCherry fluorescence in the petal, which contains actively dividing cells and is useful to capture fluorescent signals with lower background (Fig. 1C-H). In NLS-mCherry-positive nuclei, AtCENH3-mClover and ZmCENH3-mClover marked dot-like centromere regions. As a control, the exchange of the four contiguous amino acids EALV located N-terminal of $\alpha 2$ helix following Loop1, which distinguish CENH3, to HAVL from H3.3 caused a dramatic reduction of the mClover signal with only faint dot-like localization (Fig. 1C, AtCENH3^{HAVL}). Contrasting with AtCENH3 and ZmCENH3, MpCENH3-mClover signal were undetectable (Fig. 1C), suggesting that MpCENH3 misses an angiosperm-specific sequence necessary for its deposition on the centromeres.

CENH3 deposition on *A. thaliana* centromeres requires its LoopN- $\alpha 1$ region

We next sought to define specific regions of CENH3 important for the lineage-specific amino-acid recognition in the *A. thaliana* cell. By generating a series of transgenic *A. thaliana* expressing chimeric CENH3 proteins, which consist of each block from AtCENH3 and MpCENH3 (Fig. 1A and Supplementary Fig. S2). Swapping one block of AtCENH3 by that of MpCENH3 indicated that LoopN- $\alpha 1$ (block 3) and $\alpha 2$ (block 5) were required for targeting *A. thaliana* centromere (Fig. 1D). The N-terminal tail from AtCENH3 was not required for centromere targeting of MpCENH3 (Fig. 1E), although the N-terminal tail is important for full CENH3 functions in *A. thaliana* (Ravi and Chan

2010, Maheshwari et al. 2015). In human cells, a domain spanning Loop1 to $\alpha 2$ (designated as CATD) of CENP-A conferred centromere targeting and function when swapped into canonical human H3 (Black et al. 2007). In contrast, a corresponding domain of AtCENH3 failed to complement *A. thaliana cenH3-1* phenotype when swapped into H3.3 (Ravi et al. 2010). We also assessed the impact of CATD (block 4 and 5) of AtCENH3 on localization in *A. thaliana* cell by using MpCENH3 as a donor. The CATD and a larger domain containing the CATD and the $\alpha 3$ region (block 6) of AtCENH3 were not sufficient for centromere targeting (Fig. 1F and Supplementary Fig. S2). On the contrary, mClover signal of a chimeric AtCENH3 having CATD of MpCENH3 was detected in the nucleus and weakly focused on dot-like regions resembling centromeres (Fig. 1F and Supplementary Fig. S2), suggesting that regions outside of CATD are involved in escort and reserve of plant CENH3 at the nucleoplasm through lineage-specific amino-acid recognition.

To determine sufficient regions for *A. thaliana* centromere targeting in the plant CENH3 framework, we analyzed further chimeric CENH3 variants (Fig. 1G). In the presence of the LoopN- $\alpha 1$ (block 3) of AtCENH3, additional AtCENH3 regions, such as N-terminal tail and αN regions (block 1-2) or $\alpha 2$ region (block 5), conferred the capability to target centromeres in the lineage-specific manner. We searched angiosperms' CENH3-specific amino acids responsible for targeting *A. thaliana* centromere and found that two AtCENH3 mutants possessing single amino acid substitution, I107Q and S148T, abolished centromere targeting (Fig. 1H).

To complement these results obtained in the *A. thaliana* cell, we also observed tobacco BY-2 cultured cell lines harboring each chimeric CENH3 construct (Supplementary Fig. S2 and S3). We obtained fundamentally similar results in the BY-2

220 system to the *A. thaliana* system: LoopN- α 1 (block 3) of AtCENH3 was the necessary
 221 region for centromere targeting. The I107A mutant, which has a single substitution by
 222 MpCENH3-type amino acid, abolished specific targeting to the centromere in the BY-2
 223 nucleus, while the S148T mutant was localized at centromere at comparable level to
 224 wild-type AtCENH3 (Supplementary Fig. S3). In summary, by the swapping strategy
 225 using AtCENH3 and MpCENH3 sequences, we found that the CENH3's LoopN- α 1
 226 region in combination with additional regions, such as α 2, is the key for the centromere
 227 targeting in *A. thaliana* and BY-2 cells and indicated the lineage-specific amino-acid
 228 recognition mechanism of plant CENH3.

229

230 **The general H3/CENH3 chaperone NASP contributes to embryogenesis in *A.*** 231 ***thaliana***

232 In most eukaryotes, CENH3 is the indispensable protein by the common function in
 233 centromere specification to form the kinetochore, whereas it evolves rapidly. In
 234 accordance with this divergence, distantly-related or non-conserved histone chaperones,
 235 HJURP in mammals, Scm3 in yeasts, and CAL1 in flies have been identified as
 236 species-specific chaperons for CENH3/CENP-A/Cse4/CID (Stoler et al. 2007, Camahort
 237 et al. 2007, Dunleavy et al. 2009, Foltz et al. 2009, Zhou et al. 2011, Phansalkar et al.
 238 2012, Chen et al. 2014). Although no specific chaperone that specifically deposit CENH3
 239 have been identified in the plant lineage, one of the homologs of histone chaperones
 240 found in animals and yeasts, NASP, has been reported to be a general H3 chaperone that
 241 escorts CENH3 as well as H3.1 and H3.3 into the nucleoplasm in *A. thaliana* (Maksimov
 242 et al. 2016, Le Goff et al. 2020).

243 We initially tried to identify a novel specific chaperone that functions in a

species-specific CENH3 recognition by our reverse-genetic mini-screening of CENH3-related factors, including unknown function genes co-expressed with *CENH3*, but we failed to get putative knockout mutants showing the embryo lethality similar to *cenh3* mutant. We also generated CRISPR/Cas9-mediated mutants for homologs of histone chaperones in the transgenic plant line *GCH3*, that expresses *GFP-CENH3* in *cenh3-1* mutant background (Ravi et al. 2010). We obtained a heterozygous 10-bp deletion mutant of *NASP* gene in line *GCH3* (*nasp^{cri}/NASP GCH3*) but never obtained a homozygous *nasp^{cri} GCH3* mutant from progeny of the heterozygous line. In self-pollinated *nasp^{cri}/NASP GCH3* siliques approximately 25% of developing seeds were white embryoless seeds or collapsed brown seeds, indicative of the embryonic defect (Fig. 2A-D). About 22% of ovules in self-pollinated *nasp^{cri}/NASP GCH3* siliques had an arrested embryo at 1- to 2-cell stages, compared with ovules of wild-type Col-0 and *GCH3* containing a normal heart stage embryo (Fig. 2E, F). This phenotype was not observed after genetic complementation in T1 plants carrying the *NASPpro::NASP-mRuby2* (native promoter-driven genomic *NASP* fused with red fluorescent protein gene *mRuby2*) (Fig. 2D, *At^{genomic}*). From these results, we concluded that the 10-bp deletion recessive mutation was responsible for the loss of function *NASP* and the observed embryogenesis defect in homozygous *nasp^{cri} GCH3* seeds. Our results first indicate that *NASP* is indispensable in the *A. thaliana* development in a certain condition, at least *GCH3* background.

NASP is required for early embryo development after the first zygotic division of the *A. thaliana* *GFP-CENH3* line

To clarify the contribution of *NASP* to embryogenesis and CENH3 deposition after

268 fertilization, we obtained *GCH3* plant lines with homozygous *nasp^{cri}* mutation
 269 complemented with hemizygous *NASPpro::NASP-mRuby2* transgene. By observing red
 270 fluorescence of NASP-mRuby2 at the nucleus, we were able to distinguish whether
 271 zygotes and early embryos in self-pollinated pistils consisted of cells of functional *NASP*
 272 (homozygous or hemizygous *NASPpro::NASP-mRuby2*) or *nasp^{cri} GCH3* mutant (Fig. 3).
 273 At 2 days after pollination (2 DAP), most NASP-mRuby2-positive embryos normally
 274 developed at 4- or 8-cell stages, whereas NASP-mRuby2-negative ones were at 1- or
 275 2-cell stages (Fig. 3A, C). At 4 DAP, NASP-mRuby2-positive embryos further developed
 276 at about the early heart stage without any obvious morphological defects, whereas
 277 NASP-mRuby2-negative ones were still at the 1- or 2-cell stages (Fig. 3B). Normally
 278 developing NASP-mRuby2-positive embryo cells exhibited GFP-CENH3 signals, while
 279 the NASP-mRuby2-negative embryo cells had no GFP-CENH3 signals in the nucleus
 280 (Fig. 3A, B). These results suggested that the absence of NASP caused failure of CENH3
 281 deposition after fertilization causing the early arrest of the embryo at the 1- or 2-cell
 282 stages.

283

284 **Species-specific sequences of NASP involve its function in embryogenesis**

285 Taking advantage of the *nasp^{cri}/NASP GCH3* line that showed the clear embryogenesis
 286 phenotype (Fig. 2C, E), we assessed the importance of NASP protein and its domains in
 287 terms of species-specific amino-acid sequences by the complementation test. As similar
 288 to the assessment for CENH3 as shown in Fig. 1, we focused on NASP from *A. thaliana*,
 289 a Brassicaceae species *Capsella rubella*, *Z. mays*, and *M. polymorpha*. The introduction
 290 of coding sequences for AtNASP and CrNASP but not ZmNASP and MpNASP restored
 291 the abnormal seed phenotype of *nasp^{cri}/NASP GCH3* (Fig. 2D), suggesting that NASP

292 required a certain degree of species-specific amino-acid sequence for its function.

293 We compared amino-acid sequences of NASP from human and these four plant
 294 species (Fig. 4A). NASP proteins include four tetratricopeptide repeat (TPR) motifs and a
 295 C-terminal extension containing a nuclear localization signal (Dunleavy et al. 2007,
 296 Bowman et al. 2015, Liu et al. 2021, Bao et al., 2022, Liu et al. 2022). The NASPs
 297 contains an acidic amino acid-rich stretch as an interrupted loop in the second TPR,
 298 which seems to be variable between species, and a less conserved N-terminal region
 299 without any predicted protein structures (Fig. 4A). To examine the requirement of each
 300 NASP region for the function in *A. thaliana* embryogenesis, we introduced AtNASP
 301 variants fused with the cyan fluorescent protein mTurquoise2 into *nasp^{cri}/NASP GCH3*.
 302 Compared with full-length wild-type AtNASP (*At*) as a control, AtNASP lacking the
 303 N-terminal region [*At* (ΔN)] complemented the abnormal seed phenotype despite lower
 304 frequency and degree, and AtNASP lacking the acidic region [*At* (ΔDE)] did not at all
 305 (Fig. 4B). Some T1 line of AtNASP having N-terminal region of MpNASP [*At* (N^{Mp})]
 306 just slightly decreased the abnormal seed phenotype, while more than half of T1 lines of
 307 AtNASP having the acidic region of MpNASP [*At* (DE^{Mp})] restored the phenotype as
 308 wild-type AtNASP (Fig. 4B). Consistent of this, the abnormal embryo phenotype was
 309 restored by the introduction of *At* (DE^{Mp}) but not *At* (N^{Mp}) or MpNASP (*Mp*) (Fig. 4C).
 310 This analysis suggested that the acidic region was required for NASP function. Our
 311 observations are also consistent with the importance of the acidic region of human NASP
 312 for the binding to H3 to make the complex with H3–H4–ASF1 (Bowman et al. 2017).

313

314 ***A. thaliana* NASP interacts with CENH3 through lineage-specific sequence contexts**

315 Our complementation test implied the requirement of lineage-specific sequence contexts

for the NASP function. Since amino-acid sequences of canonical histone H3 are completely the same among land plant species (Supplementary Fig. S1A), we postulated that the lineage-specific NASP-CENH3 interaction would be one of explanations for the functionality of NASP variants. To assess whether CENH3 variants, including chimera of AtCENH3 and MpCENH3, interacted with AtNASP, we performed semi-quantitative interaction assay by a bimolecular fluorescence complementation (BiFC) assay in *N. benthamiana* leaf epidermal cells. As a control, we also assessed the interaction between CENH3 variants and *A. thaliana* ASF1A, a conserved histone chaperone for H3/H4 (Zhu et al. 2011, Min et al. 2019, Zhong et al. 2022).

In our BiFC assay employing the P2A-NLS-mCherry (PNC) module as used in Fig. 1, YFP fluorescence (BiFC signal) should be observed if AtASF1A or AtNASP fused to nYFP interacted with H3.3 or CENH3 variants fused to cYFP-PNC, while mCherry fluorescence certifies comparable translation level of H3.3 or CENH3 variants (Fig. 5). The BiFC assay indicated that AtASF1A interacted with both AtCENH3 and MpCENH3 but AtNASP did only with conspecific AtCENH3 (Fig. 5A). AtASF1A interacted with H3.3, ZmCENH3, and all of At/Mp-chimeric CENH3 variants we tested, despite slightly different relative BiFC intensity (Fig. 5B). In contrast, AtNASP interacted with AtCENH3, ZmCENH3, and some chimeric CENH3 variants but not the others as well as MpCENH3 (Fig. 5C). When a single region, LoopN- α 1 (block 3) or α 2 (block 5) of AtCENH3, which were required for the targeting to *A. thaliana* centromere (Fig. 1D), was swapped by that of MpCENH3, these chimeric CENH3 variants retained the interaction capability with AtNASP (Fig. 5C_i). The introduction of neither block 3 nor blocks 3/5 of AtCENH3 conferred the interaction capability to MpCENH3 framework (Fig. 5C_ii). Interestingly, CENH3 variants with blocks 1/2/3 or blocks 2/3/4

340 of AtCENH3 interacted with AtNASP, while that of MpCENH3 did not (Fig. 5C_iii).
 341 This clear contrast demonstrated that the NASP-CENH3 interaction relied on
 342 angiosperm-specific sequence contexts of CENH3's α N and LoopN- α 1. Together with
 343 the result that the LoopN- α 1 region of AtCENH3 played the dominant role in the
 344 centromere targeting of *A. thaliana* and tobacco BY-2 cells (Fig. 1, Supplementary Fig. 2
 345 and 3), our data suggest that NASP could recognize CENH3 through the LoopN- α 1
 346 region and escort it into the nucleoplasm for the deposition on the centromere
 347 nucleosome.

348 A previous study using structural and biochemical approaches for AtNASP and
 349 canonical H3 showed that AtNASP has distinct two sets of amino-acids clusters for the
 350 binding to H3's N-terminal (a.a. 21–59 containing α N) and α 3 (a.a. 116–135) regions,
 351 respectively, *in vitro* (Liu et al. 2022, Fig. 4A, Supplementary Fig. S1C, D), similar to the
 352 binding mode of human NASP to H3 (Bowman et al. 2017, Liu et al. 2021, Bao et al.,
 353 2022). Although this knowledge is the case for canonical H3, AtNASP might also bind to
 354 CENH3 via these regions, which are not completely conserved between AtCENH3 and
 355 MpCENH3 (Supplementary Fig. S1C, D). Since our interaction assay using *N.*
 356 *benthamiana* implied that AtNASP required AtCENH3's LoopN- α 1 in addition to α N,
 357 but not α 3, for the species-specific NASP-CENH3 interaction (Fig. 5C), the recognition
 358 of CENH3 by NASP could be mediated by different dynamics *in planta*. Altogether, our
 359 results indicated that NASP acts as a key chaperone that bind CENH3 in a species-specific
 360 manner in plants. It should be noted that the completely-arrested embryonic phenotype by
 361 the loss-of-function *nasp* mutation was only observed in the *GCH3* background, so the
 362 GFP tag of CENH3 might interfere interaction with some other components important for
 363 normal CENH3 supply network, including another general H3 chaperone ASF1 (Fig. 5B,

364 Cook et al. 2011).

365

366 **Conclusions**

367 In this study, we defined *A. thaliana* CENH3's domains important for centromere
368 targeting and identified NASP as a chaperone that binds to CENH3 through a domain
369 around LoopN- α 1 in a species-specific sequence manner. In addition, we demonstrated
370 that NASP is involved in early embryo development and CENH3 deposition after
371 fertilization. Although NASP seems to be neither a CENH3-specific chaperone nor a
372 sole chaperone that recognizes CENH3, we propose the hypothesis that NASP
373 contributes to *de novo* deposition of CENH3 on the centromere after fertilization. This
374 function requires species-specific sequences, suggesting coevolution of NASP with
375 CENH3 after fertilization. Our findings help to understand how the CENH3 protein is
376 deposited at the centromere in the zygote nucleus and contribute to haploid induction
377 technology by manipulation of CENH3 deposition network.

378

379

380 **Materials and Methods**

381 **Plant materials**

382 The *A. thaliana* Col-0 was used for the expression and localization experiment of CENH3
383 variants. The *GFP-CENH3* in *cenh3-1* (*GCH3*) line was kindly provided by Dr. Simon
384 Chan's laboratory (Ravi et al. 2010). For transformation of *A. thaliana* plants, T-DNA
385 constructs were introduced by floral dip using *Agrobacterium tumefaciens* strain GV3101.
386 *A. thaliana* plants were germinated on half-strength Murashige-Skoog (MS) solid
387 medium containing antibiotics (if applicable) under continuous light at 22°C and

388 transferred on soil for further growth under long-day condition in plant growth rooms.
 389 The *N. benthamiana* plant was grown from directly sowing seeds on soil in plant growth
 390 rooms under long-day condition at 22-26°C. Transformation of the tobacco BY-2 cultured
 391 cell was performed as previously described (Mayo et al. 2006) using *A. tumefaciens* strain
 392 EHA105. Several transformants per each construct were selected via several round of
 393 subculturing of callus on modified LS medium containing 1.5% agar, 50 mg/L kanamycin
 394 and 100 mg/L cefotaxime sodium. Each BY-2 callus was observed by suspending in
 395 liquid modified LS medium.

396

397 **Plasmids and sequences**

398 All plasmids were generated by standard molecular techniques and listed in
 399 Supplementary Table S1. The *P2A-NLS-mCherry* sequence was amplified from a plasmid
 400 obtained from Dr. Tomokazu Kawashima (Kawashima et al. 2014). The *nasp^{cri}* mutant
 401 was generated by CRISPR/Cas9 system based on pKIR1.1 (Tsutsui and Higashiyama
 402 2017). Gene and protein sequence data can be obtained from The Arabidopsis
 403 Information Resource (TAIR; <http://arabidopsis.org>) for *A. thaliana* and from Phytozome
 404 (<https://phytozome-next.jgi.doe.gov/>) for *C. rubella*, *Z. mays*, and *M. polymorpha*.

405

406 **Confocal microscopy and image analysis**

407 Confocal images of petal cells, ovules, and BY-2 cells of transgenic lines were taken
 408 using an LSM-780-DUO-NLO system (Zeiss). Images were acquired using ZEN 2010
 409 software (Zeiss) and processed using Fiji software (<http://fiji.sc/> Fiji).

410

411 **Establishment of the *nasp^{cri}* mutant line**

412 T1 seeds transformed with HTv985, pKIR1.1 vector containing guide RNA (gRNA)
 413 sequence (5'-GTCTGAATCGGTTGTTTCAGGG-3') that targets the first exon of *NASP*
 414 gene (Fig. 2A), were selected by red fluorescence of seeds. One mutant candidate (#3)
 415 was obtained by genomic PCR using primers [NASP+3F
 416 (5'-GGTTGAAGAATCAGCTTC-3') and NASP+440R
 417 (5'-CATGATTAGATCATTCGAG-3')] and direct Sanger sequencing, which showed a
 418 disrupted sequence around the gRNA target site. From seeds harvested from this
 419 candidate T1 individual, Cas9-free plants were selected based on the absence of red
 420 fluorescence of seeds and confirmed later by no hygromycin resistant seedling in the next
 421 generation. Genomic PCR and sequencing analysis identified the 10-bp deletion in the
 422 *nasp^{cri}/NASP* heterozygous plant (#3-16). The siblings from #3-16 plant were used for
 423 phenotypic analysis and transformation for complementation tests.

424 Based on the 10-bp deletion, NASP+141R (5'-AGTGCCTCCCTGAACAAC-3')
 425 for wild-type *NASP* and *nasp^{cri}* F (5'-CGTCTGAATCGGAGGCA-3') for 10-bp deletion
 426 *nasp^{cri}* allele were designed (Fig. 2B). Primers for wild type allele [NASP-1162F
 427 (5'-AATACTAAGCGAGCCATC-3') and NASP+141R] and primers for *nasp^{cri}* allele
 428 (*nasp^{cri}* F and NASP+440R) were used to exclusively amplify endogenous *NASP* gene
 429 and *nasp^{cri}* allele, respectively.

430

431 **Evaluation of embryo phenotype by clearing in chloral hydrate solution**

432 The carpel walls were removed from siliques five days after hand pollination using
 433 tweezers and 27-gauge needle (Terumo, Tokyo, Japan). The remaining tissues,
 434 ovules/seeds attached to the septum, were fixed and decolorized in ethanol/acetic acid
 435 (9:1) and rehydrated with ethanol series (80%, 60%, 40%, and 20%) and distilled water.

436 After rehydration, the samples were cleared in chloral hydrate solution (chloral hydrate,
437 glycerol, and distilled water in the weight ratio of 8:1:3). Differential interference contrast
438 (DIC) microscopy analysis of cleared ovules/seeds was performed with a Nikon Ts2R
439 microscope.

440

441 **Evaluation of embryo development with fluorescence by ClearSee**

442 Two or four days after hand pollination, siliques were dissected as described above and
443 fixed by immersing in 4% paraformaldehyde/PBS and applying a reduced pressure with a
444 vacuum pump. About 1 hour after the fixation, samples were washed with PBS twice and
445 cleared with ClearSeeAlpha solution to reduce the formation of brown pigment in the
446 ClearSee-treated ovule (Kurihara et al. 2021). 2–3 weeks after treatment in the dark at
447 room temperature, the samples were stained with ClearSee solution (Kurihara et al. 2015)
448 containing 0.1 mg/mL calcofluor white for 1 hour and then washed with ClearSee
449 solution for 1 hour. The silique sample was split at the center of the septum to separate
450 two lines of placenta with seeds into one line of that, which facilitates the observation of
451 embryo phenotype without overlapping each seed tissue. For confocal observation using
452 LSM-780-DUO-NLO system (Zeiss), images were acquired with a 40× objective lens
453 (LD C-Apochromat 40x/1.1 W Corr, Zeiss) under 488/561 nm excitation and
454 490-561/588-695 nm emission wavelengths for mClover and mRuby2 fluorescence and
455 sequentially under 405 nm excitation and 415-490 nm emission wavelength for
456 calcofluor white.

457

458 **BiFC assay**

459 Transient expression in *N. benthamiana* leaves by agroinfiltration was performed as

previously described (Takeuchi and Higashiyama 2016). Equal amounts of *A. tumefaciens* strain GV3101 cultures for nYFP and cYFP-PNC constructs and the p19 silencing suppressor were collected in a tube and resuspended in infiltration buffer (10 mM MgCl₂, 10 mM MES, pH 5.6, and 150 μM acetosyringone). Three to four hours after incubation at 26°C, the mixed suspensions were infiltrated in four to five-week-old *N. benthamiana* leaves. Two days after infiltration, the leaf samples were subjected to confocal microscope observation.

To virtually normalize protein production levels among CENH3 variants fused to cYFP-PNC sequence, the relative BiFC intensity was BiFC signal (YFP fluorescence) divided by NLS-mCherry signal. For each channel of images, background was subtracted and fluorescent intensity of nuclei region were measured. For each BiFC combination, the relative intensity of 9–36 nuclei from 3 replicate samples was calculated and plotted using R studio (<https://cran.ism.ac.jp/>). Statistical analysis was also performed using R studio (“multcomp” package; <https://CRAN.R-project.org/package=multcomp>) based on the Dunnett’s multiple comparison test.

Data Availability

The data underlying this study, including nucleotide and protein sequences, are available in the article and supplementary data.

Funding

This work was supported by the Japan Society for the Promotion of Science on Overseas

484 Research Fellowship (No. 27-601), Grant-in-Aid for Scientific Research on Innovative
485 Areas (18H04834), and the Sumitomo Foundation Grant for Basic Science Research
486 Projects (170663) to H.T.; Grant-in-Aid for Scientific Research on Innovative Areas,
487 Advanced Bioimaging Support (ABiS, 16H06280) to T.H and Nagoya University Live
488 Imaging Center. The Austrian Science Fund supported F.B. (I 2163, P28320).

489

490

491 **Acknowledgments**

492 We thank Simon Chan's laboratory for *GFP-CENH3* in *cenh3-1* seed; Hiroki Tsutsui for
493 pKIR1.1 vector; Tomokazu Kawashima for plasmids and cDNAs used for plasmid
494 construction; Akihisa Osakabe and other members of Berger lab (GMI, Vienna, Austria)
495 for helpful discussion; Daisuke Kurihara for technical assistance of BY-2 experiment;
496 Minako Ueda for technical advice on observation of zygote and early embryo; and
497 Yoshikatsu Sato and Nagoya University Live Imaging Center for Zeiss
498 LSM-780-DUO-NLO microscope system.

499

500

501 **Author Contributions**

502 H.T. conceived and designed this study under supervision of F.B.; H.T. generated the
503 research materials; H.T. and S.N. performed the experiments and analyzed the data; H.T.
504 wrote the manuscript with input from S.N., T.H., and F.B.

505

506

507 **Disclosures**

508 Conflicts of interest: No conflicts of interest declared.

509

510

511 **References**

512 Bao, H., Carraro, M., Flury, V., Liu, Y., Luo, M., Chen, L., et al. (2022) NASP maintains
513 histone H3-H4 homeostasis through two distinct H3 binding modes. *Nucleic Acids*
514 *Res.* 50: 5349–5368.

515 Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., et
516 al. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a
517 functional de novo kinetochore. *J Cell Biol.* 194: 229–243.

518 Bassett, E.A., DeNizio, J., Barnhart-Dailey, M.C., Panchenko, T., Sekulic, N., Rogers,
519 D.J., et al. (2012) HJURP Uses Distinct CENP-A Surfaces to Recognize and to
520 Stabilize CENP-A/Histone H4 for Centromere Assembly. *Dev Cell.* 22: 749–762.

521 Black, B.E., Jansen, L.E.T., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J. V., et al.
522 (2007) Centromere Identity Maintained by Nucleosomes Assembled with Histone
523 H3 Containing the CENP-A Targeting Domain. *Mol Cell.* 25: 309–322.

524 Borg, M., Jiang, D., and Berger, F. (2021) Histone variants take center stage in shaping
525 the epigenome. *Curr Opin Plant Biol.* 61.

526 Bowman, A., Koide, A., Goodman, J.S., Colling, M.E., Zinne, D., Koide, S., et al. (2017)
527 sNASP and ASF1A function through both competitive and compatible modes of
528 histone binding. *Nucleic Acids Res.* 45: 643–656.

529 Bowman, A., Lercher, L., Singh, H.R., Zinne, D., Timinszky, G., Carlomagno, T., et al.
530 (2015) The histone chaperone sNASP binds a conserved peptide motif within the

531 globular core of histone H3 through its TPR repeats. *Nucleic Acids Res.* 44:
532 3105–3117.

533 Camahort, R., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., and Gerton, J.L.
534 (2007) Scm3 Is Essential to Recruit the Histone H3 Variant Cse4 to Centromeres
535 and to Maintain a Functional Kinetochores. *Mol Cell.* 26: 853–865.

536 Campos, E.I., Fillingham, J., Li, G., Zheng, H., Voigt, P., Kuo, W.H.W., et al. (2010) The
537 program for processing newly synthesized histones H3.1 and H4. *Nat Struct Mol*
538 *Biol.* 17: 1343–1351.

539 Chen, C.C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., et al.
540 (2014) CAL1 is the Drosophila CENP-A assembly factor. *J Cell Biol.* 204: 313–329.

541 Cho, U.S., and Harrison, S.C. (2011) Recognition of the centromere-specific histone
542 Cse4 by the chaperone Scm3. *Proc Natl Acad Sci U S A.* 108: 9367–9371.

543 Cook, A.J.L., Gurard-Levin, Z.A., Vassias, I., and Almouzni, G. (2011) A Specific
544 Function for the Histone Chaperone NASP to Fine-Tune a Reservoir of Soluble
545 H3-H4 in the Histone Supply Chain. *Mol Cell.* 44: 918–927.

546 Dunleavy, E.M., Pidoux, A.L., Monet, M., Bonilla, C., Richardson, W., Hamilton, G.L.,
547 et al. (2007) A NASP (N1/N2)-Related Protein, Sim3, Binds CENP-A and Is
548 Required for Its Deposition at Fission Yeast Centromeres. *Mol Cell.* 28: 1029–1044.

549 Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., et al.
550 (2009) HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of
551 CENP-A at Centromeres. *Cell.* 137: 485–497.

552 Filipescu, D., Szenker, E., and Almouzni, G. (2013) Developmental roles of histone H3
553 variants and their chaperones. *Trends Genet.* 29: 630–640.

554 Finn, R.M., Ellard, K., Eirín-López, J.M., and Ausió, J. (2012) Vertebrate nucleoplasmin

555 and NASP: Egg histone storage proteins with multiple chaperone activities. *FASEB*
556 *J.* 26: 4788–4804.

557 Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., et al.
558 (2009) Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by
559 HJURP. *Cell.* 137: 472–484.

560 Le Goff, S., Keçeli, B.N., Jeřábková, H., Heckmann, S., Rutten, T., Cotterell, S., et al.
561 (2020) The H3 histone chaperone NASP SIM3 escorts CenH3 in Arabidopsis. *Plant*
562 *J.* 101: 71–86.

563 Henikoff, S., and Dalal, Y. (2005) Centromeric chromatin: What makes it unique? *Curr*
564 *Opin Genet Dev.* 15: 177–184.

565 Hirsch, C.D., Wu, Y., Yan, H., and Jiang, J. (2009) Lineage-specific adaptive evolution
566 of the centromeric protein cenH3 in diploid and allotetraploid oryza species. *Mol*
567 *Biol Evol.* 26: 2877–2885.

568 Ingouff, M., Rademacher, S., Holec, S., Šoljić, L., Xin, N., Readshaw, A., et al. (2010)
569 Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic
570 reprogramming in arabidopsis. *Curr Biol.* 20: 2137–2143.

571 Jiang, D., and Berger, F. (2017) Histone variants in plant transcriptional regulation.
572 *Biochim Biophys Acta - Gene Regul Mech.* 1860: 123–130.

573 Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., et
574 al. (2015) Point mutation impairs centromeric CENH3 loading and induces haploid
575 plants. *Proc Natl Acad Sci U S A.* 112: 11211–11216.

576 Kawashima, T., Lorković, Z.J., Nishihama, R., Ishizaki, K., Axelsson, E., Yelagandula,
577 R., et al. (2015) Diversification of histone H2A variants during plant evolution.
578 *Trends Plant Sci.* 20: 419–425.

579 Kawashima, T., Maruyama, D., Shagirov, M., Li, J., Hamamura, Y., Yelagandula, R., et
580 al. (2014) Dynamic F-actin movement is essential for fertilization in *Arabidopsis*
581 *thaliana*. *Elife*. 3: 1–18.

582 Khorasanizadeh, S. (2004) The Nucleosome. *Cell*. 116: 259–272.

583 Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., et al.
584 (2020) A variety of changes, including CRISPR/Cas9-mediated deletions, in
585 CENH3 lead to haploid induction on outcrossing. *Plant Biotechnol J*. 18:
586 2068–2080.

587 Kurihara, D., Mizuta, Y., Nagahara, S., and Higashiyama, T. (2021) ClearSeeAlpha:
588 Advanced Optical Clearing for Whole-Plant Imaging. *Plant Cell Physiol*. 62:
589 1302–1310.

590 Kurihara, D., Mizuta, Y., Sato, Y., and Higashiyama, T. (2015) ClearSee: A rapid optical
591 clearing reagent for whole-plant fluorescence imaging. *Dev*. 142: 4168–4179.

592 Lermontova, I., Schubert, V., Fuchs, J., Klatter, S., Macas, J., and Schubert, I. (2006)
593 Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G2 and
594 requires the presence of the histone fold domain. *Plant Cell*. 18: 2443–2451.

595 Liu, C.P., Jin, W., Hu, J., Wang, M., Chen, J., Li, G., et al. (2021) Distinct histone H3-H4
596 binding modes of sNASP reveal the basis for cooperation and competition of histone
597 chaperones. *Genes Dev*. 35: 1610–1624.

598 Liu, Y., Chen, L., Wang, N., Wu, B., Bao, H., and Huang, H. (2022) Structural basis for
599 histone H3 recognition by NASP in *Arabidopsis*. *J Integr Plant Biol*.

600 Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., et al. (2017) Systematic
601 comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep*.
602 7: 1–9.

603 Luger, K., Mä Der, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997)
604 Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of
605 the nucleosome core particle at 2.8 Å resolution. *Nature*. 389: 251–260.

606 Maheshwari, S., Ishii, T., Brown, C.T., Houben, A., and Comai, L. (2017) Centromere
607 location in Arabidopsis is unaltered by extreme divergence in CENH3 protein
608 sequence. *Genome Res*. 27: 471–478.

609 Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L.
610 (2015) Naturally Occurring Differences in CENH3 Affect Chromosome
611 Segregation in Zygotic Mitosis of Hybrids. *PLoS Genet*. 11: 1–20.

612 Maksimov, V., Nakamura, M., Wildhaber, T., Nanni, P., Ramström, M., Bergquist, J., et
613 al. (2016) The H3 chaperone function of NASP is conserved in Arabidopsis. *Plant J*.
614 88: 425–436.

615 Malik, H.S., and Henikoff, S. (2009) Major Evolutionary Transitions in Centromere
616 Complexity. *Cell*. 138: 1067–1082.

617 Malik, H.S., Vermaak, D., and Henikoff, S. (2002) Recurrent evolution of DNA-binding
618 motifs in the Drosophila centromeric histone. *Proc Natl Acad Sci U S A*. 99:
619 1449–1454.

620 Marimuthu, M.P.A., Maruthachalam, R., Bondada, R., Kuppu, S., Tan, E.H., Britt, A., et
621 al. (2021) Epigenetically mismatched parental centromeres trigger genome
622 elimination in hybrids. *Sci Adv*. 7: 1–19.

623 Mayo, K.J., Gonzales, B.J., Mason, H.S. (2006) Genetic transformation of tobacco NT1
624 cells with *Agrobacterium tumefaciens*. *Nat Protoc*. 1: 1105–1111.

625 Medina Pritchard, B., Lazou, V., Zou, J., Byron, O., Abad, M.A., Rappsilber, J., et al.
626 (2020) Structural basis for centromere maintenance by Drosophila CENP A

627 chaperone CAL 1 . *EMBO J.* 39: 1–21.

628 Min, Y., Frost, J.M., and Choi, Y. (2019) Nuclear Chaperone ASF1 is Required for
629 Gametogenesis in *Arabidopsis thaliana*. *Sci Rep.* 9: 13959.

630 Morris, C.A., and Moazed, D. (2007) Centromere Assembly and Propagation. *Cell.* 128:
631 647–650.

632 Nagaki, K., Terada, K., Wakimoto, M., Kashiwara, K., and Murata, M. (2010)
633 Centromere targeting of alien CENH3s in *Arabidopsis* and tobacco cells.
634 *Chromosom Res.* 18: 203–211.

635 Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., et
636 al. (2021) The genetic and epigenetic landscape of the *Arabidopsis* centromeres.
637 *Science.* 374: eabi7489.

638 Osakabe, A., Tachiwana, H., Matsunaga, T., Shiga, T., Nozawa, R.S., Obuse, C., et al.
639 (2010) Nucleosome formation activity of human somatic Nuclear Autoantigenic
640 Sperm Protein (sNASP). *J Biol Chem.* 285: 11913–11921.

641 Palladino, J., Chavan, A., Sposato, A., Mason, T.D., and Mellone, B.G. (2020) Targeted
642 De Novo Centromere Formation in *Drosophila* Reveals Plasticity and Maintenance
643 Potential of CENP-A Chromatin. *Dev Cell.* 52: 379-394.e7.

644 Phansalkar, R., Lapierre, P., and Mellone, B.G. (2012) Evolutionary insights into the role
645 of the essential centromere protein CAL1 in *Drosophila*. *Chromosom Res.* 20:
646 493–504.

647 Pusarla, R.H., and Bhargava, P. (2005) Histones in functional diversification: Core
648 histone variants. *FEBS J.* 272: 5149–5168.

649 Ravi, M., and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated
650 genome elimination. *Nature.* 464: 615–618.

651 Ravi, M., Kwong, P.N., Menorca, R.M.G., Valencia, J.T., Ramahi, J.S., Stewart, J.L., et
652 al. (2010) The rapidly evolving centromere-specific histone has stringent functional
653 requirements in *Arabidopsis thaliana*. *Genetics*. 186: 461–471.

654 Ravi, M., Marimuthu, M.P.A., Tan, E.H., Maheshwari, S., Henry, I.M., Marin-Rodriguez,
655 B., et al. (2014) A haploid genetics toolbox for *Arabidopsis thaliana*. *Nat Commun*.
656 5: 1–8.

657 Richardson, R.T., Alekseev, O.M., Grossman, G., Widgren, E.E., Thresher, R., Wagner,
658 E.J., et al. (2006) Nuclear autoantigenic sperm protein (NASP), a linker histone
659 chaperone that is required for cell proliferation. *J Biol Chem*. 281: 21526–21534.

660 Richardson, R.T., Batova, I.N., Widgren, E.E., Zheng, L.X., Whitfield, M., Marzluff,
661 W.F., et al. (2000) Characterization of the histone H1-binding protein, NASP, as a
662 cell cycle-regulated somatic protein. *J Biol Chem*. 275: 30378–30386.

663 Rosin, L., and Mellone, B.G. (2016) Co-evolving CENP-A and CAL1 Domains Mediate
664 Centromeric CENP-A Deposition across *Drosophila* Species. *Dev Cell*. 37:
665 136–147.

666 Rosin, L.F., and Mellone, B.G. (2017) Centromeres Drive a Hard Bargain. *Trends Genet*.
667 33: 101–117.

668 Roure, V., Medina-Pritchard, B., Lazou, V., Rago, L., Anselm, E., Venegas, D., et al.
669 (2019) Reconstituting *Drosophila* Centromere Identity in Human Cells. *Cell Rep*.
670 29: 464-479.e5.

671 Sanei, M., Pickering, R., Kumke, K., Nasuda, S., and Houben, A. (2011) Loss of
672 centromeric histone H3 (CENH3) from centromeres precedes uniparental
673 chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci U S A*.
674 108: E498–505.

675 She, W., Grimanelli, D., Rutowicz, K., Whitehead, M.W.J., Puzio, M., Kotliński, M., et
676 al. (2013) Chromatin reprogramming during the somatic-to-reproductive cell fate
677 transition in plants. *Dev.* 140: 4008–4019.

678 Stoler, S., Rogers, K., Weitze, S., Morey, L., Fitzgerald-Hayes, M., and Baker, R.E.
679 (2007) Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for
680 G2/M progression and Cse4 localization. *Proc Natl Acad Sci U S A.* 104:
681 10571–10576.

682 Takeuchi, H., and Higashiyama, T. (2016) Tip-localized receptors control pollen tube
683 growth and LURE sensing in *Arabidopsis*. *Nature.* 531: 245–248.

684 Talbert, P.B., and Henikoff, S. (2010) Histone variants ancient wrap artists of the
685 epigenome. *Nat Rev Mol Cell Biol.* 11: 264–275.

686 Talbert, P.B., and Henikoff, S. (2017) Histone variants on the move: Substrates for
687 chromatin dynamics. *Nat Rev Mol Cell Biol.* 18: 115–126.

688 Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., and Henikoff, S. (2002) Centromeric
689 localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell.*
690 14: 1053–1066.

691 Tsutsui, H., and Higashiyama, T. (2017) pKAMA-ITACHI vectors for highly efficient
692 CRISPR/Cas9-mediated gene knockout in *Arabidopsis thaliana*. *Plant Cell Physiol.*
693 58: 46–56.

694 Yuan, J., Guo, X., Hu, J., Lv, Z., and Han, F. (2015) Characterization of two CENH3
695 genes and their roles in wheat evolution. *New Phytol.* 206: 839–851.

696 Zhang, W., Lee, H.R., Koo, D.H., and Jiang, J. (2008) Epigenetic modification of
697 centromeric chromatin: Hypomethylation of DNA sequences in the
698 CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell.* 20:

699 25–34.

700 Zhong, Z., Wang, Y., Wang, M., Yang, F., Thomas, Q.A., Xue, Y., et al. (2022) Histone
701 chaperone ASF1 mediates H3.3-H4 deposition in Arabidopsis. *Nat Commun.* 13.

702 Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., et al. (2011) Structural
703 basis for recognition of centromere histone variant CenH3 by the chaperone Scm3.
704 *Nature.* 472: 234–238.

705 Zhu, Y., Weng, M., Yang, Y., Zhang, C., Li, Z., Shen, W.H., et al. (2011) Arabidopsis
706 homologues of the histone chaperone ASF1 are crucial for chromatin replication and
707 cell proliferation in plant development. *Plant J.* 66: 443–455.

708

709

710 **Legends to Figures**

711 **Fig. 1.** CENH3 protein regions required for targeting *A. thaliana* centromere. (A)
712 Sequence alignment of histone-fold domain of *H. sapiens*, *A. thaliana*, *Z. mays*, and *M.*
713 *polymorpha* CENH3 proteins. Protein secondary structure and blocks dividedly analyzed
714 in this study are shown below the alignment. See also Supplementary Figure S1A and B
715 for sequence alignments of the canonical H3 and N-terminal tail of CENH3. (B)
716 Schematic of the expression construct to produce CENH3 variants fused to mClover and
717 NLS-mCherry from a single transcript. Co-translational “self-cleavage” during
718 translation of the 2A peptide sequence (*P2A*) leads to generate two separate proteins.
719 (C-H) Localization assessment of CENH3 variants in the *A. thaliana* petal cell. Block
720 bars drawn above each image and titles guide the protein structure of each variant with
721 colors. Each figure panel consisting of multiple variants represents comparisons which
722 examine canonical H3 (H3.3), CENH3 of *A. thaliana* (AtCENH3), *Z. mays* (ZmCENH3),

723 and *M. polymorpha* (MpCENH3), and AtCENH3 substituted with HAVL
724 (AtCENH3^{HAVL}) (C); necessary single AtCENH3 region (D); importance of N-terminal
725 tail (E) and CATD (F); sufficient AtCENH3 region (G); and necessary amino acid of
726 AtCENH3 (H). The images are representative nuclei from more than three images
727 capturing multiple nuclei. See also Supplementary Figure 2 and 3 for summary of this
728 assay and representative images of similar assay using tobacco BY-2 cells.

729

730 **Fig. 2.** Embryo lethal phenotype of *nasp^{cri}* knockout mutant in *GCH3* line. (A) Schematic
731 of *NASP* gene structure with a guide RNA (gRNA) target site and genotyping primers.
732 Gray boxes are exons of *NASP* gene. The bold line indicates a region used for
733 complementation constructs. (B) Primer design to exclusively amplify endogenous *NASP*
734 gene and *nasp^{cri}* mutant allele. (C) Representative images of developing seeds in pistils of
735 a *GCH3* background line (*GFP-CENH3* in *cenh3-1*), a heterozygous *nasp^{cri}/NASP* in
736 *GCH3*, and a complemented line with *NASP-mRuby2* transgene. (D) The percentage of
737 abnormal seeds per developing seeds in self-pollinated siliques. For T1 lines, siblings
738 from a *nasp^{cri}/NASP GCH3* plant were used for the introduction of each orthologous
739 *NASP* gene fused to *mRuby2*, and T1 plants with *nasp^{cri}/NASP* genotype were selected by
740 genomic PCR. Each dot represents the sum value of five siliques for each independent T1
741 plant. (E) Embryo development in ovules from *GCH3* or *nasp^{cri}/NASP GCH3* five days
742 after pollination. A heart-stage embryo in *GCH* ovule and an early arrested embryo at
743 1-cell stage in self-pollinated *nasp^{cri}/NASP GCH3* ovule. (F) The proportion of embryo
744 phenotype. The phenotype was categorized based on observation as follows: developed
745 (normal), about heart-stage embryo; developed (abnormal), globular to heart-stage
746 embryo with distorted shape; arrested, 1- or 2-cell stage embryo; unfertilized,

747 non-elongated egg cell or no egg cell structure due to degeneration; ND, not defined.
748 209–321 ovules/seeds from 4 to 6 siliques were analyzed for each genotype.

749

750 **Fig. 3.** Requirement of NASP for early embryogenesis and *de novo* CENH3 deposition in
751 *GCH* line. (A, B) Confocal images of ClearSee-treated ovules inside the pistil of
752 homozygous *nasp^{cri} GCH3* with hemizygous *NASPpro::NASP-mRuby2* transgene at
753 2 days after pollination (A) and 4 days after pollination (B). The presence (+) or absence
754 (–) of mRuby2 red fluorescence (shown as magenta) in embryo nuclei indicate
755 NASP-complemented embryo or homozygous *nasp^{cri}* mutant embryo, respectively. The
756 GFP-CENH3 signals at centromeres were observed in the embryo and endosperm of
757 NASP-mRuby2-positive (+) ovules and in the maternal integuments of ovules regardless
758 of NASP-mRuby2 existence in embryo cells (+ and –). Note that asterisks indicate
759 unrelated objectives with autofluorescence in ovular integument cells. Green,
760 GFP-CENH3; cyan, cell walls stained with calcofluor white. (C) The number of ovules
761 with each stage of zygote/embryo from three independent hemizygous lines and one
762 homozygous line for *NASPpro::NASP-mRuby2* transgene in homozygous *nasp^{cri} GCH3*.
763 Counting was conducted using confocal images of ClearSee-treated ovules ($n = 54–101$)
764 from 3 siliques at 2 days after pollination.

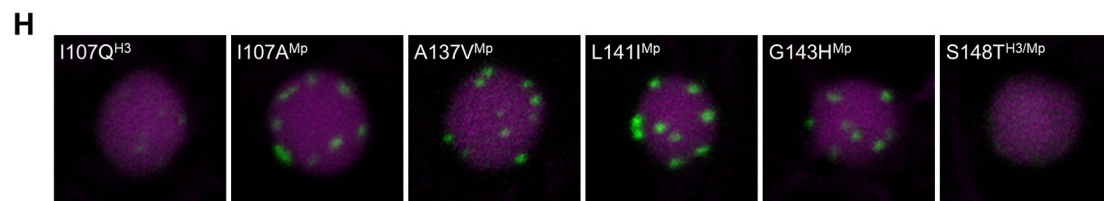
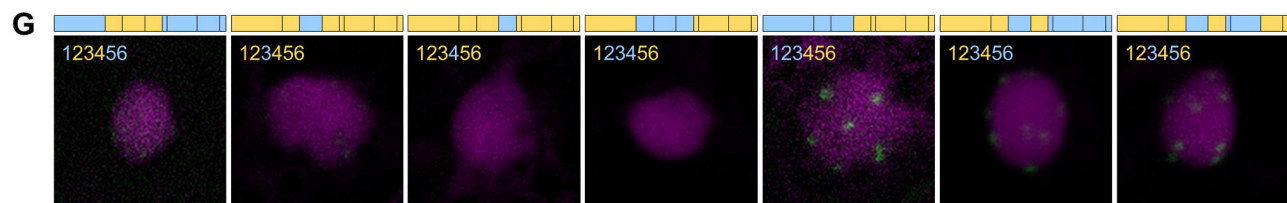
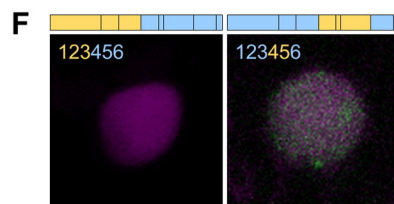
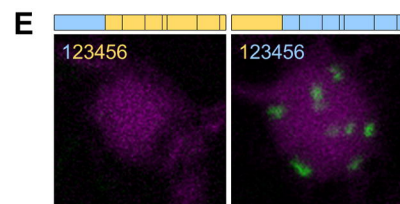
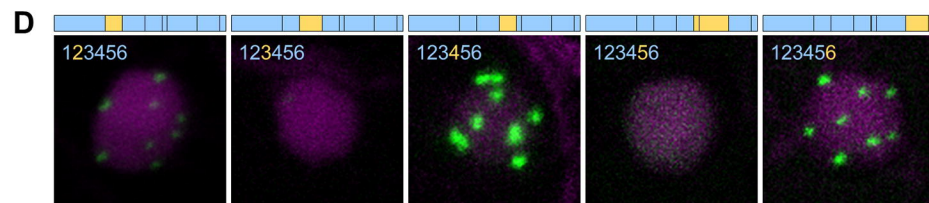
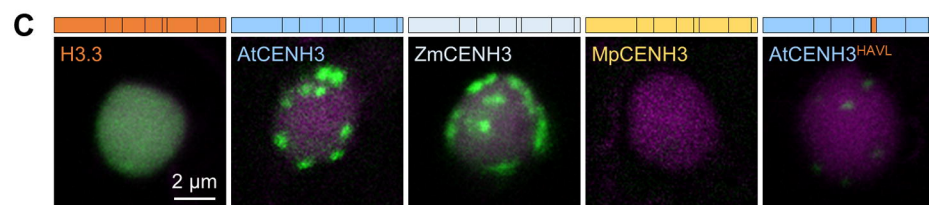
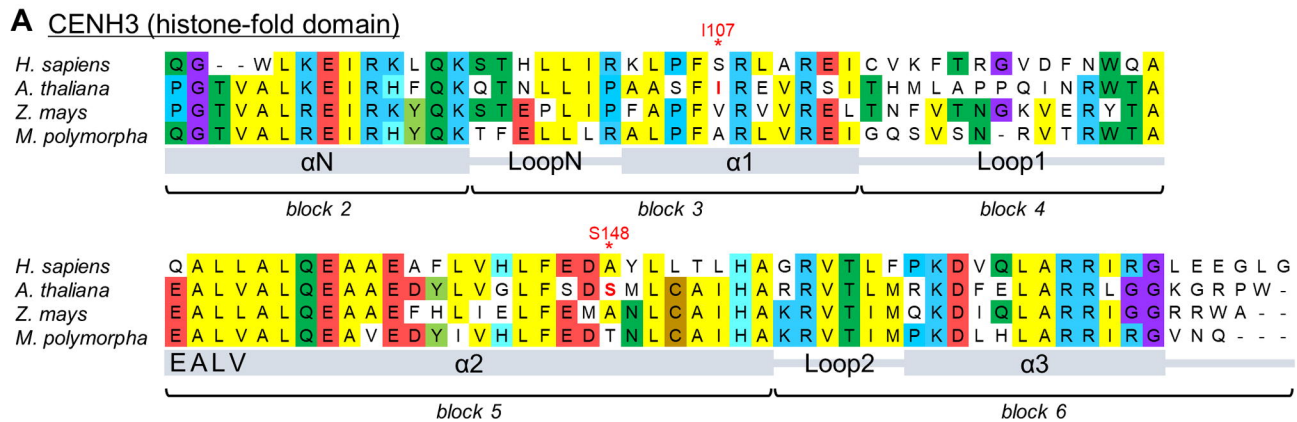
765

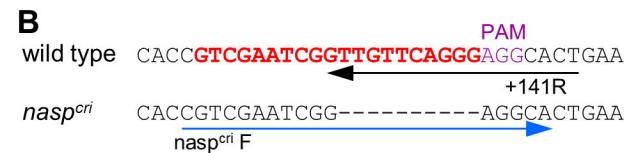
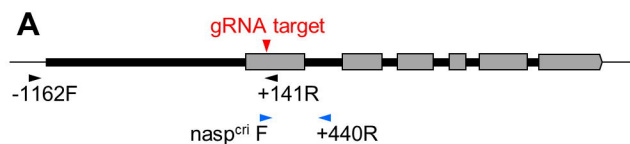
766 **Fig. 4.** NASP protein domains required for the function in *A. thaliana* embryo
767 development. (A) Sequence alignment of NASP proteins of *H. sapiens* (HsNASP), *A.*
768 *thaliana* (AtNASP), *Z. mays* (ZmNASP), and *M. polymorpha* (MpNASP). Red squares
769 and triangles below the alignment indicate amino acids of AtNASP that has been shown
770 to mediate binding to canonical H3's N-terminal and $\alpha 3$ regions, respectively (Liu et al.

2022). See also Supplementary Figure S1C and D for sequence alignments of these binding regions of H3 in comparison with AtCENH3 and MpCENH3. (B) The percentage of abnormal seeds per developing seeds in self-pollinated siliques from T1 lines. As same as Fig. 2D, siblings from a *nasp^{cri}/NASP GCH3* plant were used for the introduction of each *NASP* variant fused to *mTurquoise2*, and T1 plants with *nasp^{cri}/NASP* genotype were selected by genomic PCR. Each dot represents the sum value of five siliques for each independent T1 plant. (C) The proportion of embryo phenotype. The phenotype was categorized as noted for Fig. 2F. 101–246 ovules/seeds from 3 to 7 siliques were analyzed for each genotype.

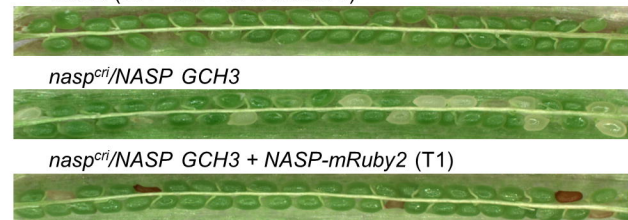
780

Fig. 5. Interaction assay for CENH3 variants and AtASF1A or AtNASP using a BiFC system with P2A-NLS-mCherry (PNC) module. (A) Confocal images of *N. benthamiana* leaf epidermal cells of BiFC assay between AtASF1A-nYFP or AtNASP-nYFP and cYFP-PNC (–), AtCENH3-cYFP-PNC (At), or MpCENH3-cYFP-PNC (Mp). Equivalent NLS-mCherry signal implies comparable levels of the three cYFP fusion proteins. Nuclei with negative or positive BiFC signals were marked with x signs or arrowheads, respectively. Scale bars, 20 μ m. (B, C) The relative BiFC intensity (BiFC/mCherry) for H3.3, AtCENH3, MpCENH3, ZmCENH3, or At/Mp-chimeric CENH3 variants with AtASF1A (B) and AtNASP (C). Each dot represents the value of each nucleus. Red lines show mean of the values. *P* values shown in gray are estimated by the Dunnett’s test compared with negative control (–). For AtCENH3, MpCENH3, and At/Mp-chimeric CENH3 variants, block bars on the right indicate the structure of each variant with colors (light blue, AtCENH3; yellow, MpCENH3), and red asterisks indicate CENH3 variants showing significant BiFC signal.

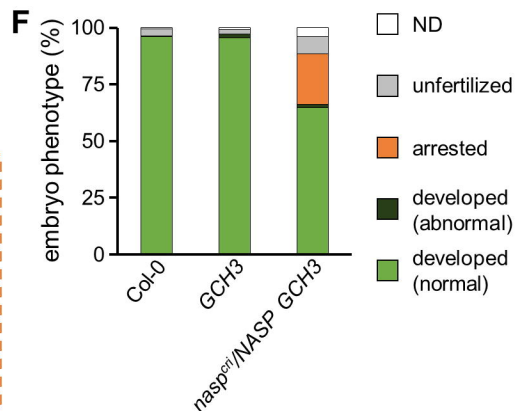
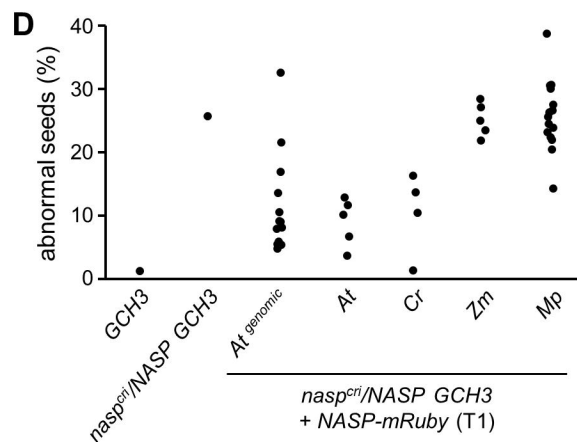
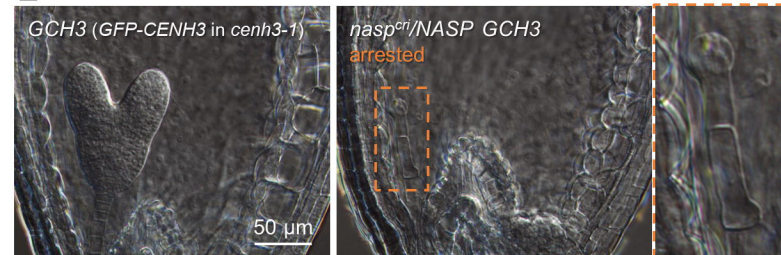


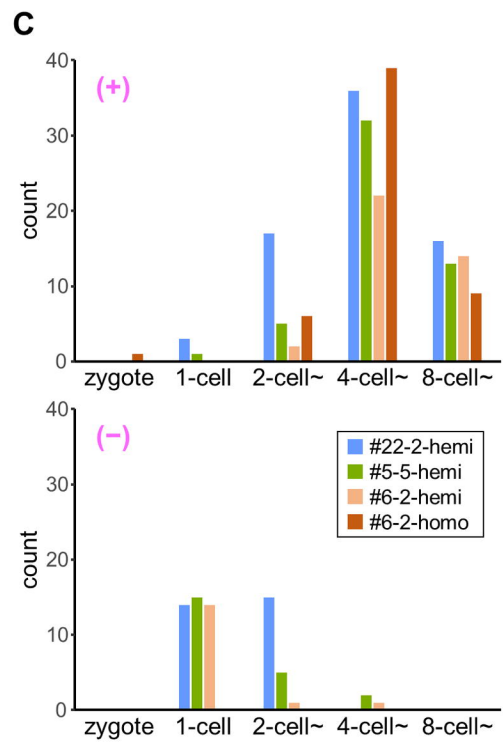
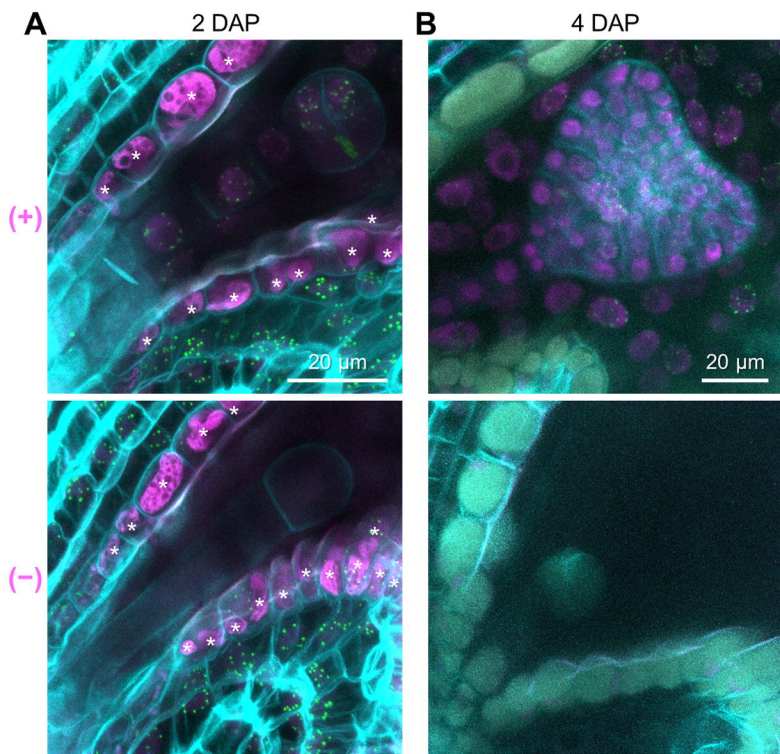


C *GCH3* (GFP-CENH3 in *cenh3-1*)



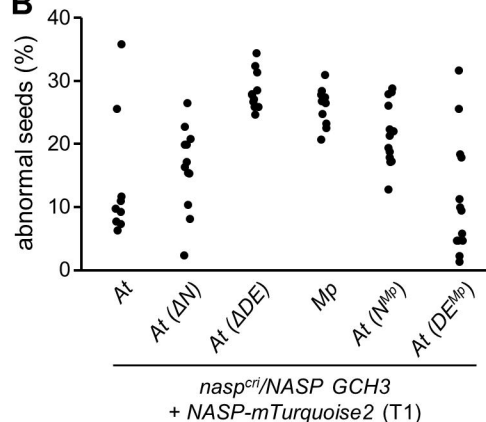
E





A

		N-terminal region (N)		TPR1	
HsNASP	1	-----MAMESTATAVAEALVSADKIEDVPAPSTSADKVESLDVDSEAKKLGLGKHKHVMGDIPAAV		63	
AtNASP	1	MVEESASASEASVVIQTLTEPATEIAQTLLEPNLASIEATVESVVGQGTSTCNNDANNNNNAADSAAETVCDDEEREKTLFAEELTEKGSVFLEKENDFAEAV		100	
CrNASP	1	MVEEAAASASEASVTEAVPTLTLEPVIQTLLEPNQASIEATVESAVQGGTSTCNNN--NNNAADSAAITDVCVEREKTIEFADBLTEKGSVFLEKQDFGEAV		98	
ZmNASP	1	-----MASSSENAAVPEEEPQAPPDPDTEPTTEAAPGEADEPQTLERAQBLFDRGSKAIEDDEFDVAV		64	
MpNASP	1	-----MPAMAIVPDADYKDEEQVEEQALPLPLFISQRLYDLGCAAISSDSYEEAT		53	
		TPR2-1		acidic region (DE)	
HsNASP	64	NAFQEAASLLGKKYGETANECEGAFFFYGKSLLEIARMENGVLGNALEGVHVHEEEGEKTEDESLVENNDNI DETEGSSE-----DDKENDKT		151	
AtNASP	101	DCFSRALEIRVAHYGELDAECINAYRYGLALLAKAQAEADPTGNMPPKKEGEVQOESS-----NGESLAPSVVSGDPERQGSSSGQEGSGGKDG		190	
CrNASP	99	DCFSRALEIRVEHYGELDAECVNAYRYGSALLAKAQAEADPTGNVPPKKEGEVQOESSSKDDSVKNTVNGESLAAVSVSSNPERQGSSSGQEGAGGIEQG		198	
ZmNASP	65	DCLSRALAIRASHYGELAECASTYFRYGCALLYKAQESDVLGNVPKSLPNEDSVKSTATKDDSGTSKVSQDADAASSEKVDAAEEGHNSNGKNQENG		164	
MpNASP	54	DCFSKALEIRVNHYGELAECASTYKYGCALLYKVQAESEALCDNRNEEPPAHKPSKEEGLSALAGGSK-----GPVDDDEEFFDDE		135	
		TPR2-2		TPR3	
HsNASP	152	EEMPNDSVLENKLSQENEFEEIGNLELAWDMLDIAKTIIFKRQETKFAQLYAAQAHKLKKEVSVSESNVYQAVEEFQSCNLCQEQYLEAHIRLLAETHYQL		251	
AtNASP	191	EDGEDCQDDLLSDADGDADEDESLLDMAWKMLDIARVITTDKQSTETMEK--VDILCSLAEVSLERDDIESSLSDYKNAISILERLVEPDSRRTAEINFRI		288	
CrNASP	199	EDGENCHDDLLSDADGD--EDDSLLDMAWKMLDIARAITTDKQADTMVK--VDILCAIAEISIERDDIESSLSDYKKAISILERLVEPDSRHTAEINFRI		294	
ZmNASP	165	NGEVEKDDDDVDDDEKMGFEEDNDLLSWKMLDIARAIIVEKSDNIMEK--VKIYSALAEVATREEDMDNSLDYMKALSMLEHLVEPDHRRVVELNFRI		262	
MpNASP	136	GDGKSEDDGEDNDGEVDGEEEGDIELAWKILETARVTHSKQPARSLDE--VDVITLADISLEKENFKTCYADYEQALQILQDLVEPDSRVIASTWYKM		233	
		TPR4			
HsNASP	252	GLAYGYNQYDEAVAQFSKISIEVIENRMMAVINEQVKEAEGSSAEYKKEIEELKELLPEIREKIEDAKESQSRSGNVAELALKATLVESSTSGFTPGGGGSS		351	
AtNASP	289	CICLETGCQKEAIPYCCAKALLICKARMERT--SNEIKGASGATSSTVSEIDEIGIQSSNPYIDKSASDKVEIGDLAGLAEDLEKKLEDLKKQAENP		386	
CrNASP	295	CICLETGCQKEAMPYCCAKALLICKARMERT--SNEIKGASGATSSTVSEIDEIGIQSSNPYIDKSTSDKEAEIEVMAGLAEDLEKKLEDLKKQAENP		392	
ZmNASP	263	CLVYELVSKIGEALISYCAKASISLCKSRISQSI-----KSSKDALLAGIDGASAAAAEAGGSEKSAIEKELE--QLTSILPDLEKKLEDLSE--ANP		348	
MpNASP	234	ALAQCFDNRFECALEACVKAIVSICEMRCQRITKEQCELLKAEQGGSGNDGEMKISSEEAENVGETVYEDLVKAVEDLKFARVDAEIKEIEESVLVDL		333	
				predicted NLS	
HsNASP	352	VSMIASRKPTDGASSNCVTDISHLVRRKKRKEEESPRKDDAKAKQKEPEVNGGSGDAVPSGNEVSENMEEEAENQAESRAAEGTVEAGATVESTAC--		449	
AtNASP	387	KQVLAEELMGVMVSAKPN--ASDKVVPAEAEMSSSRMGTNTNFGKDLSEPTVSTAHTGAA--GGGAASGVTHLGVVGRGVRRVLMNTTISSEASAKKPALEF		483	
CrNASP	393	KQVLAEELMGVMVSAKASEKASEEAVPAEAEMSSSRMGTANTNFGKDLSEPTVSTAHTGAAAGGASSGVTHLGVVGRGVRRVLMNTTISSEASAKKPAITGP		492	
ZmNASP	349	SADMDEMVKAIIVSRV---TDVMPKAASFSSQMAITSSNGF---DSSVMSTAATGGS---SGSTVTDLVGVGRGVKRASIKPISAEPAKKPALDPS		435	
MpNASP	334	REKVEELKQVIVLPSLLEALKAEENPDAAITMSEAVSAVFKRISRVEGTSSMGETVAAFP--GSSSPSEPVTHLGIIVGRGIRKATPIPIVEKELLPDPVQPAK		432	
HsNASP	449	-----		449	
AtNASP	484	SDKADGNSS-----		492	
CrNASP	493	SDKADGNSS-----		501	
ZmNASP	436	SQQGDSSINSEVVPATQTGDESVSK		460	
MpNASP	433	RSAEDILTTGGSGSETQIGFGGAS--		455	

B**C**