1 *Title*: Structural diversity and distribution of NMCP-class nuclear lamina proteins in

2 streptophytic algae

3

4 *Authors and Affiliations*:

5 Brendan S. Kosztyo[1], and Eric J. Richards[2*]

6 [1] Cornell University, College of Agriculture and Life Sciences, Ithaca, NY 14853

7 [2] Boyce Thompson Institute, Ithaca, NY 14853

8

9 **Author for Correspondence*:

10 Eric J. Richards

11 Boyce Thompson Institute

12 533 Tower Road

13 Ithaca, NY 14853

14 USA

15 Ph: 607-254-1234

**ABSTRACT**

Nuclear Matrix Constituent Proteins (NMCPs) in plants function like animal lamins, providing the structural foundation of the nuclear lamina and regulating nuclear organization and morphology. Although they are well-characterized in angiosperms, the presence and structure of NMCPs in more distantly related species, such as streptophytic algae, are relatively unknown. The rapid evolution of NMCPs throughout the plant lineage has caused a divergence in protein sequence that makes similarity-based searches less effective. Structural features are more likely to be conserved compared to primary amino acid sequence; therefore, we developed a filtration protocol to search for diverged NMCPs based on four physical characteristics: intrinsically disordered content, isoelectric point, number of amino acids, and the presence of a central coiled-coil domain. By setting parameters to recognize the properties of bona fide NMCP proteins in angiosperms, we filtered seven complete proteomes from streptophytic algae species and identified strong NMCP candidates in five taxa in the Classes Zygnematophyceae, Charophyceae, and Klebsormidophyceae. Through analysis of these proteins, we observed structural variance in domain size between NMCPs in algae and land plants, as well as a single block of amino acid conservation. Our analysis indicates that NMCPs are absent in the Mesostigmatophyceae. The presence versus absence of NMCP proteins in different streptophytic algae allows us to draw inferences about the evolution of NMCP proteins and their relationship to different forms of mitosis throughout the plant lineage.

## SIGNIFICANCE STATEMENT

All eukaryotic organisms contain a membrane-bound nucleus, which holds a cell's DNA, but plants and animals have distinct sets of proteins that make up the scaffolding essential for nuclear integrity and organization – a structure called the nuclear lamina. The major nuclear lamina proteins in plants and animals share structural features, and we exploited these characteristics to identify a key class of nuclear lamina proteins in streptophytic algae, allowing us to chart the distribution of these proteins across photosynthetic organisms and to gain insight into the evolution of nuclear organization and dynamics during cell division. Our results indicate that the major class of nuclear lamina protein in plants evolved independently from structurally similar nuclear lamina proteins in animals.


## KEYWORDS

- Streptophyte
- Open mitosis
- Nuclear organization
- Proteome filtration
- Coiled-coil domain
- Intrinsically disordered domain

54 **INTRODUCTION**

55 In eukaryotes, the nucleus is central to the integration and execution of cell functions, including
56 storing the genetic material and regulating its expression. A complex of proteins called the
57 nuclear lamina (NL), which resides directly below the inner nuclear membrane, is responsible for
58 the organization of the nucleus and its indirect attachment to the cytoskeleton, as well as for
59 genome and transcription factor regulation (Gruenbaum and Foisner, 2015). In metazoans,
60 lamins are the major structural protein of the NL and the founding member of the intermediate
61 filament superfamily (Dechat *et al.*, 2010).

62

63 Despite being ancestral and well conserved, lamins are not found constitutively throughout the
64 eukaryotic lineage. The absence of lamin homologs could be explained by an extreme divergence
65 in certain eukaryotic lineages (Dittmer and Misteli, 2011). Alternatively, the NL could have
66 independent origins, being composed of distinct proteins recruited to form the NL through
67 convergent evolution (Koreny and Field, 2016). Distinguishing between these two hypotheses
68 requires assembly of a comprehensive catalog of lamin homologs and potential functional
69 analogs across divergent eukaryotic lineages.

70

71 In the non-metazoan *Dictyostelium discoideum*, NE-81 represents an evolutionary precursor to
72 the A- and B-type lamins found in vertebrates (Batsios *et al.*, 2012; Odell and Lammerding,
73 2023). This *Dicytostelium* protein has the structural features of metazoan lamins, with which it
74 shares primary amino acid sequence similarity (Krüger *et al.*, 2012). Further, transgenic
75 expression of NE-81 in a lamin A/C deficient cell line partially complements the mechanical
76 defects exhibited by nuclei in those cells (Odell, Gräf and Lammerding, 2023). In other non-

77 metazoan groups, distinct NL proteins exist that are difficult to reconcile as lamin homologs. For
78 example, in Trypanosomes, NUP-1 and NUP-2 are NL proteins that appear to function
79 analogously to lamins, but these proteins lack sequence similarity with lamins and share only
80 some structural features (e.g., long coiled-coil domains in NUP-1; see Figure 1) (DuBois *et al.*,
81 2012; Maishman *et al.*, 2016; Padilla-Mejia *et al.*, 2021).
82
83 The situation in plants is more complex. Plants lack recognizable lamins yet nuclei in plant cells
84 have a NL structure (Masuda *et al.*, 1993; Fiserova, Kiseleva and Goldberg, 2009). The major
85 component of this structure in plant cells are Nuclear Matrix Constituent Proteins (NMCPs), first
86 discovered in the angiosperm *Daucus carota* (carrots) (Masuda *et al.*, 1997; Kimura, Kuroda and
87 Masuda, 2010; Ciska and Moreno Díaz de la Espina, 2013). Functional studies in *Arabidopsis*
88 *thaliana*, demonstrate that NMCPs (also called CRWNs, for CROWED NUCLEI, in this
89 species) are involved in the regulation of nuclear size and shape as well as heterochromatin
90 organization (Dittmer *et al.*, 2007; Sakamoto and Takagi, 2013; Wang, Dittmer and Richards,
91 2013; Sakamoto *et al.*, 2022). NMCPs in angiosperms are essential for survival (Wang, Dittmer
92 and Richards, 2013; Blunt *et al.*, 2023), but the sole *NMCP* gene can be deleted without affecting
93 viability in the liverwort *Marchantia polymorpha*, despite the fact that disruption of three-
94 dimensional chromatin organization results (Wang *et al.*, 2021). These functional studies indicate
95 that NMCPs play similar roles to lamins, but these major NL proteins in plants do not share
96 significant amino acid sequence similarity with lamins. Nor do they contain diagnostic features
97 of lamins, such as a C-terminal CAAX box or an immunoglobulin-fold domain (Ciska *et al.*,
98 2019). Nonetheless, NMCPs and lamins share structural characteristics, as illustrated in Figure 1.
99 The most obvious similarity is that NMCPs and lamins have a tripartite structure defined by a

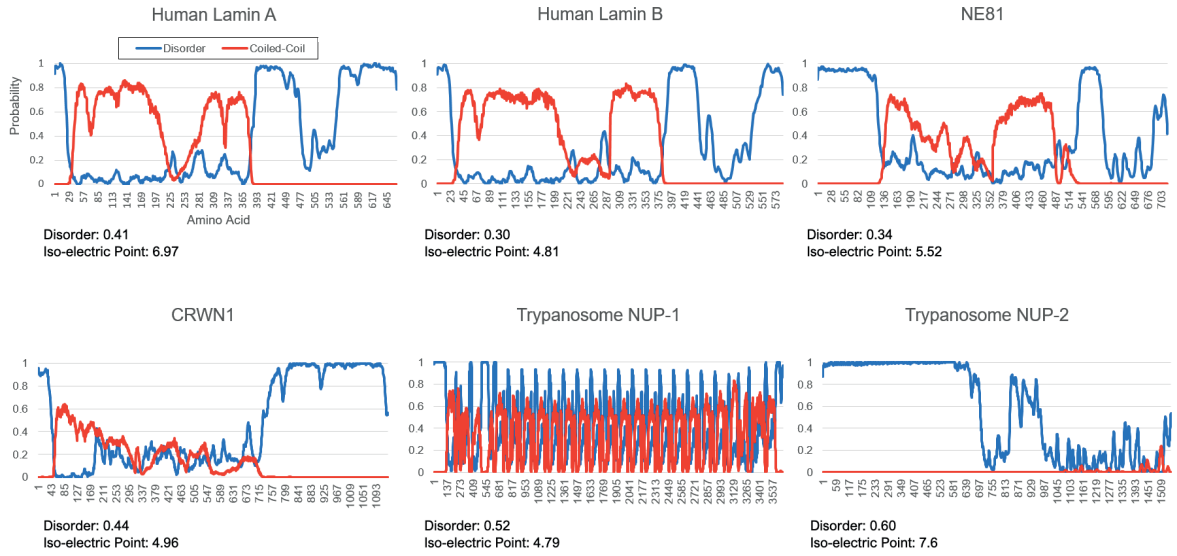
100 long central coiled-coil domain. Another feature shared by NMCPs and lamins is the high
101 percentage of intrinsically disordered regions (IDRs) in the N- and C-terminal domains (Huang
102 *et al.*, 2021; Reimann *et al.*, 2023; Qin *et al.*, 2011). These structural similarities suggest that
103 NMCPs could be highly diverged lamin homologs, but these similarities are also consistent with
104 the convergent evolution hypothesis.
105
106 To determine if NMCPs are distant lamin homologs or evolutionarily distinct, we broadened our
107 search for NMCPs in the 'green lineage,' encompassing land plants and photosynthetic algae
108 (Merchant *et al.*, 2007). All land plants examined to date contain NMCPs, including
109 angiosperms, gymnosperms, ferns, liverworts and mosses (Poulet *et al.*, 2016; Ciska *et al.*,
110 2019). Recent similarity-based studies have shown that the presence of NMCPs in the green
111 lineage extends as far as streptophytic algae, a sister group to land plants, but obvious NMCPs
112 are absent in more distantly related green algae (Chlorophyta) (Koreny and Field, 2016; Ciska *et*
113 *al.*, 2019).
114
115 Recognizable NMCPs are extremely divergent, suggesting that the primary amino acid sequence
116 similarity might not be an effective approach to find ancestral orthologs (Ciska et al., 2019).
117 Moreover, unlike mammalian lamins, NMCPs lack conserved domains, such as diagnostic post-
118 translational modification sites, a C-terminal CaaX prenylation motif, and an immunoglobulin
119 domain, making sequenced-based identification even more challenging (Ciska and Moreno Díaz
120 de la Espina, 2013). We reasoned that overall structural organization would more likely be
121 conserved and therefore more useful in searching for distantly related proteins. Accordingly, we
122 designed a method of proteomic filtration using protein properties rather than amino acid

123 sequence similarity to search for highly diverged NMCP orthologs in the green lineage. One
124 criterion involved filtering proteins based on polypeptide size (> 800 amino acids) and acidic
125 isoelectric point (IEP, pH 4.0 to 6.0). A second criterion was the presence of an extended coiled-
126 coil domain. A third critical criterion of our filtration protocol was a high incidence of
127 intrinsically disordered regions (IDRs), diagnostic of all characterized NMCPs and also found in
128 animal lamins (see Materials and Methods). Intrinsically disordered regions (IDRs) are domains
129 within proteins that lack a defined three-dimensional shape (Piovesan et al., 2022). One
130 possibility is that having a malleable shape allows for plasticity in binding, not attainable by
131 ordered proteins, and dynamic structural ordering (Nishiyama *et al.*, 2018). Phase separation is
132 another emerging theory for IDR function, in which cellular components transition between fluid
133 and more rigid states of matter, often occupying a medial state (Molliex et al., 2015). IDRs
134 promote these phase transitions by acting as linkers and stimulating sticky interactions with other
135 proteins, mediated by weak hydrophobic interactions (Banjade et al., 2015; Boeynaems et al.,
136 2018). Phase separation and condensate formation from such interactions among proteins could
137 contribute to these proteins' organization of the nuclear lamina.
138
139 We focused our search on well-characterized genomes and deduced proteomes of streptophytic
140 algae, which comprise diverse photosynthetic organisms in sister lineages to land plants, with the
141 aim of understanding the evolutionary history of the main component of the plant NL. Our
142 results indicate that NMCP-class NL proteins are restricted to a subset of streptophytic algae, and
143 we found no evidence for transitional states with features more closely related to animal lamins.
144 These findings support the view that organisms in the plant kingdom contain a convergently

145 evolved NL composed of diverse NMCP proteins, rather than specialized plant-specific lamin

146 orthologs.

147




148

149 **Figure 1.** Structural comparison of NL proteins in different eukaryotes

150

151

152 **RESULTS**

153 We began our study by selecting sequenced genomes from seven different taxa in four different

154 Classes of streptophytic algae (see Figure 2). In addition, we examined one green algal

155 (chlorophytic) species (*Chlamydomonas reinhardtii*), one moss species (*Physcomitrium patens*),

156 and one angiosperm species (*Arabidopsis thaliana*) as reference taxa bracketing the streptophytic

157 algae.

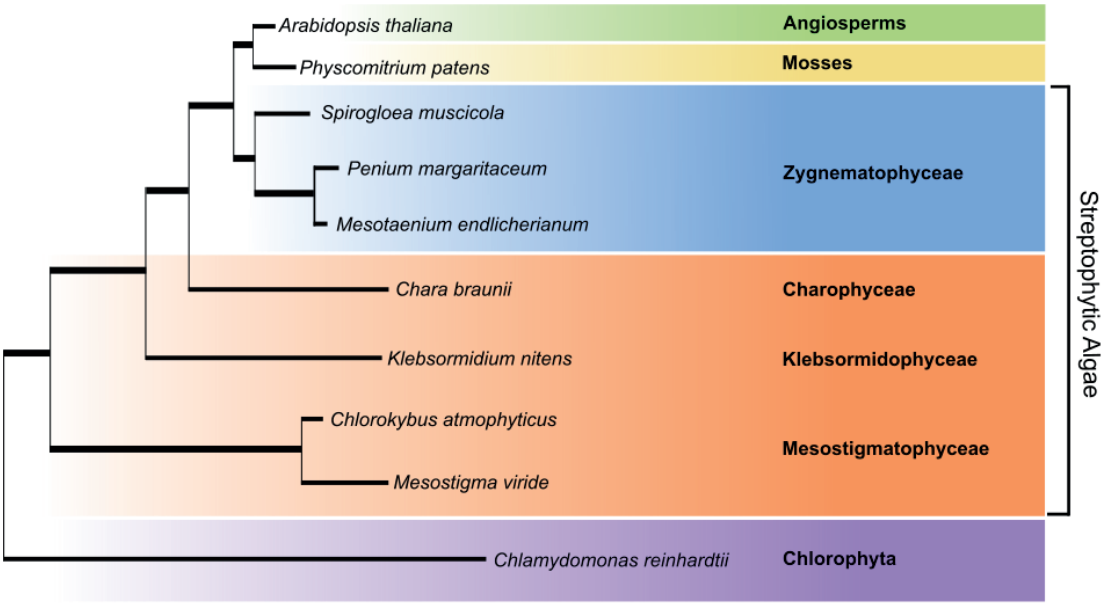



**Figure 2.** Phylogenetic relationships among species that were used in this study.

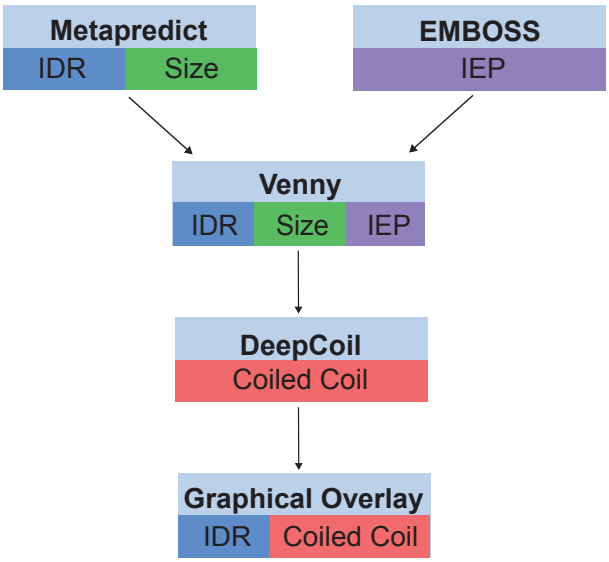



**Figure 3.** Summary of proteome filtration pathway

164

165 Deduced proteomes from the sequenced genomes were processed following the filtration steps
166 diagrammed in Figure 3. The first characteristic used to filter each proteome was the presence of
167 intrinsically disordered regions (IDRs), which are diagnostic of NMCP and lamin proteins (see
168 Materials and Methods). We use the program Metapredict to locate regions predicted to have
169 high disordered content and to filter for protein size. NMCP proteins are typically very large; in
170 all known land plants NMCPs are ≥1000 amino acids, therefore, we selected a cutoff of 800
171 amino acids for filtration. Independently, we used the program Emboss to filter for protein
172 isoelectric point (IEP). Known NMCPs have an acidic IEP, around 5.3. We chose a threshold of
173 IEP between 4-6 to filter the deduced proteomes. The results from these three steps were
174 collected, and the tool Venny: Venn Diagrams (https://bioinfogp.cnb.csic.es/tools/venny/) was
175 used to select proteins that satisfied all three filters.

176

177 With the condensed list, we used the program DeepCoil to generate the probability of finding a
178 coiled-coil domain at each amino acid. These results were expressed graphically and analyzed
179 visually for similarity to the NMCP coiled-coil region. The proteins selected had a central coiled-
180 coil region that spanned at least 100 amino acids. The coiled-coil and IDR scores were overlayed
181 for each protein, and the composite graphical representations were visually inspected to generate
182 a list of finalists.

183

184 We applied this filtration protocol to the *A. thaliana* proteome and recovered seven protein
185 "finalists," four of which were bona fide NMCPs (i.e., CRWN1-4). The remaining three proteins
186 are predicted to be filament-like proteins, including TCS1 (TRICHOME CELL SHAPE 1),
187 which is a cortical microtubule binding protein (Chen *et al.*, 2016). Two additional proteins of

188 unknown function were found (At1g47900 and the SMAD/FHA domain protein encoded by
189 *At2g45460*). In the moss, *Physcomitrium patens*, three finalists were identified by the filtration
190 protocol, two of which are proteins that show a high degree of amino acid similarity with *A.*
191 *thaliana* NMCPs. These two proteins were previously identified as NMCP homologs in *P.*
192 *patens* (Wang, Dittmer and Richards, 2013; Ciska *et al.*, 2019). The third protein
193 (XP_024361755) shows a high degree of similarity with Arabidopsis TCS1 and is a likely moss
194 homolog of this protein. These findings suggest that the filtration protocol is robust and identifies
195 known NMCPs in land plant proteomes, but also retrieves a small number of additional proteins
196 that are likely false positives not involved in nuclear lamina formation.
197
198 Consistent with this expectation, the filtration protocol identified six proteins from
199 *Chlamydomonas reinhardtii,* a member of the green algae family, that lacks known NMCPs and
200 was used as a negative control. Four of these proteins are characterized to have roles outside of
201 the nucleus, including a kinesin motor domain-containing protein (A0A2K3CRT1), a predicted
202 tropomyosin (A0A2K3DS11), a centrosomal protein (A0A2K3D6E5), and a protein implicated
203 in Golgi vesicle trafficking (A0A2K3D1Z8). The remaining two Chlamydomonas proteins
204 identified by our filter did not have a predicted or characterized function.
205
206

207 **Table 1.** Stepwise protein count through filtration

| Species | Proteome | Disorder | Size | IEP | Size Disorder IEP | Coiled Coil (SDI) | Filtered Finalists | NMCP Candidates |
|---|---|---|---|---|---|---|---|---|
| *Arabidopsis thaliana* | 48266 | 21874 | 3539 | 12777 | 596 | 21 | 7 | 4 |
| *Physcomitrium patens* | 48022 | 28682 | 8900 | 12134 | 1985 | 19 | 3 | 2 |
| *Mesotaenium endlicherianum* | 11080 | 7022 | 1476 | 3047 | 296 | 26 | 5 | 1 |
| *Mesostigma viride* | 16232 | 8231 | 3045 | 4090 | 566 | 49 | 11 | 0 |
| *Spirogloea muscicola* | 27137 | 15841 | 3549 | 7006 | 640 | 42 | 13 | 0 |
| *Chara braunii* | 35672 | 21605 | 8157 | 11176 | 2223 | 19 | 5 | 1 |
| *Penium margaritaceum* | 52333 | 36805 | 2094 | 13075 | 612 | 14 | 1 | 1 |
| *Chlorokybus atmophyticus* | 8693 | 4339 | 851 | 2168 | 125 | 9 | 7 | 0 |
| *Klebsormidium nitens* | 16283 | 7846 | 1305 | 4957 | 544 | 128 | 10 | 1 |
| *Chlamydomonas reinhardtii* | 19526 | 12939 | 5051 | 4440 | 1033 | 46 | 6 | 0 |

208

209 After confirming the filtration protocol can identify NMCPs from whole proteomes with a
210 modest number of false positives, seven streptophytic algal species were analyzed using the same
211 approach. Once all four physical characteristics (size, disorder, IEP, and coiled-coil region
212 similarity) were filtered, finalists were analyzed to examine their structural organization. Proteins
213 with confirmed coiled-coil domains were put in a graphical overlay that combined coiled-coil
214 and intrinsically disorder region prediction on a per-residue basis. Graphical analysis allowed for
215 the protein to be visualized holistically and compared to angiosperm NMCP, to assess
216 organizational similarity.

217

218 Each species filtered returned between 1 and 13 finalist proteins after this graphical analysis. We
219 know from the filtration of the control proteomes that a few false positives can be expected.
220 Therefore, another step was needed to differentiate NMCP candidate proteins from the false
221 positives. To supplement our sequence-agnostic approach, we applied multiple sequence
222 alignments to identify amino acid motifs that distinguished among the finalists generated by the
223 filtration protocol. As shown in Figure 4, a single protein from four of the seven streptophytic
224 algae species examined contained a 50-amino acid region of similarity in the coiled-coil region
225 (see the rightmost column of Table 1). The region corresponds to a short, degenerate but
226 conserved region in the N-terminus of the coiled-coil region previously identified in angiosperm
227 NMCPs (Masuda *et al.*, 1997).
228

```
Arabidopsis CRWN1   DPRILPEKISELEKELFEYQHSMGLLLIEKKEWSSQYEALQQAFEEVNEC
Physcomitrium       DRAALHARIAALEAELYDYQYNMGLLLLQRKTWSSQVDELKAAVADAQGT
Mesotaenium         ERGALQARVEELEDELYEYQYNMGVLLMERAEWQPKFDDARVAHAEAEEA
Penium              ERDELMLKVADLGAQLHDTQYNMGLLLMEQAKWNTQYEDARAALAEAEED
Chara               ERNFMVDHIKQLEDELYKYQYHMGLLLMERKSLLADNAKLDSMKKEYEEI
Klebsormidium       EKEELQRQLDDVERELADYRYNLGVLILEHKKCKPQIDELEKALQRTREE
                    :   :  ::  :  :* . :: :*:*::::     .           .
```

229

230 **Figure 4.** Multiple sequence alignment of a weakly conserved block within algal NMCPs
231

232 After identifying this region of sequence similarity in the coiled-coil domain, we wanted to
233 determine if the filtration protocol had missed any other NMCP candidates containing this
234 sequence. The 50-amino acid region of the *P. patens* NMCP protein shown in Figure 4 was used
235 as an independent query in a series of local BLAST searches against the proteome of each algal
236 species. The candidate proteins observed in *M. endlicherianum, C. braunii, P. margaritaceum,*
237 and *K. nitens* returned as significant matches (see Supplementary Table 1). In *P. margaritaceum,*

238 a second protein was identified (pm015243g0010), which was annotated as a putative NMCP in
239 the existing genome annotation (Jiao *et al.*, 2020). This protein has a very high sequence
240 similarity with the NMCP candidate identified in *P. margaritaceum,* however, both the N- and
241 C-terminal domains are truncated relative to the bona fide candidate (pm001739g0030).
242 Consequently, the shorter protein was discarded in our filtration protocol because it lacked
243 significant intrinsically disordered regions at the N- and C-termini. Other top "hits" returned
244 from the local BLAST searches had insignificant scores. Existing annotation and additional
245 protein family analysis of these less significant "hits" revealed relationships to characterized
246 proteins outside the nuclear lamina. These findings suggest that there is only one NMCP protein
247 in *M. endlicherianum, C. braunii,* and *K. nitens*, and possibly two in *P. margaritaceum*, but we
248 found no evidence for NMCPs in *S. muscicola, C. atmosphyticus,* nor *M. viride*.
249
250 We next investigated whether these absences might result from genome misannotation. Using the
251 50-amino acid sequence of *P. patens,* we searched, using tBLASTn, both the transcriptomes and
252 the genomes of three streptophytic algae taxa apparently lacking NMCPs. The transcriptome and
253 genome scans of *C. atmosphyticus* and *M. viride* returned a few proteins in the search with
254 marginally significant e-values; these proteins were the same as those returned in the BLAST
255 search against the algal proteomes (see Supplementary Table 1). These results strengthen our
256 initial conclusion that NMCP proteins are absent in *C. atmosphyticus,* and *M. viride*.
257
258 For *S. muscicola*, scans against the annotated transcriptome failed to return NMCP candidate
259 proteins; however, tBLASTn searches against the genome identified three independent sequence
260 scaffolds with putative exonic sequences capable of encoding a peptide with similarity to the

261 conserved 50-amino acid region (see Supplementary Figure 1A). By analyzing RNA-seq reads
262 aligned to genome scaffolds, we reconstructed the gene structure of three nearly complete NMCP
263 genes in *S. muscicola* (see Supplementary Figure 1B and Supplementary File 1). The three
264 deduced Spirogloea NMCP proteins share high identity (>80%) and contain a characteristic
265 extended coiled-coil region and a C-terminal domain with a high intrinsic disorder score
266 (Supplementary Figure 1C). Gaps in genome sequence obscure the true nature of the N-terminal
267 domain, but each gene is expressed based on paralog-specific polymorphisms in the RNA-seq
268 reads. A whole genome triplication was observed in *S. muscicola* (Cheng *et al.*, 2019), consistent
269 with our finding of three highly similar NMCP paralogs in this species.
270

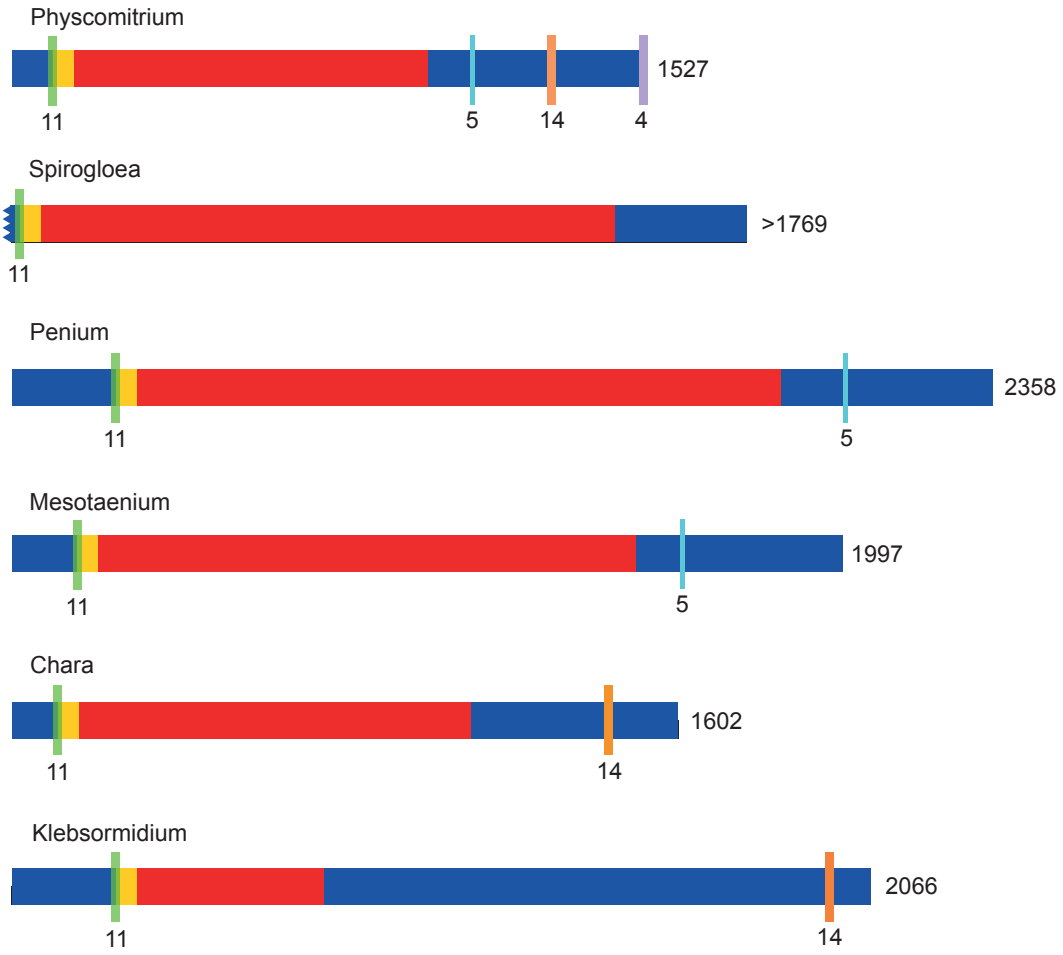
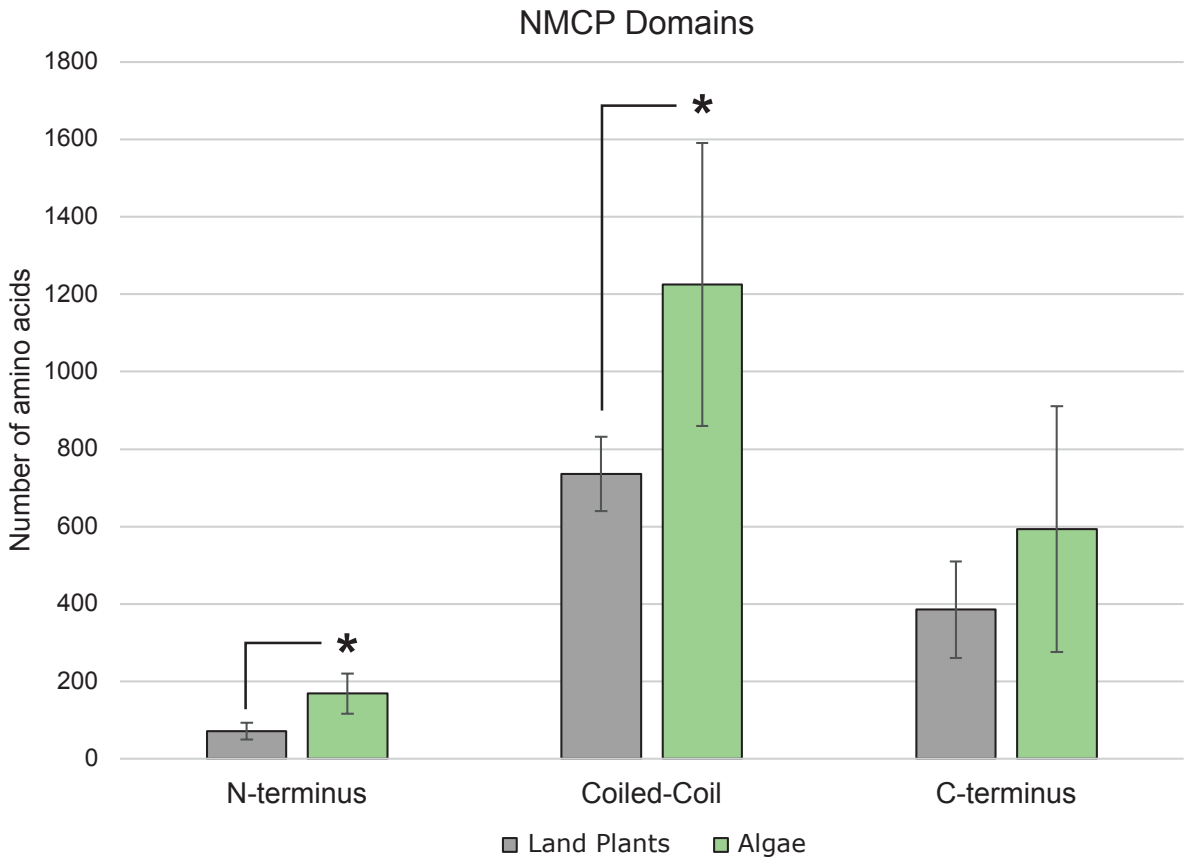



**Figure 5.** Conserved NMCP motifs across algal candidates

We next examined the structural similarities among the four streptophytic algal NMCPs recovered by our filtration protocol. Although these proteins have a similar domain organization, they share only low amino acid sequence similarity, with the highest degree of sequence identity in the coiled-coil domain (21-29%) (Supplementary Figure 2). The four streptophytic algal proteins contain some of the amino acid motifs outside of the coiled-coil region that were originally identified by Ciska et al. (2019) as conserved among NMCP proteins in land plants and algae (Figure 5 and Supplementary Figure 3).

The length of NMCP proteins and their constituent domains varies between angiosperms and streptophytic algae. Figure 5 illustrates a trend of increased size of NMCPs, particularly the intrinsically disordered region, moving from angiosperms to moss to streptophytic algae (Supplementary Figure 4). When the sizes of the three domains of the tripartite NMCPs are compared within and between land plants and streptophytic algae, we noted a larger variance among the algal proteins (see Figure 6). The size of the N-terminal domain averages over 150 amino acids for the four algal proteins identified here, roughly twice the size of N-terminal domains of typical land plant NMCPs (Figure 5). A similar trend is observed for both the coiled-coil and the C-terminal region in streptophytic algal NMCPs compared to land plant NMCPs, although the differences in this sample comparison are not statistically significant (Figure 6). These comparisons demonstrate that considerable variation in NMCP domain size is tolerated in streptophytic algae while preserving the overall tripartite organization of the protein.

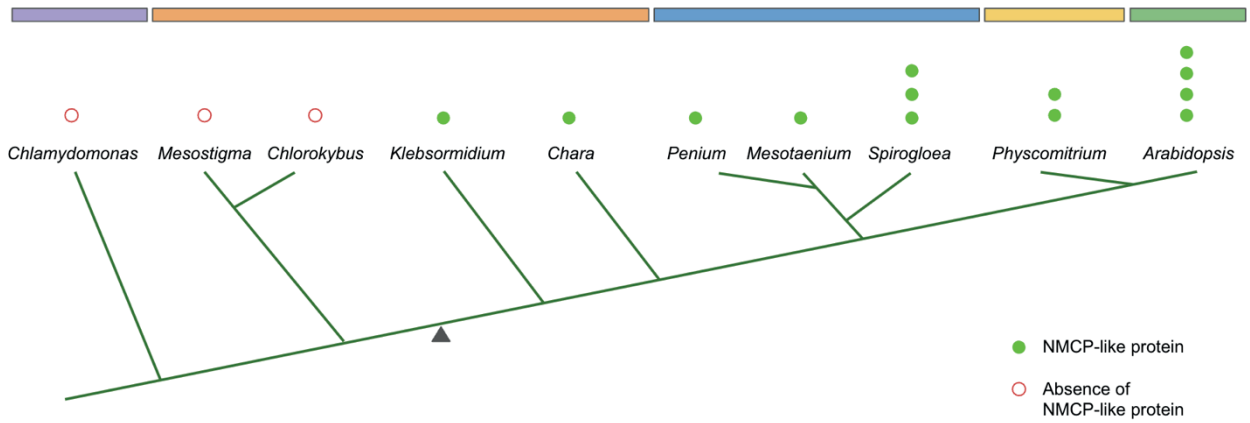



296

297 **Figure 6.** Comparison of the size of each tripartite domain in NMCPs between land plants and

298 algae

299




300

301 **Figure 7.** Distribution of NMCP proteins in the green lineage

302 **DISCUSSION**

303 NMCPs in land plants resemble lamins in structural organization, as illustrated in Figure 1. We
304 reasoned that we might be able to identify transitional stages in the evolution of NMCPs by
305 looking more closely for structural homologs in streptophytic algae in sister lineages of land
306 plants. Would it be possible to start with validated NMCP-class proteins in land plants and
307 bridge out through algae toward the Last Eukaryotic Common Ancestor to identify a connection
308 to lamins? Or are NMCPs truly plant-specific (that is, limited to the 'green lineage')? If the
309 latter, can we identify NMCPs in all lineages of photosynthetic organisms?

310

311 We developed a filtering protocol that considers the physical characteristics of proteins as
312 criteria to search a proteome for potential homologs without dependence on sequence similarity.
313 NMCP candidates could be identified in streptophytic algae through the analysis of four
314 parameters: intrinsically disordered region content, peptide size, isoelectric point, and the
315 presence of an extended coiled-coil domain. Candidate proteins were assessed by visualizing the
316 organization of the tripartite structure with reference to known angiosperm NMCPs, such as
317 *Arabidopsis* CRWN proteins. The filtration protocol (Figure 3) reduced the number of candidates
318 to a small pool (Table 1), and we introduced a final step based on the presence of a weakly
319 conserved region of 50 amino acids (Figure 4).

320

321 Through this process, we identified NMCP proteins in five streptophytic algal taxa:
322 *Klebsormidium nitens*; *Chara braunii*; *Mesotaenium endlicherianum*; *Penium margaritaceum*;
323 and *Spirogloea muscicola* (Supplementary Table 2). The *Spirogleoa muscicola* genome, which
324 underwent a recent triplication (Cheng *et al.*, 2019), encodes three closely related NMCPs

325 (Supplementary Figure 1 and Supplementary File 1), but the remaining four taxa contain only a
326 single NMCP protein. Although we found a second NMCP paralog candidate in *Penium*, this
327 protein contains only the coiled-coil region. It is possible that this second predicted *Penium*
328 protein was derived from a full-length progenitor NMCP or that the unusual protein structure
329 stems from misannotation of the gene.
330
331 Despite retention of a characteristic tripartite structure, the size of the domains varies
332 considerably among the five algal NMCPs (Figure 5), and the overall size of the algal proteins
333 tends to be larger than their counterparts in land plants (Figure 6). Further, the algal NMCPs
334 share very limited amino acid similarity (Supplementary Figure 2), suggesting that primary
335 sequence can diverge significantly if the proteins maintain the structural characteristics that
336 define the NMCP family. Nonetheless, a few short sequence motifs shared among NMCP
337 proteins can be identified among the algal proteins (see Figures 4 and 5) (Ciska *et al.*, 2019).
338 Absent, however, is a well-conserved C-terminal motif, which is present in many NMCP
339 proteins in land plants and implicated in ABA regulation (Supplementary Figure 3) (Zhao *et al.*,
340 2016). The function of the few short amino acid motifs shared by algal NMCPs is unknown; it is
341 possible that these regions represent sites for interaction with core nucleoskeletal or nuclear
342 membrane proteins.
343
344 A previous study by Koreny and Field (2016) uncovered NMCP candidates in streptophytic
345 algae in the genera *Coleochaete, Klebsormidium, Nitella,* and possibly *Spirogyra*, and Ciska et
346 al. (2019) added additional analysis. These NMCP proteins were identified using a similarity-
347 based search from transcriptomic datasets, but the limitations of such an approach were

348 acknowledged by Koreny and Field (2016), citing the extreme divergence of these proteins. In
349 addition, similarity among even unrelated coiled-coil regions could lead to a high background of
350 false positives (Koreny and Field, 2016). We examined the three novel algal NMCP candidates
351 identified by Koreny and Field (2016) using the structural characteristics described in this study
352 (note that the previous study also found the protein from *Klebsormidium* uncovered in our
353 filtration protocol). Two of the three NMCP candidates conform to our structural criteria
354 (JV767595 from *Nitella mirabilis*; and GB-SL01031242 from *Coleochaete orbicularis*);
355 however, the *Spirogyra* protein (GB-SM01021289 from *Spirogyra pratensis*) is a relatively short
356 protein (634 amino acids) that lacks significant intrinsically disordered regions.
357
358 While NMCPs are clearly present in different streptophytic algae, these nuclear lamina proteins
359 are not found in all streptophytes. We conclude that NMCPs are absent from the proteomes of
360 *Chlorokybus atmophyticus* and *Mesostigma viride* based on the following results. First, we
361 analyzed the proteins in our finalist candidate list shown in Table 1 for the possibility that they
362 might represent previously unrecognized NMCP variants. However, these candidates were
363 related to proteins with roles outside the nuclear lamina and therefore are unlikely to be highly
364 diverged NMCP homologs. Second, to identify homologs that might not have the diagnostic
365 tripartite domain organization, we searched the seven complete proteomes using a query
366 corresponding to a 50-amino acid block in the coiled-coil region from a *Physcomitrium patens*
367 NMCP (Figure 4). No new NMCP candidates were recovered by this approach from proteomes
368 of *Chlorokybus atmophyticus* or *Mesostimga viride*. We did, however, identify an additional
369 NMCP-like protein from *Penium margaritaceum* that eluded the filtering protocol because it
370 lacked disordered domains as noted above. Third, to account for NMCP proteins potentially

371 overlooked due to errors in genome annotation, we performed tBLASTn searches with the same
372 query against algal transcriptomes, as well as genome sequences, and found no new candidates.
373 These findings argue that *Chlorokybus atmophyticus* and *Mesostimga viride* lack NMCP
374 proteins.
375
376 Mapping the presence and absence of NMCP proteins relative to the phylogeny of streptophytic
377 algae demonstrates that NMCPs do not appear at the boundary between green algae
378 (Chlorophyta) and streptophytic algae. The streptophytes most closely related to green algae,
379 *Mesostigma viride* and *Chlorokybus atmophyticus,* do not have a recognizable NMCP. Moving
380 toward lineages more closely related to land plants, NMCPs first appear in Klebsormidophyceae,
381 represented by *Klebsormidium nitens*. Taking a further step toward land plants, NMCPs are
382 present in the Charophyceae. Within the Zygnematophyceae, we found NMCPs in all three
383 different taxa examined. These data indicate that NMCP proteins were first acquired in the green
384 lineage at the branch point distinguishing the extant Classes Mesostigmatophyceae and
385 Klebsormidophyceae (Figure 7). Our results, which demonstrate that NMCP-class nuclear
386 lamina proteins are present in some but not all taxa in the green lineage, are consistent with the
387 convergent evolution model for the nuclear lamina in streptophytes. We found no evidence for
388 intermediate NMCP variants, resembling ancestral lamins, corresponding to more basal nodes of
389 the green lineage. Thus, we conclude that NMCPs are plant-specific proteins with an
390 independent origin and are not highly diverged lamins.
391
392 The distribution of NMCP proteins in streptophytic algae provides the opportunity to test the
393 hypothesis that NMCPs play a role in the control of nuclear dynamics disassembly and

394 reassembly during mitosis, analogous to lamins (Goldman *et al.*, 2002). According to this
395 hypothesis, NMCPs would be required to carry out an open mitosis, such as that which occurs in
396 land plants, where the nuclear envelope breaks down before chromosomal segregation and
397 reassembles after division (Rose, 2007). In closed mitosis, the nuclear envelope remains intact
398 throughout division, and the nucleus partitions into two daughter nuclei during cytokinesis
399 (Heath, 1986). Consistent with the hypothesis, green algae (Chlorophyta), which undergo closed
400 mitosis, lack NMCPs (Heath, 1986; Koreny and Field, 2016). Bracketed by green algae and land
401 plants, the streptophytic algae display a range of different mitotic types, including open, semi-
402 closed, and closed. In some cases, streptophytes that contain NMCPs, including in the genus
403 *Klebsormidium, Nitella*, and *Chara*, exhibit an open mitosis (Heath, 1986). The
404 Zygnematophyceae have a semi-closed mitosis, characterized by partial disassembly of the
405 nuclear envelope (Van Den Hoek, Mann and Jahns, 1995). This consideration suggests that all
406 taxa in Zygnematophyceae should contain an NMCP to mediate a partial nuclear breakdown, and
407 our results correspond to this prediction (Figure 7). The state of the nucleus during mitotic
408 division, however, is not necessarily predicted by presence or absence of an NMCP homolog. In
409 certain streptophytes more distantly related to land plants, open mitosis occurs without NMCPs.
410 *Mesostigma* is believed to be the final genus included in the Streptophyta at the boundary
411 between the Chlorophyta (Rodríguez-Ezpeleta *et al.*, 2007). Its sister group, represented in our
412 study by *Chlorokybus atmophyticus*, has an open mitosis but does not have an NMCP (Lokhorst,
413 Sluiman and Star, 1988).
414
415 Therefore, no simple correlation exists between the presence/absence of NMCPs and the form of
416 mitosis in streptophytic algae. The specific roles played by NMCPs in nuclear dynamics remain

417 to be defined, but we anticipate that investigations exploiting the diversity of nuclear types and
418 nuclear lamina components in streptophytic algae will help elucidate these roles.
419
420 **MATERIALS AND METHODS**
421 Each proteomic filtration consisted of four general steps, compounding the previous results
422 before advancing. Four physical characteristics were assessed for filtration: Amino acid chain
423 length, intrinsically disordered regions, isoelectric point, and coiled-coil patterns. In Arabidopsis,
424 CRWN proteins are well studied and provide a convenient positive control and a point of
425 reference for the filtration methods. For example, the CRWN1 protein is 1132 amino acids long,
426 and a threshold of 800 was chosen to account for potential candidates with fewer amino acids
427 than CRWN1. The coiled-coil region of CRWN1 is about 700 amino acids long and has a short
428 N-terminus of about 50 amino acids.
429
430 Metapredict, a machine learning-based intrinsic disorder prediction software, generates IDR
431 predictions for each protein from the amino acid sequence (Emenecker, Griffith and Holehouse,
432 2021). For all proteomic datasets, IDRs were predicted for each individual protein, and values
433 for angiosperms and gymnosperms were compared to the existing literature to confirm their
434 accuracy. Metapredict returned disorder probability scores describing the confidence of a
435 disordered region on a per residue basis across each protein for a complete proteome. These
436 scores were consolidated to a per protein basis by converting per residue scores to a binary
437 system. The developers of Metapredict recommend that a score of 0.3 constitutes significant
438 disorder; subsequently, the threshold of the binary systems was set at 0.3. Taking the binary
439 results per residue, each protein was assigned a percentage of residues satisfying the threshold.

440 The output reported the number of total residues, as well as the number of disordered amino
441 acids and total number of amino acids, allowing for amino acid chain length to be filtered
442 simultaneously with disorder.
443
444 The proteins' isoelectric points were calculated in parallel with filtration for overall size and
445 disorder content. The isoelectric point (IEP) of a protein, or pKi, describes the average pKa of
446 each amino acid residue over an entire protein. The IEP allows for the visualization of a protein's
447 optimal conditions to function. Lamin isoelectric points are typically acidic, ranging from about
448 4-6 pKi (Krohne and Benavente, 1986). The bioinformatics package EMBOSS was used for
449 local isoelectric point prediction in UNIX (Rice, Longden and Bleasby, 2000). Each protein
450 returned a single score which was then filtered to satisfy the acidic range of lamin pKi.
451
452 Filtered lists of proteins for size, IDR, and IEP were compiled and consolidated using Venny's
453 biological Venn diagram tool to select for proteins that satisfied all three criteria
454 (https://bioinfogp.cnb.csic.es/tools/venny/). A FASTA file of only the selected proteins was then
455 generated and filtered for the coiled-coil domain.
456
457 The presence of a long coiled-coil domain is strongly characteristic of lamin proteins. To model
458 these regions computationally, the program DeepCoil, a coiled-coil domain prediction program,
459 was used to assign each residue a probability of being part of a coiled-coil in the tertiary
460 structure (Ludwiczak *et al.*, 2019). The scores of each protein were visualized graphically and
461 analyzed for similar patterns to the coiled-coil domain of CRWN1 in *A. thaliana.* A long central
462 coiled-coil, with a disordered N-terminus of about 100 amino acids and a disordered C-terminus

463 of varying length is indicative of an NMCP-like protein. Each coiled-coil output was assessed
464 visually for these characteristics and was assigned a similarity grouping from 1 (no coiled-coil)
465 to 5 (very similar coiled-coil pattern).

**ACKNOWLEDGEMENTS**

We are grateful for the work of Rachel Christopherson, who worked on an early version of this project. We thank the Adrian Powell, Ryan Preble, and other members of the Boyce Thompson Institute Computational Biology Center for their technical support and advice. This work was supported by funding from the National Science Foundation [grant URoL-2022048 to EJR].

line

***Data Availability***

Genome and Proteome Sources

The *Arabidopsis thaliana* proteome was obtained via TAIR (www.arabidopsis.org). The proteome for *Penium margariticeum* was obtained from the Zhangjun Fei lab (Jiao *et al.*, 2020). The proteome for *Klebsormidium nitens* (Hori et al 2014), *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007), and *Physcomitrium patens* (Rensing *et al.*, 2008) were obtained via the NCBI's GenBank genome portal. The proteomic information for the remaining algae: *Mesotaenium endlicherianum* (Cheng *et al.*, 2019), *Spirogloea musicola* (Cheng *et al.*, 2019), *Mesostigma viride* (Wang *et al.*, 2020), *Chlorokybus atmophyticus* (Wang *et al.*, 2020), and *Chara braunii* (Nishiyama *et al.*, 2018) were downloaded from the JGI genome portal (https://genome.jgi.doe.gov/portal/). Links to each dataset are provided in Supplementary Table 3.

***Code Availability***

The code used for the filtration pipeline in this manuscript can be accessed at https://github.com/brendankz/Proteome-filtration.

492 **Figure Legends**

493

494 *Figure 1:* Structural comparison of NL proteins in different eukaryotes. Scores for intrinsically
495 disordered (blue) and coiled-coil (red) regions were mapped per residue for human lamin A,
496 human lamin B1, Trypanosome NUP-1 and -2, NE81 in *Dictyostelium*, and CRWN1 in
497 *Arabidopsis*. Human lamins, CRWN1, and NE81 share an overall tripartite organization, with
498 disordered regions flanking a long, central coiled-coil region. NE81 differs from human lamins
499 and CRWN proteins in having an elongated N-terminus. The Trypanosome proteins lack this
500 tripartite organization. Note that the proteins lengths are normalized and not drawn to scale.

501

502 *Figure 2:* Phylogenetic relationships among species that were used in this study. Branch lengths
503 are proportional to evolutionary distance, based on Figure 1 from Cheng *et al.* (2019). The
504 modified position of Spirogloea between land plants and Zygnematophyceae is supported by the
505 recent work of Hess *et al.* (2022). Four Classes of streptophytic algae were examined:
506 Zygnematophyceae, Charophyceae, Klebsormidophyceae, and Mesostigmatophyceae.
507 Charophyceae, Klebsormidophyceae, and Mesostigmatophyceae are closely related and grouped
508 together (in orange) because of similarity in their morphological characteristics. Green algae
509 (Chlorophyta) are more diverged from land plants (represented here by Angiosperms and
510 Mosses) than streptophytic algae.

511

512 *Figure 3:* Summary of proteome filtration pathway. Each filtration step is represented by a
513 separate box. The method used to execute each filtration step is shown in the upper half of each
514 box, and the characteristic being evaluated is shown in the lower portion. Intrinsic disorder is

515 represented in blue, size is represented in green, isoelectric point is shown in purple, and coiled-
516 coil is shown in red.
517
518 *Figure 4:* Multiple sequence alignment of a weakly conserved block within algal NMCPs. The
519 50 amino acids of the N-terminus of the coiled-coil regions of NMCP-like candidates and two
520 reference NMCPs are shown. The core of the motif is a well-conserved MGLL sequence, which
521 lies close to the N-terminal border of the extended coiled-coil domain of the proteins. Amino
522 acids in green are polar, red are non-polar, blue are negatively charged, and pink are positively
523 charged. The *Physcomitrium* protein, NMCP1, used in the multiple sequence alignment is
524 A0A2K1L4A2 encoded by the *PHYPA_003649* gene.
525
526 *Figure 5:* Conserved NMCP motifs across algal candidates. Algal candidate NMCPs and *P.*
527 *patens* NMCP1 (A0A2K1L4A2_PHYPA) aligned showing the differences in domain size and
528 presence of conserved regions identified in Ciska et al. (2019). Spirogloea is represented by
529 NMCP A in this figure. Regions in blue represent intrinsic disordered domains, while regions in
530 red represent the coiled-coil domain. The total length of each protein is indicated to the right, #
531 of amino acids; domains are drawn to scale. The yellow box represents the 50-amino acid motif
532 at the beginning of the coiled-coil domain described in this study (Figure 4). The head domain
533 characterized as motif 11 in Ciska et al. (2019) is represented in green. The tail motif 5 is
534 represented in aqua, motif 14 is represented in orange, and motif 4 is represented in purple (see
535 Supplementary Figure 3).
536

537 *Figure 6:* Comparison of the size of each tripartite domain in NMCPs between land plants and
538 algae. Length of different domains (# of amino acids) of NMCP proteins from land plants
539 compared to streptophytic algae. Representative proteins from land plants (n=8) are shown in
540 gray, including *A. thaliana* CRWN1 and 4; *S. lycopersicum* NMCP1B and 2; *Salvinia cucullata*
541 NMCP1 and 2; and both *P. patens* NMCPs. NMCP proteins from streptophytic algae (n=5) are
542 shown in green and are composed of the identified NMCP candidates from the filtering protocol
543 in *S. muscicola*, *M. endlicherianum, C. braunii*, and *K. nitens*, and *P. margaritaceum*. Error bars
544 indicate standard deviation. The asterisk over the algae N-terminus and coiled-coil domain
545 indicates a significant difference in domain sizes assessed using a Tukey HSD test ($p<0.01$).
546
547 *Figure 7:* Distribution of NMCP proteins in the green lineage. Presence or absence of NMCP or
548 NMCP-like protein by species and number of proteins identified. Color scheme of groups
549 corresponds to that shown in Figure 2. Branch lengths are proportional to the tree shown in
550 Figure 2. The hypothesized acquisition of an NMCP protein is indicated by a triangle.
551
552 *Table 1:* Genomic and proteomic file sources. Links to proteome sources for all species
553 considered in this study.
554
555 *Table 2:* Stepwise protein count through filtration. This table summarizes the protein number
556 present after each filtration step was executed. *Proteome* describes the total number of proteins
557 before filtration. *Disorder* corresponds to the proteins satisfying the Metapredict threshold of 0.3.
558 *Size* shows the number of proteins over 800 amino acids in length. IEP represents the proteins
559 that had an isoelectric point in the range of 4-6 pKi. *Size, Disorder, IEP* describes the number of

560 proteins that satisfy all three requirements. *Coiled-Coil* shows the number of proteins with a
561 structurally similar coiled-coil region via DeepCoil (EMBL) per residue predictions, of the
562 proteins that were included in SDI. *Filtered Finalists* describes the number of proteins, of those
563 that had satisfied the previous requirements, with a similar graphical distribution of disorder per
564 residue, which also have protein duplicates removed to describe the number of discrete proteins
565 considered. *Candidates* combines the data from the coiled-coil analysis, mapping the per residue
566 disorder scores with the per residue coiled-coil pattern to look for graphical similarities to a
567 typical NMCP map and ensure both motifs are accounted for and is compared with a multiple
568 sequence alignment to known NMCPs.
569
570

571 **REFERENCES**

572

573 Batsios, P. *et al.* (2012) 'A lamin in lower eukaryotes?', *Nucleus (Austin, Tex.)*, 3(3), pp. 237–
574 243. Available at: https://doi.org/10.4161/nucl.20149.

575

576 Blunt, E.L. *et al.* (2023) 'The nuclear lamina is required for proper development and nuclear
577 shape distortion in tomato', *Journal of Experimental Botany*, 74(18), pp. 5500–5513. Available
578 at: https://doi.org/10.1093/jxb/erad294.

579

580 Chen, L. *et al.* (2016) 'TCS1, a Microtubule-Binding Protein, Interacts with KCBP/ZWICHEL
581 to Regulate Trichome Cell Shape in Arabidopsis thaliana', *PLoS genetics*, 12(10), p. e1006266.
582 Available at: https://doi.org/10.1371/journal.pgen.1006266.

583

584 Cheng, S. *et al.* (2019) 'Genomes of Subaerial Zygnematophyceae Provide Insights into Land
585 Plant Evolution', *Cell*, 179(5), pp. 1057-1067.e14. Available at:
586 https://doi.org/10.1016/j.cell.2019.10.019.

587

588 Ciska, M. *et al.* (2019) 'Evolutionary history and structure of nuclear matrix constituent proteins,
589 the plant analogues of lamins', *Journal of Experimental Botany*, 70(10), pp. 2651–2664.
590 Available at: https://doi.org/10.1093/jxb/erz102.

591

592 Ciska, M. and Moreno Díaz de la Espina, S. (2013) 'NMCP/LINC proteins', *Plant Signaling &*
593 *Behavior*, 8(12), p. e26669. Available at: https://doi.org/10.4161/psb.26669.

Dechat, T. *et al.* (2010) 'Nuclear Lamins', *Cold Spring Harbor Perspectives in Biology*, 2(11), p. a000547. Available at: https://doi.org/10.1101/cshperspect.a000547.

Dittmer, T.A. *et al.* (2007) 'LITTLE NUCLEI Genes Affecting Nuclear Morphology in Arabidopsis thaliana', *The Plant Cell*, 19(9), pp. 2793–2803. Available at: https://doi.org/10.1105/tpc.107.053231.

Dittmer, T.A. and Misteli, T. (2011) 'The lamin protein family', *Genome Biology*, 12(5), p. 222. Available at: https://doi.org/10.1186/gb-2011-12-5-222.

DuBois, K.N. *et al.* (2012) 'NUP-1 Is a Large Coiled-Coil Nucleoskeletal Protein in Trypanosomes with Lamin-Like Functions', *PLOS Biology*, 10(3), p. e1001287. Available at: https://doi.org/10.1371/journal.pbio.1001287.

Emenecker, R.J., Griffith, D. and Holehouse, A.S. (2021) 'Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure', *Biophysical Journal*, 120(20), pp. 4312–4319. Available at: https://doi.org/10.1016/j.bpj.2021.08.039.

Fatafta, H. *et al.* (2021) 'Chapter Five - Molecular simulations of IDPs: From ensemble generation to IDP interactions leading to disorder-to-order transitions', in V.N. Uversky (ed.) *Progress in Molecular Biology and Translational Science*. Academic Press (Dancing Protein

616 Clouds: Intrinsically Disordered Proteins in the Norm and Pathology, Part C), pp. 135–185.
617 Available at: https://doi.org/10.1016/bs.pmbts.2021.06.003.
618
619 Fiserova, J., Kiseleva, E. and Goldberg, M.W. (2009) 'Nuclear envelope and nuclear pore
620 complex structure and organization in tobacco BY-2 cells', *The Plant Journal*, 59(2), pp. 243–
621 255. Available at: https://doi.org/10.1111/j.1365-313X.2009.03865.x.
622
623 Goldman, R.D. *et al.* (2002) 'Nuclear lamins: building blocks of nuclear architecture', *Genes &*
624 *Development*, 16(5), pp. 533–547. Available at: https://doi.org/10.1101/gad.960502.
625
626 Hawkins, J., Davis, L. and Bodén, M. (2007) 'Predicting Nuclear Localization', *Journal of*
627 *Proteome Research*, 6(4), pp. 1402–1409. Available at: https://doi.org/10.1021/pr060564n.
628
629 Heath, I.B. (1986) 'Nuclear Division: A Marker for Protist Phylogeny?', *Progress in*
630 *Protistology*, 1, pp. 115–162.
631
632 Hess, S. *et al.* (2022) 'A phylogenomically informed five-order system for the closest relatives of
633 land plants', *Current Biology*, 32(20), pp. 4473-4482.e7. Available at:
634 https://doi.org/10.1016/j.cub.2022.08.022.
635
636 Hori, K. *et al.* (2014) 'Klebsormidium flaccidum genome reveals primary factors for plant
637 terrestrial adaptation', *Nature Communications*, 5, p. 3978. Available at:
638 https://doi.org/10.1038/ncomms4978.

639

640 Huang, S. *et al.* (2021) 'A phase-separated nuclear GBPL circuit controls immunity in plants',
641 *Nature*, 594(7863), pp. 424–429. Available at: https://doi.org/10.1038/s41586-021-03572-6.
642

643 Jiao, C. *et al.* (2020) 'The Penium margaritaceum Genome: Hallmarks of the Origins of Land
644 Plants', *Cell*, 181(5), pp. 1097-1111.e12. Available at: https://doi.org/10.1016/j.cell.2020.04.019.
645

646 Kimura, Y., Kuroda, C. and Masuda, K. (2010) 'Differential nuclear envelope assembly at the
647 end of mitosis in suspension-cultured Apium graveolens cells', *Chromosoma*, 119(2), pp. 195–
648 204. Available at: https://doi.org/10.1007/s00412-009-0248-y.
649

650 Koreny, L. and Field, M.C. (2016) 'Ancient Eukaryotic Origin and Evolutionary Plasticity of
651 Nuclear Lamina', *Genome Biology and Evolution*, 8(9), pp. 2663–2671. Available at:
652 https://doi.org/10.1093/gbe/evw087.
653

654 Krohne, G. and Benavente, R. (1986) 'The nuclear lamins. A multigene family of proteins in
655 evolution and differentiation', *Experimental Cell Research*, 162(1), pp. 1–10. Available at:
656 https://doi.org/10.1016/0014-4827(86)90421-0.
657

658 Krüger, A. *et al.* (2012) 'Characterization of NE81, the first lamin-like nucleoskeleton protein in
659 a unicellular organism', *Molecular Biology of the Cell*, 23(2), pp. 360–370. Available at:
660 https://doi.org/10.1091/mbc.E11-07-0595.
661

662 Lemieux, C., Otis, C. and Turmel, M. (2007) 'A clade uniting the green algae Mesostigma viride
663 and Chlorokybus atmophyticus represents the deepest branch of the Streptophyta in chloroplast
664 genome-based phylogenies', *BMC Biology*, 5(1), p. 2. Available at: https://doi.org/10.1186/1741-
665 7007-5-2.
666
667 Liang, Z. *et al.* (2020) 'Mesostigma viride Genome and Transcriptome Provide Insights into the
668 Origin and Evolution of Streptophyta', *Advanced Science (Weinheim, Baden-Wurttemberg,*
669 *Germany)*, 7(1), p. 1901850. Available at: https://doi.org/10.1002/advs.201901850.
670
671 Lokhorst, G.M., Sluiman, H.J. and Star, W. (1988) 'The Ultrastructure of Mitosis and
672 Cytokinesis in the Sarcinoid Chlorokybus Atmosphyticus (Chlorophyta, Charophyceae)
673 Revealed by Rapid Freeze Fixation and Freeze Substitution', *Journal of Phycology*, 24, pp. 237–
674 248.
675
676 Ludwiczak, J. *et al.* (2019) 'DeepCoil-a fast and accurate prediction of coiled-coil domains in
677 protein sequences', *Bioinformatics (Oxford, England)*, 35(16), pp. 2790–2795. Available at:
678 https://doi.org/10.1093/bioinformatics/bty1062.
679
680 Maishman, L. *et al.* (2016) 'Co-dependence between trypanosome nuclear lamina components in
681 nuclear stability and control of gene expression', *Nucleic Acids Research*, 44(22), pp. 10554–
682 10570. Available at: https://doi.org/10.1093/nar/gkw751.
683

Masuda, K. *et al.* (1993) 'Residual structure and constituent proteins of the peripheral framework of the cell nucleus in somatic embryos from Daucus carota L.', *Planta*, 191(4), pp. 532–540. Available at: https://doi.org/10.1007/BF00195755.

Masuda, K. *et al.* (1997) 'Peripheral Framework of Carrot Cell Nucleus Contains a Novel Protein Predicted to Exhibit a Long α-Helical Domain', *Experimental Cell Research*, 232(1), pp. 173–181. Available at: https://doi.org/10.1006/excr.1997.3531.

Merchant, S.S. *et al.* (2007) 'The Chlamydomonas genome reveals the evolution of key animal and plant functions', *Science (New York, N.Y.)*, 318(5848), pp. 245–250. Available at: https://doi.org/10.1126/science.1143609.

Nishiyama, T. *et al.* (2018) 'The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization', *Cell*, 174(2), pp. 448-464.e24. Available at: https://doi.org/10.1016/j.cell.2018.06.033.

Odell, J., Gräf, R. and Lammerding, J. (2023) 'Heterologous expression of Dictyostelium discoideum NE81 in mouse embryo fibroblasts reveals conserved mechanoprotective roles of lamins', *Molecular Biology of the Cell*, p. mbc.E23-05-0193. Available at: https://doi.org/10.1091/mbc.E23-05-0193.

705 Odell, J. and Lammerding, J. (2023) 'Lamins as structural nuclear elements through evolution',
706 *Current Opinion in Cell Biology*, 85, p. 102267. Available at:
707 https://doi.org/10.1016/j.ceb.2023.102267.
708
709 Padilla-Mejia, N.E. *et al.* (2021) 'A hub-and-spoke nuclear lamina architecture in trypanosomes',
710 *Journal of Cell Science*, 134(12), p. jcs251264. Available at: https://doi.org/10.1242/jcs.251264.
711
712 Peter, A. and Stick, R. (2015) 'Evolutionary aspects in intermediate filament proteins', *Current*
713 *Opinion in Cell Biology*, 32, pp. 48–55. Available at: https://doi.org/10.1016/j.ceb.2014.12.009.
714
715 Poulet, A. *et al.* (2016) 'Exploring the evolution of the proteins of the plant nuclear envelope',
716 *Nucleus*, 8(1), pp. 46–59. Available at: https://doi.org/10.1080/19491034.2016.1236166.
717
718 Qin, Z. *et al.* (2011) 'Structure and stability of the lamin A tail domain and HGPS mutant',
719 *Journal of structural biology*, 175(3), pp. 425–433. Available at:
720 https://doi.org/10.1016/j.jsb.2011.05.015.
721
722 Reimann, T.M. *et al.* (2023) 'The large GTPase AtGBPL3 links nuclear envelope formation and
723 morphogenesis to transcriptional repression', *Nature Plants*, 9(5), pp. 766–784. Available at:
724 https://doi.org/10.1038/s41477-023-01400-5.
725

726 Rensing, S.A. *et al.* (2008) 'The Physcomitrella genome reveals evolutionary insights into the
727 conquest of land by plants', *Science (New York, N.Y.)*, 319(5859), pp. 64–69. Available at:
728 https://doi.org/10.1126/science.1150646.
729
730 Rice, P., Longden, I. and Bleasby, A. (2000) 'EMBOSS: the European Molecular Biology Open
731 Software Suite', *Trends in genetics: TIG*, 16(6), pp. 276–277. Available at:
732 https://doi.org/10.1016/s0168-9525(00)02024-2.
733
734 Rodríguez-Ezpeleta, N. *et al.* (2007) 'Phylogenetic Analyses of Nuclear, Mitochondrial, and
735 Plastid Multigene Data Sets Support the Placement of Mesostigma in the Streptophyta',
736 *Molecular Biology and Evolution*, 24(3), pp. 723–731. Available at:
737 https://doi.org/10.1093/molbev/msl200.
738
739 Rose, A.-K. (2007) 'Open mitosis: nuclear envelope dynamics.', in *In Cell Division Control in*
740 *Plants, ed.* Heidelberg, Germany: Springer-Verlag, pp. 207–30.
741
742 Sakamoto, Y. and Takagi, S. (2013) 'LITTLE NUCLEI 1 and 4 regulate nuclear morphology in
743 Arabidopsis thaliana', *Plant & Cell Physiology*, 54(4), pp. 622–633. Available at:
744 https://doi.org/10.1093/pcp/pct031.
745
746 Sakamoto, T. *et al.* (2022) 'Two-step regulation of centromere distribution by condensin II and
747 the nuclear envelope proteins', *Nature Plants*, 8(8), pp. 940–953. Available at:
748 https://doi.org/10.1038/s41477-022-01200-3.

Uversky, V.N. (2013) 'Intrinsically Disordered Proteins', in S. Maloy and K. Hughes (eds) *Brenner's Encyclopedia of Genetics (Second Edition)*. San Diego: Academic Press, pp. 124–126. Available at: https://doi.org/10.1016/B978-0-12-374984-0.01698-3.

Van Den Hoek, Mann, D.G. and Jahns, H.M. (1995) 'Chlorophyta: Class 10. Zygnematophyceae', in *Algae: An Introduction to Phycology*. Cambridge University Press, pp. 461-473.

Wang, N. *et al.* (2021) 'Characterization of a Plant Nuclear Matrix Constituent Protein in Liverwort', *Frontiers in Plant Science*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2021.670306

Wang, S. *et al.* (2020) 'Genomes of early-diverging streptophyte algae shed light on plant terrestrialization', *Nature Plants*, 6(2), pp. 95–106. Available at: https://doi.org/10.1038/s41477-019-0560-3.

Zhao, W. *et al.* (2016) 'The Arabidopsis CROWDED NUCLEI genes regulate seed germination by modulating degradation of ABI5 protein', *Journal of Integrative Plant Biology*, 58(7), pp. 669–678. Available at: https://doi.org/10.1111/jipb.12448.