Estimation of the SNP mutation rate in two vegetatively propagating species of duckweed

George Sandler[1], Magdalena Bartkowska[1], Aneil F. Agrawal[1,2]*, Stephen I. Wright[1,2]*

[1]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada; [2]Center for Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

Author for correspondence

George Sandler

Tel: +1 416 6172717

Email: george.sandler@mail.utoronto.ca

*These authors contributed equally to this work.

Key words: Mutation rate, vegetative reproduction, duckweed, asexual, effective population size

Abstract

Mutation rate estimates for vegetatively reproducing organisms are rare, despite their frequent occurrence across the tree of life. Here we report mutation rate estimates in two vegetatively reproducing duckweed species, *Lemna minor* and *Spirodela polyrhiza*. We use a modified approach to estimating mutation rates by taking into account the reduction in mutation detection power that occurs when new individuals are produced from multiple cell lineages. We estimate an extremely low per generation mutation rate in both species of duckweed and note that allelic coverage at *de novo* mutation sites is very skewed. We also find no substantial difference in mutation rate between mutation accumulation lines propagated under benign conditions and those grown under salt stress. Finally, we discuss the implications of interpreting mutation rate estimates in vegetatively propagating organisms.

Introduction

Most research on the evolution of mutation rates has focused either on sexually reproducing eukaryotes or unicellular organisms, both of which feature a single cell phase as part of their life cycle. However, a diverse array of organisms reproduce either through clonal budding, fission or vegetative growth, whereby a single cell phase is not imposed every generation (Bell 1982). This mode of reproduction potentially allows multiple cell lineages to be transmitted from parent to offspring, complicating the process of genotyping individuals. This happens because when individuals composed of a mosaic of cells are sequenced, the mean number of sequencing reads supporting non-reference mutations is no longer 50%. Such a skew in allelic coverage makes it harder to distinguish true mutations from sequencing errors (Cibulskis *et al.* 2013), complicating the assessment of power when calculating per base pair mutation rates. Even if a cellular mutation rate can be calculated for an organism with multiple cell lineages, it becomes more challenging to use this parameter in population genetics analyses as mutations can potentially be lost within an organism before truly contributing to population level genetic diversity. Previous theoretical work modeling mutation load in organisms with multiple cell lineages has suggested that cell lineage selection can significantly reduce mutation load by purging deleterious mutations during somatic growth (Otto and Orive 1995). As most new mutations are thought to be deleterious (Sturtevant 1937; Eyre-Walker and Keightley 2007) this type of selection might skew the level of genetic diversity observed in organisms with vegetative reproduction compared to the level expected given their per base pair mutation rates.

Previous studies have investigated the rate of somatic mutations in plants where multiple cell lineages can segregate within a generation (Watson *et al.* 2016; Schmid-Siegert *et al.* 2017; Plomion *et al.* 2018; Wang *et al.* 2019). While somatic mutations can be transmitted from generation to generation in plants (Plomion *et al.* 2018; Wang *et al.* 2019), if somatic growth is followed by sexual reproduction, a single cell bottleneck is nonetheless imposed on any segregating variation within the soma, removing the persistence of multiple cell lineages across generations. This is however not the case for organisms reproducing through vegetative growth, budding or fission. Despite their frequency across the eukaryotic tree of life, almost no per-base-pair mutation rate estimates exist for organisms procreating through such modes of reproduction. One recent study in a vegetatively growing fairy-ring

64 mushroom reported very low mutation rates per mitotic cell division (Hiltunen *et al.* 2019).
65 The authors of this work used simulated mutations to assess the level of power they had to
66 detect low frequency *de novo* mutation in this dataset, improving their estimate of the fairy
67 ring mushroom mutation rate.

68 Here we report mutation rate estimates in two species of duckweed (*L. minor* and *S.*
69 *polyrhiza*). Both species are free-floating, facultatively sexual aquatic plants. While
70 duckweed can produce seed though sexual reproduction, most growth occurs vegetatively
71 via clonal budding from two pouches present in the duckweed frond (Landolt, 1986). While
72 these species are found all across the globe and likely have enormous census population
73 sizes, allozyme and genomic analyses have revealed low levels of genetic diversity within
74 local populations (Cole and Voskuil 1996; Ho 2018; Xu *et al.* 2019; Ho *et al.* 2019). Work by
75 Xu et al (2019), has estimated the per base pair mutation rate in a genotype of *S. polyrhiza*
76 grown in the field and the lab, finding an extremely low rate of mutation in both cases.
77 However, their analysis did not take into account the fact that duckweed individuals are
78 likely composed of a mosaic of cell lineages during periods of asexual growth, potentially
79 leading to an underestimate of the true mutation rate.

80 Studying two duckweed species allows us to contribute to three other questions in
81 mutation rate evolution research. First, our mutation rate estimates provide another species
82 to add to the existing set of species with mutation rate estimates that, collectively, allow for
83 testing the theory that selection against mutators should be most efficient in species with
84 large effective population sizes (Sung *et al.* 2012; Lynch *et al.* 2016). Duckweeds are useful
85 additions to this set as previous work has suggested that the effective population size (N_e) of
86 *S. polyrhiza* is on the order of a million individuals (Ho et al. 2019; Xu et al. 2019). Second,
87 the inclusion of two facultatively sexual species that differ in their degree of sexuality allow
88 us to preliminarily investigate the effect of recombination on the evolution of mutation
89 rates. Genomic and allozyme patterns have suggested that *L. minor* undergoes bouts of
90 sexual reproduction more often than *S. polyrhiza*, a pattern that is in line with flowering
91 observations of these species in the field (Hicks, 1932; Landolt, 1986). Theoretical work has
92 shown that when recombination breaks apart associations between mutator alleles (that
93 elevate mutation rate) and the mutations they produce, mutation rates can evolve in
94 several ways. On one hand selection against mutators in more sexual populations may be
95 relaxed as they no longer remain linked to new deleterious mutations (Kimura 1967; Leigh

96 1970). Alternatively mutator alleles can spread when recombination is sufficiently low if
97 they hitch hike along with any beneficial mutations they produce (André and Godelle 2006).
98 Finally, environmental stress is known to increase mutation rates in bacteria in a process
99 known as stress-induced mutagenesis (Foster 2007). A few examples of stress increasing
100 mutation rates are known in eukaryotes (Matsuba et al. 2013; Jiang et al. 2014; Sharp and
101 Agrawal 2016; but see Saxena et al. 2019), however it is unclear how general this
102 phenomenon is. We performed our experiment both under a control and salt stress
103 treatment to test whether stress-induced mutagenesis is a common phenomenon in plants.
104 We estimated the mutation rate in 46 asexually propagated mutation accumulation
105 lines, including two genotypes of *S. polyrhiza* and one genotype of *L. minor*. We report an
106 exceptionally low rate of mutation in both species of duckweed and note a pattern of
107 skewed allelic counts at *de novo* sites that suggests the presence of multiple segregating cell
108 lineages in vegetatively reproducing duckweed.
109
110 **Materials and Methods**
111 Mutation accumulation and DNA extraction
112 MA lines were started for three genotypes in April 2014 and propagated for
113 approximately 60 generations. Two *Spirodela polyrhiza* genotypes were used: GP23 from
114 Grenadier Pond, Toronto, Canada and CC from Cowan Creek, Oklahoma, USA. A single
115 genotype of *L. minor* (GPL7) was also isolated from Grenadier Pond, Toronto, Canada. For
116 each genotype (CC, GP23 and GPL7), we established 16 MA lines. We generated each line
117 from a single maternal plant, which was started by isolating two fronds from each genotype
118 culture. Because daughter fronds are generated iteratively, we grew and isolated daughters
119 tracking pedigree until a minimum of 16 daughter fronds paired by generation were
120 available for each of the three genotypes (arising from a single starting maternal frond).
121 Daughters in each frond pairs (matched for generation relative to maternal frond) were
122 assigned to one of two growth medium treatments (salt stress and control). Daughters are
123 produced from two pockets of meristem tissue on either side of the maternal frond and
124 mature daughter fronds remain attached to the maternal plant via a stipule for a short time
125 (Landolt, 1986). To ensure that each generation was propagated with a daughter frond, we
126 separated the daughter from the maternal frond as soon as the daughter began producing
127 her own frond. Each line was checked for mature daughter fronds every two days. The first

128 daughter produced was used whenever possible. MA lines were propagated in 0.5X
129 Appenroth liquid growth medium (Appenroth *et al.* 1996) at 24°C with 12 hours of light per
130 day. Generation times were similar in both species at ~2.9 and ~2.8 days under normal
131 conditions and 3.5 and ~3.3 days under salt stress for *S. polyrhiza* and *L. minor* respectively.
132 Salt stress lines were supplemented with 25mM of NaCl for *S. polyrhiza* and 50mM of NaCl
133 for *L. minor*. Prior to the start of the MA experiment, we performed growth assays to
134 establish stressful NaCl levels for both species. The chosen salinity levels caused duckweed
135 fronds to become patchy and thin but still allowed for continual asexual propagation.
136 After the termination of the MA experiment we allowed MA lines to continue
137 growing for several generations without removing any individuals to obtain enough plant
138 material to perform CTAB DNA extractions.
139
140 Sequencing and filtering
141 We sequenced the MA lines at the McGill Innovation Centre. Illumina HISeq 2000
142 sequencing with 100bp paired end reads was used for both *S. polyrhiza* genotypes while
143 Illumina HISeq 2500 sequencing with 125bp paired end reads was used for the *L. minor*
144 genotype.
145 Paired end reads were mapped to the *S. polyrhiza* and *L. minor* reference genomes
146 (Van Hoeck *et al.* 2015; Michael *et al.* 2017) using the Burrows-Wheeler aligner (BWA) 0.717
147 using the BWA-MEM option (Li and Durbin 2009). We then used Picard to remove duplicate
148 reads before calling indels using the HaploTypeCaller tool in GATK 3.7 (McKenna *et al.*
149 2010). Next, we used the IndelRealigner tool in GATK to perform indel realignment. Finally,
150 we used BCFtools (1.6) (Li 2011) to create mpileup files for the realigned output from GATK
151 and to call SNPs and short indels (indels no more than 10bp). After mapping, mean and
152 median coverage was 26, and 25 for individual *S. polyrhiza* lines, and 18, 17 for individual *L.*
153 *minor* lines. We also calculated total median coverage for each site within each genotype
154 (by summing across all individual lines), which was 436, 426, 280 for genotypes GP23, CC,
155 and GPL7 respectively.
156 We first filtered out sites with unusually high or low coverage. We did this by
157 eliminating sites that had coverage outside +/- 200x median coverage (summed across all
158 lines) in each *S. polyrhiza* genotype and +/- 100x median coverage in the *L. minor* genotype
159 due to the lower quality of the reference genome for this species. We visualised relatedness

160 between our lines using a PCA plot created from heterozygous sites present in our MA lines
161 in R (v5.3.5) (R Core Team 2019) using the package SNPRelate (Zheng *et al.* 2012). In doing
162 so we discovered two major outliers in one of our *S. polyrhiza* genotypes (CC) suggesting
163 that these two lines were cross contaminated. We subsequently removed these two lines
164 from our analysis.

165 Our next round of filtering aimed to remove low quality regions of the genome that
166 contain unusually high amounts of in-phase heterozygous variants (e.g. Figure S1). Such
167 regions likely represent collapsed duplications in the reference genome that map poorly to
168 an incorrect genomic coordinate. These variants are highly reference-biased in their allelic
169 coverage likely due to the poor mapping of reads that contain many differences relative to
170 the reference genome. To remove such regions, we first created a consensus genotype for
171 each set of lines; if more than one line in a given genotype supported the existence of a
172 heterozygote at a site, that site was designated as heterozygous in the consensus genotype.
173 We then performed a sliding window analysis on heterozygosity on each consensus
174 genotype. We used 1000bp windows with a 100bp step. After trying a variety of filtering
175 criteria, we decided to designate regions of the genome as callable if there existed no more
176 than 10 heterozygous calls in a 1000bp window in each *S. polyrhiza* genotype and no more
177 than 5 heterozygous calls per 1000bp window in the *L. minor* genotype. We used more
178 stringent criteria in *L. minor* due to the lower quality of the genome assembly. These cut-
179 offs represent a trade-off between eliminating problematic, variant-rich areas of the
180 genome and excluding well assembled genomic areas with higher than average diversity.
181 After filtering, around 100Mb of the genome was retained as callable in each of the three
182 genotypes. This filtering step greatly improved the allelic coverage of ancestral
183 heterozygous sites by removing suspected hidden duplications that map poorly to the
184 reference (Figure S2).

185 Next, we called putative *de novo* mutations in the remaining callable regions. Within
186 each set of lines, we picked sites where one line had a heterozygous genotype, with at least
187 5 reads supporting the non-reference base, but all other lines supported a homozygous
188 genotype. We then extracted such sites from the mpileup file used to call genotypes. This
189 was done as the pileup file contains reads that are filtered out during genotype calling but
190 are useful in our case as they can lead to the elimination of false positive mutations. We
191 filtered putative *de novo* mutations using the mpileup file in two ways. First, if a line other

192 than the one which contained the *de novo* mutation had any reads which supported the *de*
193 *novo* base call we excluded the site. Second, if a site with a *de novo* base call contained
194 reads with more than two non-reference bases across all samples, we also excluded the site.
195 We did this to exclude sites where a high rate of sequencing errors might have occurred. We
196 used this cut-off based on the observation that at sites where all lines supported a
197 homozygous genotype, the vast majority of sites contain no more than one alternate base
198 call (again likely due to sequencing errors which can be observed in the mpileup file). The
199 remaining putative *de novo* mutations that passed these filtering criteria were visually
200 inspected in the Integrative Genomics Viewer (IGV) (Robinson *et al.* 2011). We excluded a
201 few mutations which appeared on reads in complete linkage with other non-reference bases
202 (an indication of hidden genomic duplications) or on reads that looked like the product of
203 PCR or sequencing errors (see Figure S3-5 for examples).
204
205 Power analysis
206 To calculate the per generation mutation rate, we first needed to know how much
207 power we had to detect *de novo* mutations at our callable sites. To assess power, we first
208 obtained a list of sites where we knew we had non-zero power to detect mutations, this
209 included sites where all lines within a genotype supported a homozygous reference base call
210 and no more than one alternate base was present in the mpileup file (one less than in the
211 case if a *de novo* mutation was present). We then randomly sampled 500,000 such sites
212 from each genotype independently, and randomly chose a line where a mutation could have
213 happened. We randomly eliminated a third of the sites where one alternate base was
214 present in the mpileup file as our filtering criteria would eliminate true *de novo* mutations if
215 another line by chance contained a sequencing error which matched the *de novo* base call
216 (i.e., we assume the probability of this occurring is 1/3). Of the remaining sites, we assigned
217 how many reads would support the *de novo* base call by drawing from a binomial
218 distribution with a success rate of 50%, 34%, 28%, 20% and 10%. These different values
219 were chosen to represent a range of frequencies a mutation may be found due to the
220 inheritance of multiple cell lineages in asexual reproduction. This is similar to the approach
221 taken by Hiltunen et al. (2019) when calculating mutation rates in vegetatively growing
222 fairy-ring mushrooms.

223 Our estimate of power was the proportion of sites (out of the original 500,000) that
224 had at least 5 reads supporting the *de novo* mutation (for each possible binomial success
225 rate that we tested). We then multiplied our power estimates by the number of callable
226 sites in the genome. Then separately for each line, we multiplied the adjusted number of
227 callable sites by the number of MA generations and summed these values across all lines in
228 a given genotype (split by treatment). This provided us with a denominator for our mutation
229 rate calculation. Our final step was to divide the number of *de novo* mutations identified in
230 each genotype (split by treatment) by this denominator. The *de novo* mutation count was
231 adjusted for false positives identified during mutation validation.
232
233 Calling indels
234 We scanned our lines for *de novo* indels in the same way as we searched for point
235 mutations with one key modification. We used the same regions of the genome that we had
236 previously assessed as "callable"; however, we only considered *de novo* indels if they were
237 at least 2000 bp away from any other indel in any other line of the same genotype. This
238 filtering step was required to avoid false positive indel calls which appear due to spurious
239 mapping patterns in repetitive regions.
240
241 Mutation validation
242 We only had tissue from *S. polyrhiza* to perform validation on putative *de novo*
243 mutations as unlike *L. minor*, *S. polyrhiza* produces asexual resting stages called turions
244 which we were able utilize for long term refrigerated storage. After allowing turions to
245 germinate we extracted DNA from our MA lines using a Qiagen DNeasy kit. Afterwards, we
246 designed primers for 14 SNPs and one indel found in our two *S. polyrhiza* genotypes. We
247 performed PCR reactions using FroggaBio PCR mastermix which were then sent off for
248 Sanger sequencing at Eurofins Genomics. We inspected the Sanger sequencing
249 chromatograms in FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA;
250 http://www.geospiza.com) to see if we could detect a peak suggesting the presence of a
251 mutant base in the focal MA line and checked that the base was not present in a different
252 MA line of the same genotype as a negative control. One line from the *S. polyrhiza* CC
253 genotype (line E) contained a massive overabundance of putative *de novo* mutations (of the
254 29 initial mutations found in 14 lines of the CC genotype, 9 were identified from this line); all

255 3 variants that we attempted to Sanger sequence validate from this line failed. Ultimately,
256 we excluded this CC line from all the results presented here. Excluding CC line E, 11 total
257 SNPs that were checked with Sanger sequencing and used to correct for the false positive
258 rate in mutation detection in all our *S. polyrhiza* lines. We adjusted our mutation rate for
259 false positive by excluding mutations that failed validation, summing all mutations that
260 passed validation, and multiplying the sum of unvalidated mutations by our estimate of the
261 false positive rate. Our formula for the true positive rate is as follows:

$$TP = \frac{n_{validated}}{n_{validated} + n_{failed}}$$

262 Where $n_{validated}$ refers to the total number of mutations successfully validated in *S.*
263 *polyrhiza*, and n_{failed} refers to the total number of mutations that failed validation.
264
265 Finally, the formula for our per generation, per base pair mutation rate estimates is
266 as follows:

$$\frac{n_{validated\ in\ lines} + TP \times n_{not-checked\ in\ lines}}{\sum(\#callable\ sites\ \times \#generations)\ \times\ power\ \times 2}$$

267 where $n_{validated\ in\ lines}$ refers to the number of validated mutations in the focal set of lines,
268 and $n_{not-checked\ in\ lines}$ refers to the number of mutations not tested with Sanger
269 sequencing in the focal set of lines. The number of mutations and callable sites was summed
270 across all lines within a genotype, within a treatment. The true positive rate, TP, was
271 estimated once using all Sanger-tested mutations from across the entire experiment.
272
273 <u>Statistical analysis</u>
274 We calculated 95% confidence intervals for our mutation rates using the Agresti
275 Coull method implemented in the R package "binom" (Dorai-Raj 2014). We tested for
276 significant differences between our salt and normal treatment lines using independent Chi-
277 square tests in each genotype. To test for differences between *L. minor* and *S. polyrhiza* we
278 merged the number of callable sites (scaled by power) and number of mutations in the two
279 *S. polyrhiza* genotypes (separately for each treatment), and also applied Chi-square tests to
280 test for significance in mutation rate variation between species and conditions.
281

282 Data availability

283 Raw sequencing data is available on the NCBI SRA database under accession PRJNA659313
284 for *S. polyrhiza* and PRJNA659264 for *L. minor*. Custom scripts and additional data are
285 available at https://github.com/gsan211/duckweed_MA. Supplementary material has been
286 uploaded as a separate file.

287

288 **Results**

289

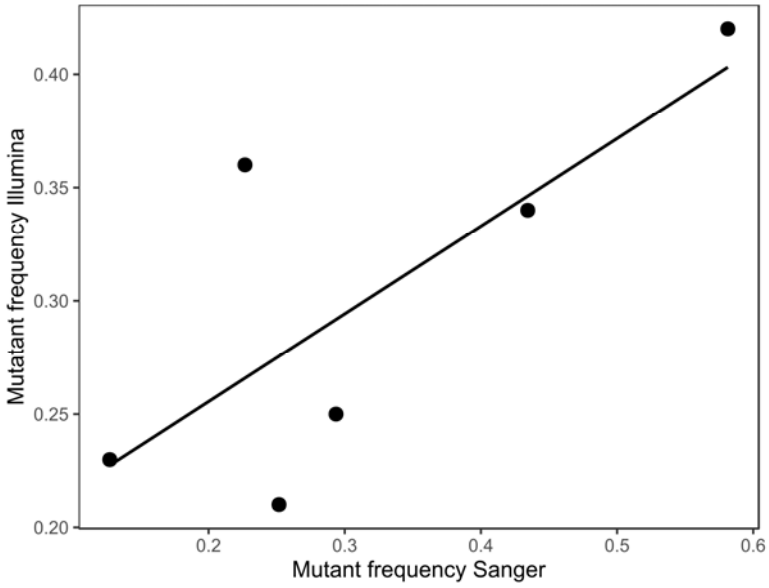
290 *De novo* mutations

291 In total, we identified 23 and 19 putative *de novo* mutations in *S. polyrhiza*
292 genotypes GP23 and CC, respectively and a further 29 mutations in *L. minor* (Supplemental
293 table 1). When inspecting the putative *de novo* mutations in IGV, we observed that most
294 mutant sites exhibited highly reference biased allelic counts (over 50% of reads support the
295 reference base call). On average, mutant sites contained 26%, 30%, and 34% mutant reads
296 with SD 8.2%, 11.6%, and 9.1% in genotypes GP23, CC (*S. polyrhiza*) and GPL7 (*L. minor)*
297 respectively. We implemented several filtering and quality control steps to ensure these
298 mapping patterns were not a result of sequencing error or genome mis-assembly. First, we
299 masked areas of the genome that had odd coverage patterns or were enriched for
300 heterozygous calls to avoid areas containing hidden genomic duplications. The heterozygous
301 variants that pass these filtering criteria and are present in the ancestral genotypes (i.e. are
302 shared by all MA lines within a genotype) show relatively normal patterns of reference and
303 non-reference allele coverage. Second, we only considered *de novo* mutations at sites with
304 minimal sequencing errors. This step eliminated problematic areas of the genome prone to
305 genotyping error. Finally, we performed Sanger sequencing validation on 11 putative *de*
306 *novo* mutations (in *S. polyrhiza*) of which 6 were validated with both positive and negative
307 controls. The pattern of reference bias in our *de novo* mutations was also present in our
308 Sanger sequencing chromatograms; the reference base peak was generally much larger than
309 the mutant peak. When we plotted Illumina mutant base frequency vs. Sanger mutant peak
310 height (standardized by reference peak height), we observed strong concordance for our 6
311 validated mutations (R^2 = 0.56; Figure 1) suggesting that these are not spurious mapping
312 patterns or sequencing errors but rather reflect a true bias in mutation abundance. When
313 we inspected our PCR products with gel electrophoresis we observed single, clean bands of

314 the expected product size so we consider it unlikely that multiple sites in the genome were
315 amplified leading to odd mapping patterns at our *de novo* mutation sites. One alternative is
316 that reference biased mapping patterns may be due to some bias in the PCR amplification
317 process. We do not believe this is likely however since our primers were designed for
318 sequences that were flanking the site of *de novo* mutations that should be identical whether
319 a *de novo* mutations is present or not.

320 We report a high false positive rate for *de novo* mutation identification in *S. polyrhiza*
321 of ~45%. This is likely a consequence of having to distinguish between true mutations with
322 low allelic counts and sequencing errors or bioinformatic artifacts that can appear at
323 similarly low frequencies. Additionally, background noise in Sanger Sequencing
324 chromatograms can obscure variants with low allelic counts, posing a potential way in which
325 false positive rates may be artificially elevated. We attempted to avoid this issue by ensuring
326 that our chromatograms had relatively low levels of background noise before confidently
327 assigning mutations as failing or passing validation.

328



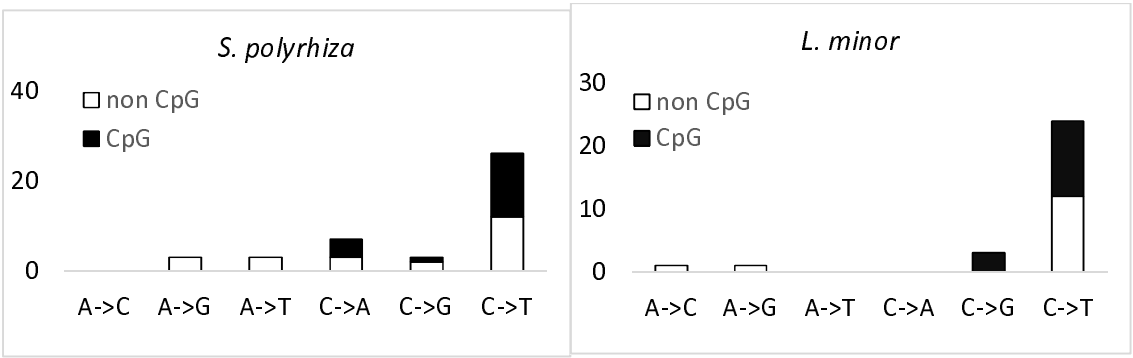
329

330 ***Figure** 1 Proportion of de novo bases from Illumina reads vs. relative peak height of mutant*
331 *base in Sanger sequencing data in Spirodela polyrhiza.* $R^2 = 0.57$.

332

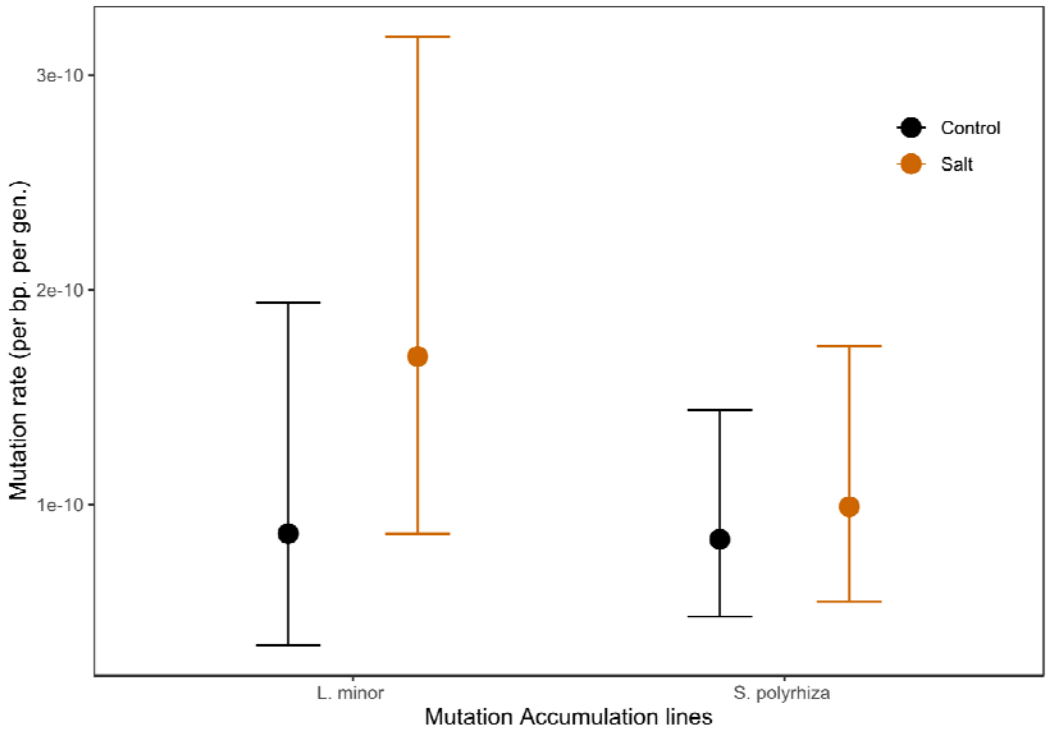
333 The majority of mutations were C -> T transitions in both species of duckweeds
334 (Figure 2) in concordance with patterns found in previous mutation rate studies (Ossowski

335 *et al.* 2010; Thomas *et al.* 2018). Transitions were more common than transversions, with
336 the average ti/tv ratio being 2.23 for *S. polyrhiza* and 4.5 for *L. minor*. This is consistent with
337 previous results reported in *S. polyrhiza* by Xu et al. (2019) where three C -> T and one C -> A
338 mutations were detected and validated. Approximately 50% of C -> T transitions occurred at
339 CpG sites in both species, a pattern that is consistent with previous evidence of elevated
340 mutation rates at such sites (Hodgkinson and Eyre-Walker 2011). We used SNPeff (Cingolani
341 *et al.* 2012) to annotate our putative de novo mutations with most SNPs being labelled as
342 intergenic (See supplemental table 1).
343 Our analysis uncovered only one putatively *de novo* short indel that turned out to be
344 a false positive based on our Sanger sequencing validation analysis.
345



346
347 ***Figure 2*** *De novo mutation spectra in two species of duckweed (Spirodela polyrhiza, Lemna*
348 *minor).*
349
350
351 Mutation rate comparisons between *S. polyrhiza* and *L. minor*
352 Our estimate of the per generation, per base pair SNP mutation rate highly depends
353 on our power to detect mutations, which in turn depends on factors such as total read
354 depth and the number of reads that support a *de novo* mutation base-call (See
355 Supplemental tables 2-5). Heterozygous SNPs are expected to be supported by around 50%
356 of reads in organisms with a single cell phase. In our case, on average, 28% (*S. polyrhiza*) and
357 34% (*L. minor*) of reads supported *de novo* heterozygous mutations. To improve our
358 estimate of the mutation rate, we assumed that we should only expect to find mutations
359 that are on average supported by 28% (in *S. polyrhiza)* or 34% (in *L.* minor) of reads at each
360 mutant site (see methods for more details of power analysis). In this case, our point

361 estimate of the mutation rate for plants grown in standard medium is 8.39E-11 for *S.*
362 *polyrhiza* and 8.66E-11 for *L. minor* (Figure 3). For plants grown in stressful salt medium our
363 point estimates for the mutation rate are 9.91E-11 for *S. polyrhiza* and 16.9E-11 for *L. minor*.
364 While the estimated mutation rate is higher for *L. minor* in both conditions, the
365 difference between the two species is non-significant (chi-square test: p = 1 and p = 0.30 for
366 control and salt stressed conditions). The difference in mutation rate in the two species is,
367 however, very sensitive to the expected frequency of reads that support the mutant base.
368 For example, if our power estimate is based on the assumption that *de novo* mutations are
369 expected to constitute only 10% of reads, the mutation rate under normal conditions
370 increases to 5.13E-10 in *S. polyrhiza* and 16.8E-10 in *L. minor*. Our mutation rate estimates
371 were not significantly different between our two individual *S. polyrhiza* genotypes (chi-
372 square test: p = 0.50 and p = 0.77 for control and salt stressed conditions). Supplemental
373 tables 3-4 give the inferred mutation rates with alternative assumptions for the expected
374 proportion of non-reference reads at sites harbouring *de novo mutations*.
375



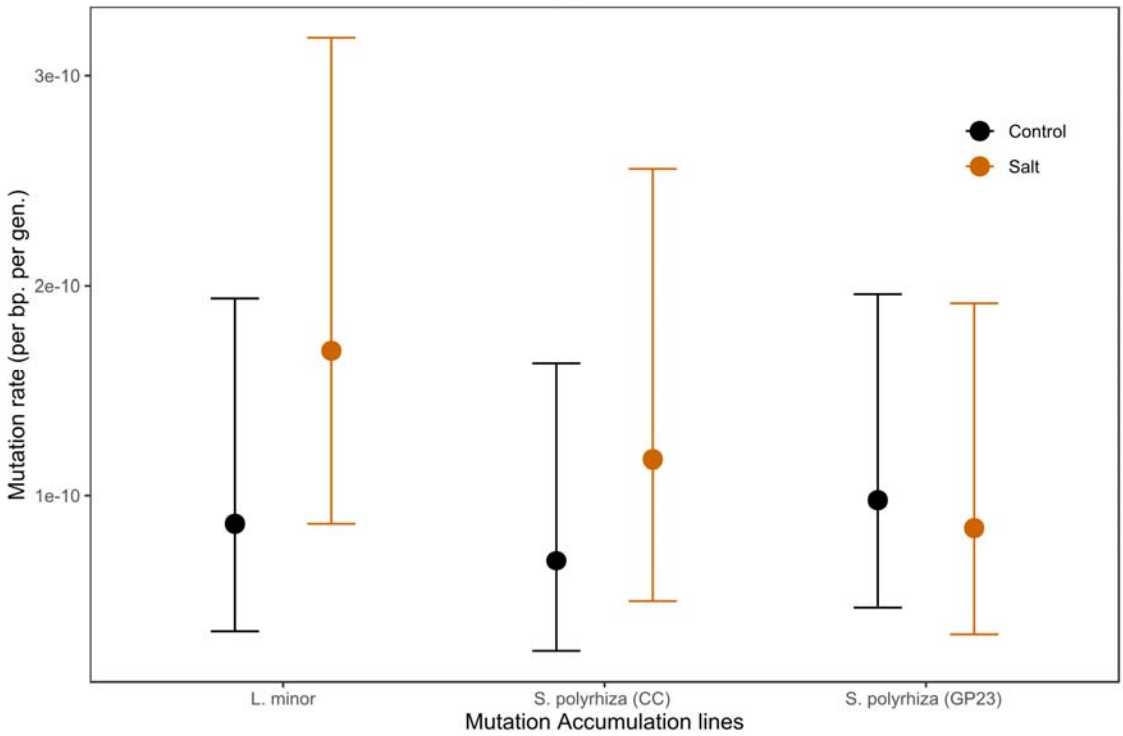
376
377 ***Figure 3*** *Mutation rate estimates in two species of duckweeds duckweed (Spirodela*
378 *polyrhiza, Lemna minor). Error bars show Agresti-Coull 95% confidence intervals.*
379

380 Effect of salt stress on mutation rate

381 There were no significant differences in the mutation rate between lines propagated
382 in normal and salt stress medium for either of our three genotypes (Figure 4, chi-square
383 test: p = 1.00, p = 0.53, p = 0.29 for genotypes GP23, CC, GPL7 respectively). There was also
384 no consistent trend in the change of mutation rate with the addition of salt stress as
385 genotype GP23 appeared to have a decreased mutation rate while genotypes CC and GPL7
386 appeared to have an increase in the mutation rate with the addition of salt stress.

387 Salt stress also did not have a significant impact on the ti/tv ratio. In two of the
388 genotypes (1.33 -> 2 *S. polyrhiza* CC and 4.5 -> 8 *L. minor*), salt stress increased the ti/tv
389 ration. In *S. polyrhiza* genotypes GP23 salt stress decreased the ti/tv ration (3.67 -> 2.00)
390 although the difference was not significant in either of the three genotypes (chi-square test:
391 p > 0.5 all cases). These results run counter to previous work in *A. thaliana* where salt stress
392 was found to increase the de-novo mutation rate ~two-fold and lower the ti/tv ratio (Jiang
393 *et al.* 2014).

394



395

396 ***Figure 4*** *Effect of salt stress on the mutation rate in two genotypes of Spirodela polyrhiza*
397 *and one genotype of Lemna minor. Error bars show Agresti Coull 95% confidence intervals.*

398

399 **Discussion**

400 Using whole genome sequencing on 46 MA lines, we report a low per base pair, per
401 generation SNP mutation rate in two species of duckweeds under two growth conditions. An
402 important result in our study is that *de novo* mutations appear to have considerably
403 reference biased genomic coverage in both duckweed species. We believe that this pattern
404 is not indicative of sequencing or genome assembly errors but rather is a by-product of
405 vegetative reproduction for several reasons. First, upon our inspection of the data in an
406 independent study of mutation rates in *S. polyrhiza* by Xu et al.; we noticed similar levels of
407 reference bias in Illumina short-read sequencing data at validated *de novo* mutations.
408 Second, our own validation with Sanger sequencing showed that mutations with a higher
409 reference bias in the Illumina dataset tended to have higher reference base peaks in their
410 Sanger sequencing chromatograms and vice versa. Moreover, after filtering, most ancestral
411 heterozygous sites that were shared by all lines in the three clonal genotypes were not
412 reference biased in such a manner and those that were, were found in regions highly
413 enriched for non-reference base calls. None of our putative *de novo* mutations are found in
414 such regions; due to the low genetic diversity in duckweed, *de novo* mutations were
415 generally the only heterozygous variants present in the immediate genomic area. This
416 means that mapping bias due to divergence from the reference is unlikely to cause the
417 strong allelic coverage bias we observed. Finally, a study of mutations in a vegetatively
418 growing fairy-ring mushroom also reported patterns of reference bias with an average allelic
419 coverage of 44% across 111 *de novo* mutations (Hiltunen *et al.* 2019). Therefore, the
420 reference bias in our mutations is most likely a signature of the segregation of multiple cell
421 lines from generation to generation in our duckweed MA experiment. Given that vegetative
422 reproduction through clonal budding is the main form of reproduction in duckweed
423 (Landolt, 1986), it seems reasonable that duckweed might not undergo a single cell phase
424 for prolonged periods of time.

425 Simple models can be helpful in clarifying how multi-cell descent during vegetative
426 reproduction may affect mutation rates and our estimation of them. Assume that clonal
427 reproduction occurs by n parental cells forming a diploid offspring (generation 1). If a
428 mutation occurs in one of these n cells, because of multi-cell descent, the offspring will be a
429 genetic mosaic (i.e., not all cells will be genetically identical). The mutation begins at a
430 frequency of $1/2n$ in the offspring. With conventional single-cell descent, where $n = 1$, a new

431 mutation is expected to be at 50% frequency, but the frequency will be lower than 50%
432 whenever $n > 1$. The n cells multiply in some (unknown) growth pattern to produce the
433 mature offspring, which then itself reproduces. In the absence of detailed knowledge of
434 developmental growth trajectories, we do not know what the representation of the original
435 n cell lineages will be in the subsequent generation.

436 We can consider two simple scenarios: If only one of these original cell lineages gives
437 rise to the next set of n cells used to produce the next generation (generation 2), then the
438 mutation will either be completely lost from the lineage (if it was not in the chosen cell
439 lineage) or it will be present in heterozygous state in all cells of future generations (if it was
440 present in the chosen cell lineage). In this first scenario, the genetic mosaicism resulting
441 from multi-cell descent persists only a single generation. Thus, in an experiment such as
442 ours, only mutations occurring in the very last generation will be affected by this issue (i.e.,
443 multi-cell descent is a trivial issue in this case). A second scenario is that all cell lineages
444 grow at equal rates and, each generation, n cells are chosen at random to form the next
445 daughter. This is conceptually similar to the process of coalescence in a population (of cells
446 here rather than individuals) of size n. Thinking backwards in time, all cells must eventually
447 trace their coalescent history to a single cellular ancestor, but it may take many generations
448 for this to happen (i.e., $2n$ generations, on average, but with variance proportional to n^2).
449 Thinking forwards in time, a mutation occurring in one of the n cells forming the generation
450 1 offspring may be present for many future generations, possibly becoming quite common
451 within individuals, before eventually being present in all future cells (with chance $1/n$) or
452 none of them (with chance $(n - 1)/n$). With even modest values of n (e.g., $8 < n < 80$), genetic
453 mosaicism can persist for many generations. Mutations that will eventually be present in all
454 lineages—as well as those that will eventually be eliminated—will thus be found below the
455 50% frequency expected in a "traditional" (i.e., non-mosaic) heterozygote. Although this
456 second scenario as formulated here is unrealistically simple (e.g., random and independent
457 choice of n cells for each generation), it illustrates how multi-cell descent can have
458 consequences across multiple generations. Though developmental growth trajectories in
459 duckweed are insufficient to formulate a more realistic model, we suspect that multiple cell
460 lineages persist across multiple generations in duckweed and is responsible for the clear
461 bias towards mutant SNPs being less than 50%. Recent work on segregating mutations in a
462 single *Zostera marina* seagrass clone has made similar arguments to this model (Yu et al.,

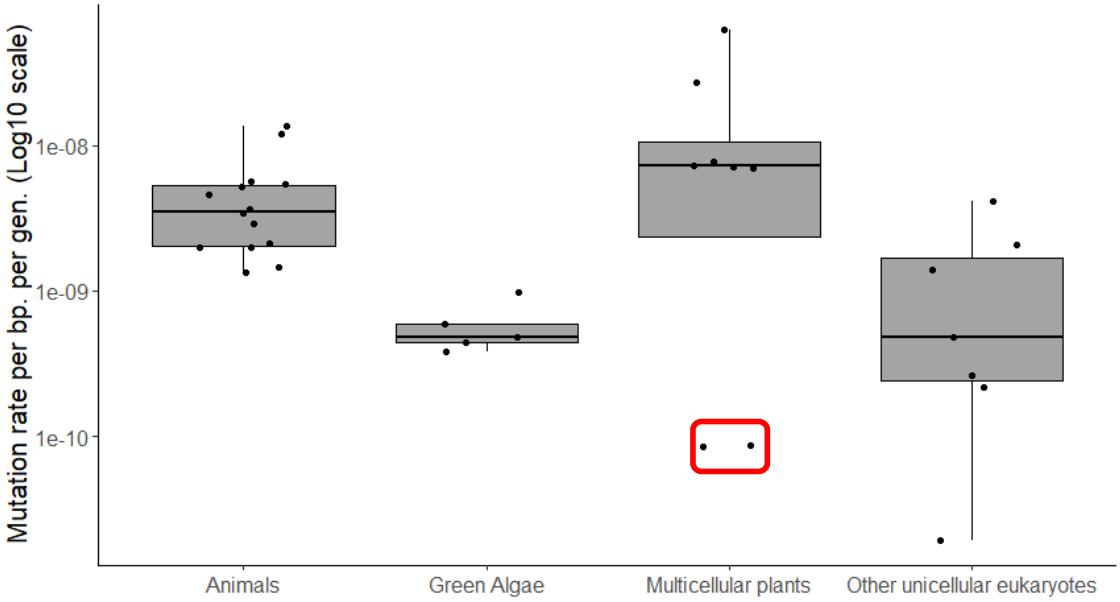
463 2020). In their study, Yu et al. uncovered a large class of reference biased SNPs that were
464 present in some but not all sampled *Z. marina* ramets. By reconstructing the genealogical
465 relationship of the sampled ramets, the authors demonstrated that such SNPs changed in
466 frequency during vegetative growth, with some reaching heterozygous fixation in specific
467 ramets. The authors argued that this data was well explained by a model of "somatic drift"
468 whereby *de novo* mutations arise at low frequency within a single cell lineage before
469 ultimately either reaching fixation or being lost.

470 Calculating mutation rates in organisms with multiple segregating cell lineages poses
471 a technical challenge due to the difficulty of assessing power when mutations are present in
472 only a subset of the cells of an individual. We implemented a method that takes into
473 account the fraction of reads we expect to support a *de novo* mutation by using observed
474 mapping patterns of putatively *de novo* mutations, similar to the approach used by Hiltunen
475 et al. (2019). This method considers the fact that we have reduced power to detect recently
476 arisen mutations that are at low frequency within their MA lines, giving a more accurate
477 estimate of the SNP mutation rate. This approach is an admittedly crude attempt to address
478 the problem. Rather than assuming the expected frequency of the mutant allele is 50%, we
479 simply choose a lower value based on the observed coverage at putative *de novo* mutations.
480 In reality, this lower value is unknown and will differ among mutations depending upon
481 when each mutation arose, i.e., there is an unknown distribution of mutation frequencies at
482 the end of the MA experiment. Nonetheless, our approach is an improvement over
483 completely ignoring the issue. Moreover, the variation in mutation rate estimates inferred
484 using different values for the assumed expected mutant frequency provides some sense of
485 the sensitivity of these estimates to this assumption. However, the issue of the unknown
486 distribution of mutation frequencies adds a caveat to the between-species comparison.
487 Because of the moderate difference in coverage between species, which affects the power
488 to detect mutations segregating at different true frequencies, the estimates for the two
489 species may be somewhat differentially affected by any bias introduced by our method.

490 A natural consequence of the lack of a single cell phase is that from a population
491 genetics perspective, the mutation rate becomes a harder parameter to interpret. On one
492 hand, we may be interested in calculating the mutation rate that captures every new
493 mutation that has arisen in a clonal bud. Alternatively, from an evolutionary perspective, we
494 might be only be interested in mutations that will not only arise in a clonal bud but also

495 persist in a future clonal descendant such that they contribute to population level genetic
496 diversity. Aside from random inter-cell lineage "drift", cell lineage selection may bias which
497 *de novo* mutations are eventually lost in a clonal lineage (Otto and Orive 1995), although
498 this process likely has a minimal effect on mutation rate estimates in most mutation
499 accumulation studies. In principle, we might be able to calculate an "evolutionarily
500 relevant" mutation rate by performing long term mutation accumulation experiments
501 allowing the majority of *de novo* mutations to either be lost or to have fixed within a clonal
502 lineage such that all cells in any clone will be either fully homozygous (in the case of loss) or
503 heterozygous (in the case of fixation). However, from a practical standpoint, it is hard to
504 know *a priori* how many generations will be necessary for this process to occur. In practice,
505 we could attempt to estimate the frequency of *de novo* mutations that have become
506 sufficiently common such that they are likely to not be lost before fixation leading to an
507 estimate of an evolutionarily relevant mutation rate. For example, we could assume that
508 mutations found in at least 50% of cells (i.e. at a frequency of at least 25% in a clonal bud)
509 are more likely than not to eventually fix in their clonal lineage, being found in 100% of cells
510 in some future generation. In reality, some of these mutations are still likely to be lost
511 before they have a chance to fix while simultaneously some mutations that are below 50%
512 frequency might still reach fixation in the future. More generally, the lower the frequency
513 cut-off we use for mutations that are "likely" to fix, the better the chance that we capture
514 all of the mutations that will reach fixation with the caveat that we will also be capturing
515 more mutations that will eventually be lost. In our study we opted to use a read number
516 cut-off rather than a frequency cut-off as we were primarily concerned about differentiating
517 true *de novo* mutations from false positives, a task that is particularly challenging in
518 organisms with low mutation rates. In practice, our filtering criteria results in us mostly
519 identifying mutations with an allelic frequency of at least 20% (61/71 *de novo* mutation
520 reported in our study). As mentioned above however, some of these mutations may still be
521 lost prior to fixation within an organism meaning that the mutation rate we report here is
522 likely inflated compared to a true "evolutionarily relevant" rate. To some extent, this
523 upward bias is counter-balanced by mutations that could eventually go to fixation but are at
524 too low frequency at the time of sequencing to be either observed or even inferred by our
525 power calculation that is based on some threshold frequency (10-50%; Tables S2-5).

526 Our estimate of the mutation rate in *S. polyrhiza* (8.39E-11 per bp per gen.) is similar
527 to the estimates reported by Xu et al. (7.92E-11 per bp per gen.), however there are two
528 important differences between our studies. First, our MA experiment was conducted only in
529 the lab, while Xu et al. placed MA lines both in the lab and in the field, observing no
530 mutations in the lab setting, likely due to the smaller number of MA generations in their
531 study. Second, Xu et al. used ancestral heterozygous sites to estimate their power to detect
532 *de novo* mutations which are not strongly reference biased in a similar manner. This
533 suggests that the estimate from Xu et al., while conducted in a more naturally realistic
534 environment, may be an underestimate of the mutation rate in the field.



535
536

537 ***Figure 5*** *Comparison of per base pair, per generation mutation rates between duckweed and*
538 *other eukaryotic species. Duckweed estimated highlighted by red box. References:* (Ossowski
539 *et al.* 2010; Lynch 2010; Denver *et al.* 2012; Schrider *et al.* 2013; Weller *et al.* 2014; Zhu *et*
540 *al.* 2014; Venn *et al.* 2014; Keightley *et al.* 2015; Uchimura *et al.* 2015; Yang *et al.* 2015;
541 Farlow *et al.* 2015; Ness *et al.* 2015; Smeds *et al.* 2016; Xie *et al.* 2016; Besenbacher *et al.*
542 2016 p.; Feng *et al.* 2017; Liu *et al.* 2017; Oppold and Pfenninger 2017; Flynn *et al.* 2017;
543 Krasovec *et al.* 2017, 2018, 2019; Hanlon *et al.* 2019; Orr *et al.* 2020)

544 The estimates of the per generation, per base pair mutation rate in *S. polyrhiza* and
545 *L. minor* are among the lowest so far for multicellular eukaryotes (Figure 5 and Table S6).
546 One potential explanation for why duckweeds have lower mutation rates than other plants
547 is the smaller number of cell divisions they likely go through compared to their larger, long-

548 lived relatives. Mutation rate studies in trees have indeed shown that while per generation
549 tree mutation rates are high (on the order of 1E-08 mutations per bp), mutation rates per
550 unit growth (a proxy for number of cell divisions), must be several orders of magnitude
551 lower (Xie *et al.* 2016; Hanlon *et al.* 2019; Orr *et al.* 2020). However, duckweeds do appear
552 to exhibit a low mutation rate compared to animals which have limited cell divisions
553 between meiotic events due to a segregated germline. Moreover, our per generation
554 duckweed mutation rate estimates fall on the lower end of values seen in green algae and
555 other unicellular eukaryotes, organisms which unlike duckweed only go through a single cell
556 division per generation (Figure 5). Overall, these patterns fit with previous work that has
557 suggested that number of cell divisions per generation alone is not enough to fully explain
558 variation in mutation rates (Lynch 2010).

559 The low mutation rate observed in our study is consistent with efficient selection
560 against mutator alleles in highly asexual organisms such as duckweed, bacteria and
561 unicellular eukaryotes (Kimura 1967; Leigh 1970). This explanation would also predict that
562 the mutation rate should be higher in *L. minor* than in *S. polyrhiza* as genomic analyses and
563 field observations suggest that *L. minor* outcrosses more frequently than *S. polyrhiza*
564 (Vasseur *et al.* 1993; Ho 2018; Xu *et al.* 2019; Ho *et al.* 2019). However, we did not observe a
565 significant difference in mutation rate between species in our study. This could either be
566 because we did not have the power to differentiate between such overall low mutation
567 rates, because the difference in rates of sexual reproduction is not large enough between
568 duckweed species, because a difference in rates of sex has arisen in recent history, or
569 because factors other than reproductive mode play a larger role in shaping mutation rate
570 evolution. The overall low mutation rate in both species is in contrast to the theoretical
571 prediction that strong linkage in asexual genomes can allow mutators to fix in asexual
572 populations if they hitchhike to fixation with beneficial mutations they produce (André and
573 Godelle 2006). Population genomic analyses in duckweed have shown that selection on
574 protein coding genes is weak as evidenced by elevated measures of π_N/π_S (Ho 2018; Xu *et al.*
575 2019; Ho *et al.* 2019). This suggests that strongly beneficial mutations might be too rare to
576 allow mutators to be selected for in duckweed.

577 Xu et al. inferred a global N_e for *S. polyrhiza* of ~1x10E-06 so it might also be possible
578 that the large effective population sizes of these species allow selection to be efficient
579 enough to lower the mutation rate more than in most multicellular eukaryotes (Sung *et al.*

580 2012; Lynch *et al.* 2016). This explanation, however, is also inconsistent with the fact that
581 measures for the efficacy of selection suggest that selection is weak in duckweed, likely due
582 to the predominance of asexual reproduction (Xu *et al.* 2019; Ho *et al.* 2019).
583 In conclusion, we report a very low SNP mutation rate in two species of duckweed
584 consistent with previous results in this group. We found that *de novo* mutations appear at
585 low frequencies within MA lines suggesting the presence of multiple segregating cell
586 lineages. We then used an approach that allows us to estimate the mutation rate when
587 multiple cell lineages are transmitted across generations. The low mutation rate of these
588 duckweeds is consistent with the idea that a higher degree of asexual reproduction leads to
589 strong selection for low mutation rates.
590
591 **Acknowledgements**
592 This work was supported by the Natural Sciences and Engineering Research Council of
593 Canada (AFA and SIW). We thank Mitch Cruzan for helpful suggestions with data analysis.
594 We thank Jade Lavallee and Victor Mollov for help with maintaining MA lines.
595
596 **Author contributions**
597 SIW and AFA conceived the project. AFA, SIW, and MB collected samples. MB performed the
598 MA experiment. GS did bioinformatics analyses. AFA, SIW, GS wrote the manuscript. All
599 authors contributed to manuscript editing. SIW and AFA contributed equally to this work.
600
601

Literature Cited:

André, J.-B., and B. Godelle, 2006 The evolution of mutation rate in finite asexual populations. Genetics 172: 611–626.

Appenroth, K.-J., S. Teller, and M. Horn, 1996 Photophysiology of turion formation and germination in Spirodela polyrhiza. Biol. Plant. 38: 95–106.

Bell, G., 1982 *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. University of California Press, Berkeley

Besenbacher, S., P. Sulem, A. Helgason, H. Helgason, H. Kristjansson *et al.*, 2016 Multi-nucleotide de novo mutations in humans. PLOS Genet. 12: e1006315.

Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe *et al.*, 2013 Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31: 213–219.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6: 80–92.

Cole, C. T., and M. I. Voskuil, 1996 Population genetic structure in duckweed (Lemna minor, Lemnaceae). Can. J. Bot. 74: 222–230.

Denver, D. R., L. J. Wilhelm, D. K. Howe, K. Gafner, P. C. Dolan *et al.*, 2012 Variation in base-substitution mutation in experimental and natural lineages of Caenorhabditis nematodes. Genome Biol. Evol. 4: 513–522.

Dorai-Raj, S., 2014 binom: *Binomial confidence intervals for several parameterizations*. R package, version 1.1-1, URL https://cran.r-project.org/web/packages/binom/index.html

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8: 610–618.

Farlow, A., H. Long, S. Arnoux, W. Sung, T. G. Doak *et al.*, 2015 The spontaneous mutation rate in the fission yeast Schizosaccharomyces pombe. Genetics 201: 737–744.

Feng, C., M. Pettersson, S. Lamichhaney, C.-J. Rubin, N. Rafati *et al.*, 2017 Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. eLife 6:.

Flynn, J. M., F. J. J. Chain, D. J. Schoen, and M. E. Cristescu, 2017 Spontaneous mutation accumulation in Daphnia pulex in selection-free vs. competitive environments. Mol. Biol. Evol. 34: 160–173.

Foster, P. L., 2007 Stress-induced mutagenesis in bacteria. Crit. Rev. Biochem. Mol. Biol. 42: 373–397.

Hanlon, V. C. T., S. P. Otto, and S. N. Aitken, 2019 Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis. Evol. Lett. 3: 348–358.

Hicks, L. E., 1932 Flower Production in the Lemnaceae. Ohio Journal of Science 32: 115–132.

Hiltunen, M., M. Grudzinska-Sterno, O. Wallerman, M. Ryberg, and H. Johannesson, 2019 Maintenance of high genome integrity over vegetative growth in the fairy-ring mushroom Marasmius oreades. Curr. Biol. 29: 2758-2765.e6.

Ho, K. H. E., 2018 The effects of asexuality and selfing on genetic diversity, the Efficacy of selection and species Persistence. [Thesis, University of Toronto].

Ho, E., M. Bartkowska, S. I. Wright, and A. Agrawal, 2019 Population genomics of the facultatively asexual duckweed Spirodela polyrhiza. New Phytol. 224: 1361–1371

648 Hodgkinson, A., and A. Eyre-Walker, 2011 Variation in the mutation rate across mammalian
649 genomes. Nat. Rev. Genet. 12: 756–766.
650 Jiang, C., A. Mithani, E. J. Belfield, R. Mott, L. D. Hurst *et al.*, 2014 Environmentally
651 responsive genome-wide accumulation of de novo Arabidopsis thaliana mutations
652 and epimutations. Genome Res. 24: 1821–1829.
653 Keightley, P. D., A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra *et al.*, 2015
654 Estimation of the spontaneous mutation rate in Heliconius melpomene. Mol. Biol.
655 Evol. 32: 239–243.
656 Kimura, M., 1967 On the evolutionary adjustment of spontaneous mutation rates*. Genet.
657 Res. 9: 23–34.
658 Krasovec, M., M. Chester, K. Ridout, and D. A. Filatov, 2018 The mutation rate and the age of
659 the sex chromosomes in Silene latifolia. Curr. Biol. 28: 1832-1838.e4.
660 Krasovec, M., A. Eyre-Walker, S. Sanchez-Ferandin, and G. Piganeau, 2017 Spontaneous
661 mutation rate in the smallest photosynthetic eukaryotes. Mol. Biol. Evol. 34: 1770–
662 1779.
663 Krasovec, M., S. Sanchez-Brosseau, and G. Piganeau, 2019 First estimation of the
664 spontaneous mutation rate in diatoms. Genome Biol. Evol. 11: 1829–1837.
665 Landolt, E., 1986 *Biosystematic investigations in the family of duckweeds (Lemnaceae),*
666 *volume 2. The family of Lemnaceae – a monographic study, volume 1.*
667 Geobotanischen Institutes der ETH, Zurich.
668 Leigh, Egbert Giles, 1970 Natural selection and mutability. Am. Nat. 104: 301–305.
669 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping
670 and population genetical parameter estimation from sequencing data.
671 Bioinformatics. 27: 2987–2993.
672 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
673 transform. Bioinformatics. 25: 1754–1760.
674 Liu, H., Y. Jia, X. Sun, D. Tian, L. D. Hurst *et al.*, 2017 Direct determination of the mutation
675 rate in the bumblebee reveals evidence for weak recombination-associated
676 mutation and an approximate rate constancy in insects. Mol. Biol. Evol. 34: 119–130.
677 Lynch, M., 2010 Evolution of the mutation rate. Trends Genet. 26: 345–352.
678 Lynch, M., M. S. Ackerman, J.-F. Gout, H. Long, W. Sung *et al.*, 2016 Genetic drift, selection
679 and the evolution of the mutation rate. Nat. Rev. Genet. 17: 704–714.
680 Matsuba, C., D. G. Ostrow, M. P. Salomon, A. Tolani, and C. F. Baer, 2013 Temperature,
681 stress and spontaneous mutation in Caenorhabditis briggsae and Caenorhabditis
682 elegans. Biol. Lett. 9: 20120334.
683 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome
684 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
685 sequencing data. Genome Res. 20: 1297–1303.
686 Michael, T. P., D. Bryant, R. Gutierrez, N. Borisjuk, P. Chu *et al.*, 2017 Comprehensive
687 definition of genome features in Spirodela polyrhiza by high-depth physical mapping
688 and short-read DNA sequencing strategies. Plant J. 89: 617–635.
689 Ness, R. W., A. D. Morgan, R. B. Vasanthakrishnan, N. Colegrave, and P. D. Keightley, 2015
690 Extensive de novo mutation rate variation between individuals and across the
691 genome of Chlamydomonas reinhardtii. Genome Res. 25: 1739–1749.
692 Oppold, A.-M., and M. Pfenninger, 2017 Direct estimation of the spontaneous mutation rate
693 by short-term mutation accumulation lines in Chironomus riparius. Evol. Lett. 1: 86–
694 92.

695 Orr, A. J., A. Padovan, D. Kainer, C. Külheim, L. Bromham *et al.*, 2020 A phylogenomic
696 approach reveals a low somatic mutation rate in a long-lived plant. Proc. R. Soc. B
697 Biol. Sci. 287: 20192364.
698 Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark *et al.*, 2010 The
699 rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana.
700 Science 327: 92–94.
701 Otto, S. P., and M. E. Orive, 1995 Evolutionary consequences of mutation and selection
702 within an Individual. Genetics 141: 1173–1187.
703 Plomion, C., J.-M. Aury, J. Amselem, T. Leroy, F. Murat *et al.*, 2018 Oak genome reveals
704 facets of long lifespan. Nat. Plants 4: 440.
705 R Core Team, 2019 R: *A language and environment for statistical computing. R Foundation*
706 *for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.
707 Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011
708 Integrative Genomics Viewer. Nat. Biotechnol. 29: 24–26.
709 Saxena, A. S., M. P. Salomon, C. Matsuba, S.-D. Yeh, and C. F. Baer, 2019 Evolution of the
710 mutational process under relaxed selection in Caenorhabditis elegans. Mol. Biol.
711 Evol. 36: 239–251.
712 Schmid-Siegert, E., N. Sarkar, C. Iseli, S. Calderon, C. Gouhier-Darimont *et al.*, 2017 Low
713 number of fixed somatic mutations in a long-lived oak tree. Nat. Plants 3: 926–929.
714 Schrider, D. R., D. Houle, M. Lynch, and M. W. Hahn, 2013 Rates and genomic consequences
715 of spontaneous mutational events in Drosophila melanogaster. Genetics 194: 937–
716 954.
717 Sharp, N. P., and A. F. Agrawal, 2016 Low genetic quality alters key dimensions of the
718 mutational spectrum. PLoS Biol. 14:.
719 Smeds, L., A. Qvarnström, and H. Ellegren, 2016 Direct estimate of the rate of germline
720 mutation in a bird. Genome Res. 26: 1211–1218.
721 Sturtevant, A. H., 1937 Essays on Evolution. I. On the Effects of Selection on Mutation Rate.
722 Q. Rev. Biol. 12: 464–467.
723 Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, 2012 Drift-barrier
724 hypothesis and mutation-rate evolution. Proc. Natl. Acad. Sci. 109: 18488–18492.
725 Thomas, G. W. C., R. J. Wang, A. Puri, R. A. Harris, M. Raveendran *et al.*, 2018 Reproductive
726 longevity predicts mutation rates in Primates. Curr. Biol. 28: 3193-3197.e5.
727 Uchimura, A., M. Higuchi, Y. Minakuchi, M. Ohno, A. Toyoda *et al.*, 2015 Germline mutation
728 rates and the long-term phenotypic effects of mutation accumulation in wild-type
729 laboratory mice and mutator mice. Genome Res. 25:1125-1134
730 Van Hoeck, A., N. Horemans, P. Monsieurs, H. X. Cao, H. Vandenhove *et al.*, 2015 The first
731 draft genome of the aquatic model plant Lemna minor opens the route for future
732 stress physiology research and biotechnological applications. Biotechnol. Biofuels 8:.
733 Vasseur, L., L. W. Aarssen, and T. Bennett, 1993 Allozymic variation in local apomictic
734 populations of Lemna minor (Lemnaceae). Am. J. Bot. 80: 974–979.
735 Venn, O., I. Turner, I. Mathieson, N. de Groot, R. Bontrop *et al.*, 2014 Strong male bias drives
736 germline mutation in chimpanzees. Science 344: 1272–1275.
737 Wang, L., Y. Ji, Y. Hu, H. Hu, X. Jia *et al.*, 2019 The architecture of intra-organism mutation
738 rate variation in plants. PLoS Biol. 17: e3000191.
739 Watson, J. M., A. Platzer, A. Kazda, S. Akimcheva, S. Valuchova *et al.*, 2016 Germline
740 replications and somatic mutation accumulation are independent of vegetative life
741 span in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 113: 12226–12231.

742 Weller, A. M., C. Rödelsperger, G. Eberhardt, R. I. Molnar, and R. J. Sommer, 2014 Opposing
743 forces of A/T-biased mutations and G/C-biased gene conversions shape the genome
744 of the nematode Pristionchus pacificus. Genetics 196: 1145–1152.
745 Xie, Z., L. Wang, L. Wang, Z. Wang, Z. Lu *et al.*, 2016 Mutation rate analysis via parent–
746 progeny sequencing of the perennial peach. I. A low rate in woody perennials and a
747 higher mutagenicity in hybrids. Proc. R. Soc. B Biol. Sci. 283: 20161016
748 Xu, S., J. Stapley, S. Gablenz, J. Boyer, K. J. Appenroth *et al.*, 2019 Low genetic variation is
749 associated with low mutation rate in the giant duckweed. Nat. Commun. 10: 1243.
750 Yang, S., L. Wang, J. Huang, X. Zhang, Y. Yuan *et al.*, 2015 Parent-progeny sequencing
751 indicates higher mutation rates in heterozygotes. Nature 523: 463–467.
752 Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie *et al.*, 2012 A high-performance
753 computing toolset for relatedness and principal component analysis of SNP data.
754 Bioinforma. Oxf. Engl. 28: 3326–3328.
755 Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov, 2014 Precise estimates of mutation rate
756 and spectrum in yeast. Proc. Natl. Acad. Sci. U. S. A. 111: E2310-2318.
757
758

759 **Supporting Information**

760

761 **Figure S1** Visualization of a low-quality genomic region that was filtered from our analyses.

762

763 **Figure S2** Allelic coverage at heterozygous sites before and after filtering.

764

765 **Figure S3** Visualization of a validated *de novo* mutation from both Illumina short read and

766 Sanger sequencing data.

767

768 **Figure S4** Visualisation of a putative *de novo* mutation that was identified as a likely

769 sequencing error

770

771 **Figure S5** Visualisation of a putative *de novo* mutation that was identified as a likely genome

772 misassembly error

773

774 **Table S1** List of *de novo* mutations identified in two species of duckweed

775

776 **Table S2** Mutation rate estimates under the observed allelic bias at *de novo* mutations

777

778 **Table S3** Mutation rate estimates under several scenarios of allelic bias at *de novo* mutant

779 sites in salt stressed mutation accumulation lines

780

781 **Table S4** Mutation rate estimates under several scenarios of allelic bias at *de novo* mutant

782 sites in control mutation accumulation lines

783

784 **Table S5** Number and fraction of callable sites under several scenarios of allelic bias at *de*

785 *novo* mutant sites

786

787 **Table S6** Summary of mutation rate estimates retrieved from the literature

788