High rates of spontaneous chromosomal duplications unravel dosage compensation by translational regulation

Marc Krasovec[1*], Remy Merret[2,3,*], Frédéric Sanchez[1], Sophie Sanchez-Brosseau[1], Gwenaël Piganeau[1*]

[1] Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650 Banyuls/Mer, France

[2] Centre National de la Recherche Scientifique, Laboratoire Génome et Développement des Plantes, UMR5096, 66860 Perpignan, France

[3] Université de Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, UMR5096, 66860 Perpignan, France

[*] Corresponding authors: krasovec@obs-banyuls.fr ; remy.merret@univ-perp.fr; gwenael.piganeau@obs-banyuls.fr

Keywords: chromosome duplication, mutation accumulation, aneuploidy, dosage compensation, translation efficiency, phytoplankton.

Running title: Spontaneous chromosome duplication rates.

ABSTRACT (203 words)

While duplications have long been recognized as a fundamental process driving major evolutionary innovations, direct estimations of spontaneous chromosome duplication rates, leading to aneuploid karyotypes, are scarce. Here, we provide the first estimations of spontaneous chromosome duplication rates in six unicellular eukaryotic species from mutation accumulation (MA) experiments. The spontaneous chromosome duplication rates reach $1x10^{-4}$ to $1x10^{-3}$ per genome per generation, which is ~4 to ~50 times less frequent than spontaneous point mutations per genome, whereas chromosome duplication events can affect 1 to 7% of the total genome size. Comparative transcriptomics between MA lines with different chromosome duplications reveals a strong positive correlation between RNA expression rate and DNA copy number. However, comparative analyses of the translation rate of mRNAs estimated by polysome profiling unravel a chromosome specific dosage compensation mechanism. In particular, one chromosome with a gene average of 2.1 excess of mRNAs is compensated by an average of ~0.7 decrease in translation rates. Altogether, our results are consistent with previous observations of a chromosome dependent effect of dosage compensation and provide evidence that it may occur during translation. These results support the existence of a yet unknown post-transcriptional mechanism orchestrating the modification of translation of hundreds of transcripts from genes located on duplicated regions in eukaryotes.

INTRODUCTION

Complete or partial chromosome duplications leading to aneuploidy karyotypes are known to be at the origin of genetic diseases, such as Trisomy 21 in humans, and is a near universal feature of tumor cells (Rajagopalan and Lengauer 2004). When a single chromosome is duplicated, an immediate issue arises: the imbalance of ploidy and gene dose within a same karyotype. This is expected to have deleterious effects because it creates an imbalance of the transcript and protein productions, which is costly and may disrupt the function of a pathway and protein interactions (Dephoure et al. 2014; Veitia and Potier 2015). The deleterious effects of aneuploidy have been documented in many different biological model systems such as yeast (Torres et al. 2007), mouse and human (Gearhart et al. 1987). In *Caenorhabditis elegans*, mutation accumulation experiments provided evidence of purifying selection against gene duplications causing excess of transcripts, as compared to gene duplications associated to invariant transcript levels (Konrad et al. 2018). Moreover, the consequences of aneuploidy on gene expression are complex and while gene expression may increase with chromosome copy number, gene expression changes may also spread outside the duplicated regions in *Arabidopsis* (Hou et al. 2018; Song et al. 2020), *Drosophila* (Devlin et al. 1988), and human cells (FitzPatrick et al. 2002). Contrasting with the before mentioned studies, aneuploidy may confer a selective advantage as a response to stress, such as resistance to drugs in *Candida albicans* (Selmecki et al. 2006) or *Saccharomyces cerevisiae* (Chen et al. 2012). Recent evidence about a lack of deleterious effect of aneuploidy is the high aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages (Scopel et al. 2021), a fifth of the sequenced strains harboring atypical aneuploidy karyotypes (Peter et al. 2018). Indeed, different dosage compensation mechanisms have evolved and may restore the ancestral gene dose leading to aneuploidy tolerance. The three targets of dosage compensation mechanisms of duplicated genes are the modification of the transcription rate, the translation rate, or the protein degradation rate of duplicated genes. The most studied mechanisms are those involved in modifying the transcription rate during the evolution of heterogametic sex chromosomes, known both in plants (Muyle et al. 2012; Muyle et al. 2017; Charlesworth 2019) and animals (Disteche 2012; Graves 2016). Different mechanisms evolved either by simulating the ancestral ploidy by doubling the expression of the genes of the single copy male X (Baker et al. 1994); or by equalizing the ploidy between the two sexes by halving the

expression – silencing of one X in female (Heard et al. 1997). Although dosage compensation is well studied in sex chromosome evolution, whether and how it evolves after a chromosome ploidy variation is unclear, particularly for autosomes (Kojima and Cimini 2019). Previous studies did not report consistent results, with a significant increase of expression for the duplicated genes in *C. albicans* (Selmecki et al. 2006), *Drosophila* (Loehlin and Carroll 2016), *S. cerevisiae* (Torres et al. 2007), mammalian cells (Williams et al. 2008); and a decrease of the expression of the two copies in yeast and mammals, suggesting a compensation at the transcriptional level (Henrichsen et al. 2009; Qian et al. 2010). In the case of chromosome duplication, evidence for dosage compensation at the transcriptional level is scarce (Stenberg et al. 2009; Hose et al. 2020), and scaling of gene expression with gene copy number has been reported in disomic yeasts (Kaya et al. 2020). Interestingly, dosage compensation has been reported to occur at the post-transcriptional level via the modification of translation efficiency in *Drosophila* (Zhang and Presgraves 2017) or at the post-translational level for 20% of the proteome in yeast via an increased degradation rate of proteins (Dephoure et al. 2014). Increase of protein degradation has also been reported to be involved in compensation mechanism in human Downs syndrome for proteins encoded on the triplicated chromosome 21 and include known stable heteromeric protein complexes (Liu et al. 2017).

Although the consequences of chromosome duplication on transcription rates - and to a lesser extent their consequences on translation rates and protein abundance - have been investigated in many model organisms, our knowledge of the rate of chromosome duplication events in eukaryotes have been yet restricted to *Saccharomyces cerevisiae* (Lynch et al. 2008; Zhu et al. 2014; Liu and Zhang 2019) and humans (Nagaoka et al. 2012; Loane et al. 2013). Here, we investigate the spontaneous chromosome duplication rate in six unicellular photosynthetic species by analyzing mutation accumulation (MA) experiment genomes. The principle of MA experiments is to follow the descendants originated from a single cell under minimal selection insured by serial bottlenecks during dozens to thousands cell divisions (Halligan and Keightley 2009). Here, MA experiments have been performed by maintaining 12 to 40 MA lines for a total of 1,595 to 17,250 generations depending of the species (Krasovec et al. 2017; Krasovec, Sanchez-Brosseau, et al. 2018; Krasovec et al. 2019). The six species include five Chlorophyta (one Trebouxiophyceae et four Mamiellophyceae) and one Bacillariophyta, ecological

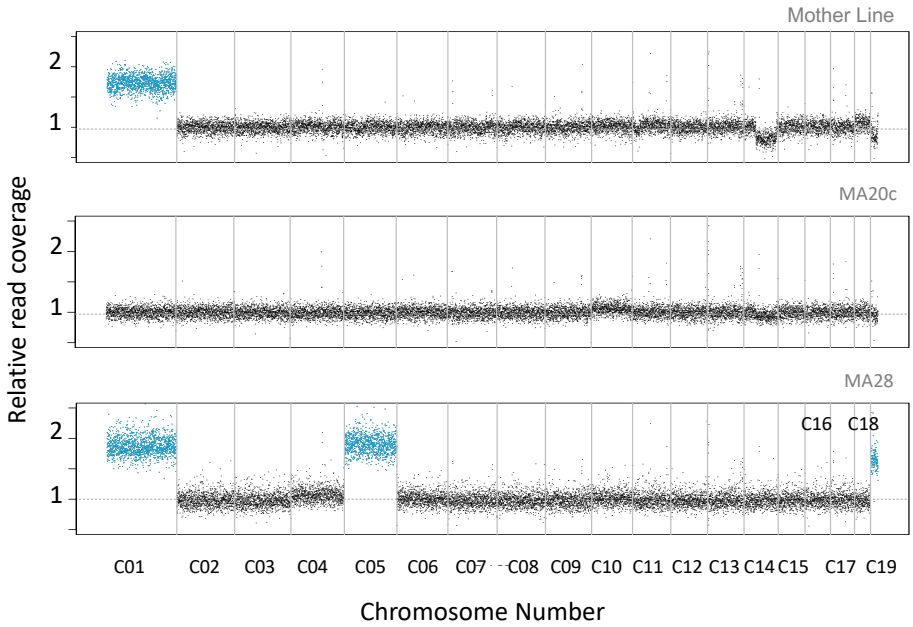
107 relevant primary producers in the sunlit ocean (de Vargas et al. 2015), with a large
108 phylogenetic spread encompassing 1.5 billion years of divergence (Yoon et al. 2004).
109 We further recovered cryopreserved MA lines to investigate the consequence of
110 chromosome duplication on the transcription and translation rates.
111

RESULTS

Whole chromosome duplication rate

In this study, we analyzed whole genome resequencing data from previous MA experiments (Krasovec et al. 2016; Krasovec et al. 2017; Krasovec, Sanchez-Brosseau, et al. 2018) in five haploid green algae species (Chlorophyta, four Mamiellophyceae and one Trebouxiophyceae) and one diploid diatom (Bacillariophyta) : *Picochlorum costavermella* RCC4223 (Krasovec, Vancaester, et al. 2018), *Ostreococcus tauri* RCC4221 (Blanc-Mathieu et al. 2014), *O. mediterraneus* RCC2590 (Subirana et al. 2013), *Bathycoccus prasinos* RCC1105 (Moreau et al. 2012), *Micromonas commoda* RCC299 (Worden et al. 2009) and *Phaeodactylum tricornutum* RCC2967 (Bowler et al. 2008; Giguere et al. 2022). Coverage was used as a proxy of copy number (Figure 1A and Supplemental figures S1 to S16), and unveiled whole chromosome duplication events in four of the six species (Table 1 and Table S1). Four whole chromosome duplications were observed in *M. commoda* (chromosomes C05, C12, C16 and C17), four in *B. prasinos* (chromosomes C04, C05, C06 and C19), one in *O. mediterraneus* (chromosome C14) and four in *P. tricornutum* (chromosomes C02, C14 and C23). The chromosome duplication events were mapped onto the genealogies of the MA lines to identify all independent chromosome duplication events (Figure 1B for *B. prasinos*). Genealogy analysis provided evidence that several chromosomes were duplicated two times independently: the MA lines Bp25c and Bp28b carry two independent duplications of chromosome C19; Mc08 and Mc09 carry two independent duplications of chromosome C17; and Pt11 and Pt10c carry two independent duplications of chromosomes C23. The probability of observing two independent whole chromosome duplication of the same chromosome in *B. prasinos*, *M. commoda* and *P. trichornutum* are 0.46, 0.47 and 0.34 respectively (see methods), and these probabilities are thus consistent with an equal probability of duplication across chromosomes. All independent duplication events inferred from coverage analysis and genealogies are summarized in Table 1 and S1. Unexpectedly, the analyses revealed that the ancestral line of the MA experiment in *B. prasinos* carried two copies of the chromosome C01. One of the copies had subsequently been lost 11 times independently over 4,145 generations, corresponding to a spontaneous chromosome loss of a duplicated chromosome of 0.006 per duplicated chromosome per generation in *B. prasinos*.

A. Whole genome read coverage for different mutation accumulation lines



B. Genealogy of the mutation accumulation experiment in *B. prasinos* RCC4222

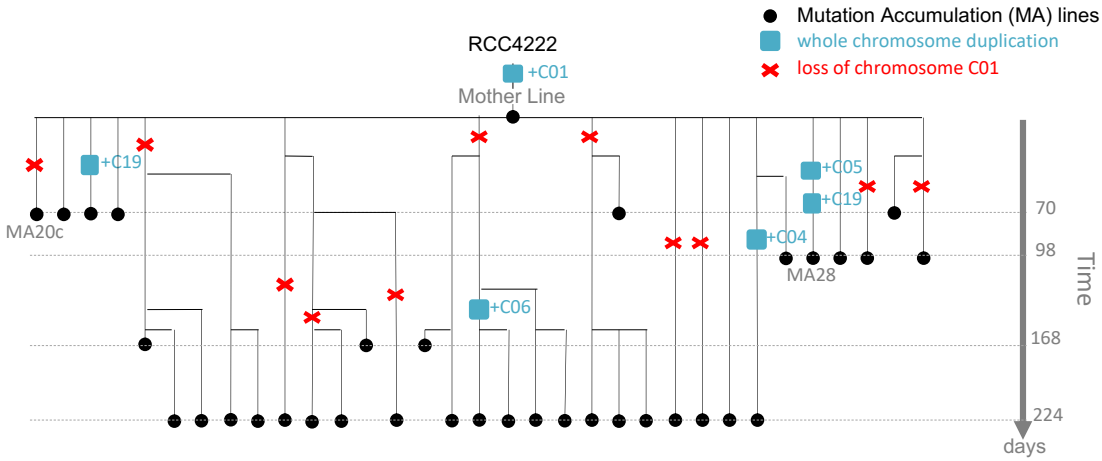


Figure 1. A. Normalized raw genomic coverage of the mother line (T0 of mutation accumulation experiment), Bp20c and Bp28 mutation accumulation lines of *Bathycoccus prasinos*. Grey lines are chromosome separators. In blue are the chromosomes in double copies. Raw coverage of all lines from all species are provided in Figures S1 to S16. **B.** Pedigree of the mutation accumulation lines from the *B. prasinos* experiment. Chromosome C01 is duplicated in the T0 line (named mother line) of the experiment. This duplication is then lost several times, and five other duplications of chromosomes C04, C05, C06 and C19 occurred.

158 **Table 1.** Spontaneous whole chromosome duplication rate in six species. N_{lines} : number of MA lines,
159 Gen : average number of generations per MA line, Tot_{Gen} : total number of generations, N_{Chrom} : number
160 of chromosomes in the ancestral karyotype, N_{WCD} is the number of independent whole chromosome
161 duplications, U_{WCDI} is the whole chromosome duplication rate per chromosome per cell division, and
162 U_{Cell} is the whole chromosome duplication rate per cell division, U_{bs} is the base substitution mutation
163 rate per cell division. The estimation of the upper limit of N_{WCD} *and* U_{cell} in *O. tauri* and *P. costavermella*
164 relies on the assumption of one duplication event.

Species	N_{lines}	Gen	Tot_{Gen}	N_{Chrom}	N_{WCD}	Duplicated Chrom	U_{WCD}	U_{cell}	U_{bs}
Bathycoccus prasinos	35	265	4,145	19	5	C04, C05, C06, C19	0.000063	0.00121	0.0046
Micromonas commoda	37	272	4,994	17	5	C05, C12, C16, C17	0.000059	0.00100	0.0171
Ostreococcus tauri	40	512	17,250	20	0	-	<0.000003	<0.00006	0.0054
Ostreococcus mediterraneus	33	272	8,380	19	1	C14	0.000006	0.00012	0.0064
Picochlorum costavermella	12	133	1,596	10	0	-	<0.00006	<0.00063	0.0119
Phaeodactylum tricornutum	36	181	6,516	25	4	C02, C14, C23	0.000025	0.00061	0.0132

165

166

167 ***Consequences of chromosome duplication on transcription***

168 To explore the consequences of the whole or partial chromosome duplication on
169 transcription, we estimated the level of transcription of a control strain (*B. prasinos*
170 RCC4222) and one MA line of *B. prasinos* by recovering one 4-year-old cryopreserved
171 MA line. This cryopreserved culture originated from MA line Bp37, and the recovered
172 culture is hereafter named Bp37B. Chromosome copy number of Bp37B and the
173 control line were estimated by whole genome resequencing. This confirmed the two
174 copies of chromosome C04 in Bp37B as in the original Bp37 (Figure 2A and S7), and
175 also revealed additional karyotypic changes: an increase in chromosome C01 copy
176 number (Figure 2A, Figure S17) in Bp37B. The heterogeneous coverage of
177 chromosome C01 led us to divide it into two regions for subsequent analyses; region
178 C01a (1.98 fold coverage) and C01b (1.35 fold coverage). The control line also
179 contained duplicated regions (Figure 2A, Figure S17), chromosome C10 and a region
180 of chromosome C02 named C02b, while C02a is in single copy. Limits of sub-regions
181 of chromosomes C01 and C02 are provided in Table S2. We interpreted coverage
182 values smaller than 2 (Figure S17), *e.g.* C01b in Bp37B, C10 and C02b as duplications
183 that are carried by a subpopulation of cells. Cultures were grown to up to 10 million
184 cells per ml prior to extraction so that polymorphism is not unexpected.

185 Comparative transcriptome analyses of Bp37B and the control line revealed that there
186 are on average twice the number of transcripts for genes located on the duplicated
187 chromosomes as compared to the control line - C04 (transcription rate *tr(i)*

188 average=2.13, median=1.95, estimated from TPM, Figure 2B, raw data in Table S3)
189 and C01a (transcription rate *tr(i)* average=2.29, median=2.02). The transcription rate
190 of genes on chromosomes C10 and C02b is also affected. Altogether, the transcript
191 production scales up with the chromosome copy number in the Bp37B and control
192 lines (Pearson correlation, rho=0.87, p-value<0.001, Figure 3A).
193
194 ***Gene-by-gene variation of the consequences of DNA copy number***
195 Dosage invariant genes (Antonarakis et al. 2004; Lyle et al. 2004) are genes for which
196 the transcription is not affected by copy number, and they may be involved in
197 aneuploidy tolerance. We investigated differential gene expression at the individual
198 gene scale with Deseq2 (Love et al. 2014) to identify candidate invariant genes. We
199 found that the transcription rate of 158 out of the 1776 genes in two copies (9%)
200 located on duplicated chromosomes or regions were not significantly higher than for
201 genes in single copy (Table S4). Two Gene Ontology categories (RNA processing,
202 GO:0006396 and DNA metabolic process GO:0006259) were over-represented in this
203 transcriptomic invariant gene set. First, genes involved in RNA processing (3.8 times
204 more frequent, p-value <0.01) including members of heteromeric protein complexes
205 such as subunits 7 and 2 of the U6 snRNA-associated Sm-like protein (Table S5,
206 invariant 158 annotations). Second, the genes involved in DNA metabolic process (2.2
207 times more frequent in the invariant gene subset, p-value <0.04) including the DNA
208 Polymerase A and the DNA primase large subunit. Notable protein complex members
209 of the invariant data set are Histone 3 and the subunit E of the translation initiation
210 factor 3. There are in total 12 genes annotated as subunits in the subset of invariant
211 genes, this is significantly more than the frequency of protein coding genes annotated
212 as subunits in the complete proteome (Fisher exact test p=0.0006). This suggests that
213 a higher proportion of these genes coding for protein forming complexes have a gene-
214 specific tailored regulated transcription that is not affected by gene copy number.
215

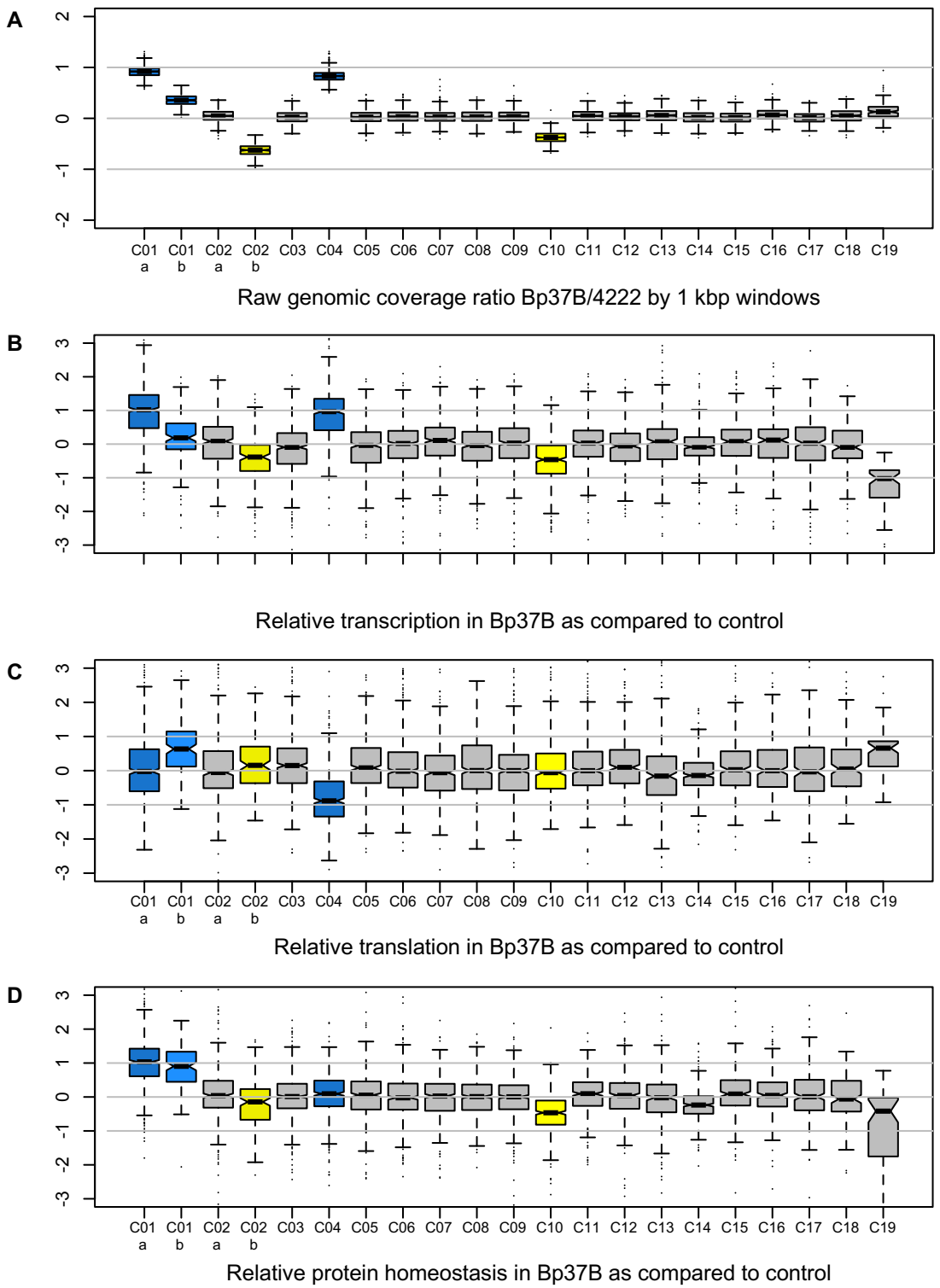


Figure 2. A. Genomic coverage of 1 kbp windows along the genome of the MA line *Bathycoccus prasinos* Bp37B. **B.** Distribution of genes transcription rate *tr(i)* of the 19 chromosomes. **C.** Distribution of genes translation efficiency *te(i)* of the 19 chromosomes. **B** and **C**. The transcription rate and translation efficiency ratio averages are different between chromosomes (ANOVA, p-value<10^{-15}). **D.** Distribution of genes protein homeostasis (transcription rate x translation rate) of the 19 chromosomes. Y axis are in log2. Blue: duplications in Bp37B. Yellow: duplications in control RCC4222.

224 ***Consequences of chromosome duplication on translation***

225 The compelling evidence of overexpression of the majority of duplicated genes
226 prompted us to investigate whether post-transcriptional processes may temperate this
227 excess of transcripts. This second hypothesis was tested in the same *B. prasinos* MA
228 line by sequencing mRNAs associated with ribosomes (polysomes) in order to
229 compare the translation efficiency *te(i)* of genes located on duplicated and non-
230 duplicated lines, by estimating the relative proportion of mRNA binded to a ribosome
231 in single versus duplicated genes. We found that the average translation efficiency of
232 the genes located on the duplicated chromosome C04 was 0.71 (median=0.54) as
233 compared to the genes on this chromosome in the control line (Figure 2C). At
234 chromosome scale (Table S7), the translation rate is 0.67 for the chromosome C04.
235 However, the translation efficiency of part C01a of the chromosome C01 was 0.93 and
236 1.54 for the part C01b. The translation efficiency of the two parts of the chromosome
237 C02 are 1.04 and 0.76 for C02a and C02b, respectively, and 0.99 for chromosome
238 C10. Translation rate modification is thus chromosome dependent (Figure 3): genes
239 located on the C04 and C01b display an important variation in translation efficiency
240 between the two lines. Last, we estimated the expected protein production in Bp37B
241 as compared to the control line by going back to the absolute translation rate for each
242 gene: that is the ratio of mRNA in polysomes in Bp37B as compared to the control
243 (Figure 2D). This predicted that genes on C04 have a similar protein production rate
244 in Bp37B and in the control line, despite a higher transcription rate as a consequence
245 of the chromosome duplication, as well for the two parts of the chromosome C02
246 (Figure 2D). However, the two parts of the chromosome C01, despite different
247 transcription and translation efficiency rates both leads to a protein level that is the
248 double of the control line. The chromosome C10 also exhibits a lower protein content
249 in Bp37B compared to control, meaning there is an excess of proteins from genes
250 linked to this chromosome in the control line because of the duplication. Altogether,
251 we observed there is no dosage compensation at the transcription level for any of the
252 duplicated regions, but that dosage compensation may occur at the translation level
253 on some duplicated regions and seems chromosome dependent (Figure 3).
254
255

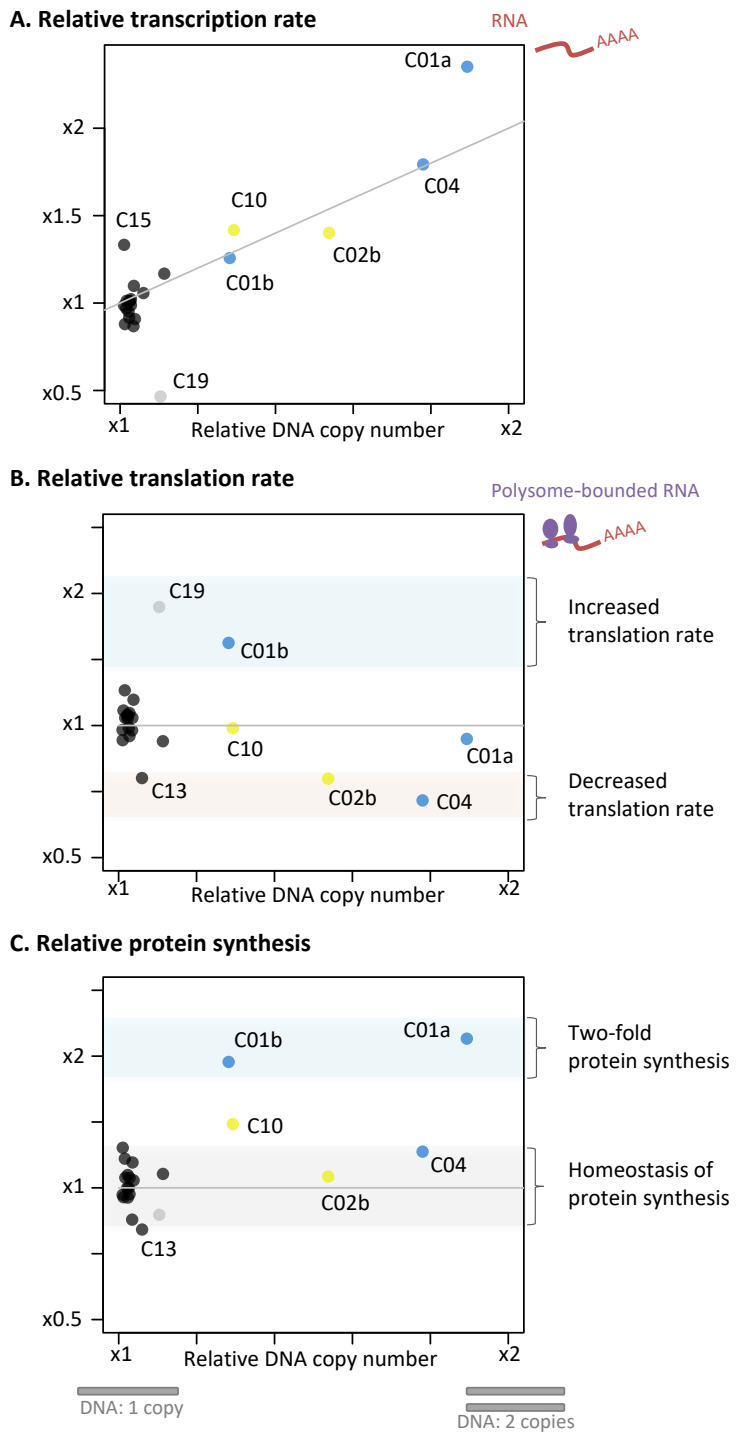


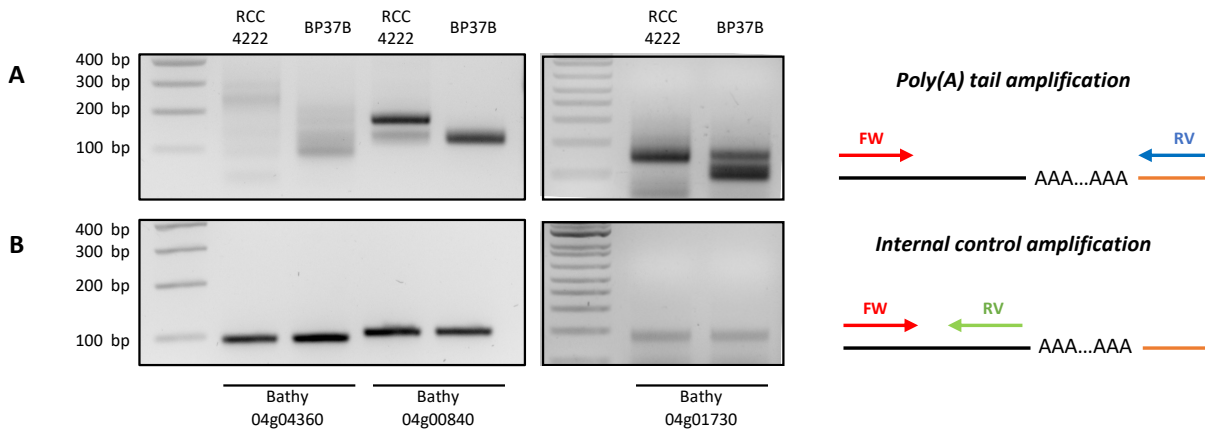
Figure 3. Raw relative genomic coverage average (Bp37B/Control) related to average transcription rate *tr(i)* (**A**), translation efficiency *te(i)* (**B**), and average protein synthesis (average transcription rate x average translation efficiency) per chromosome (**C**). Blue: chromosomes duplicated in Bp37B; Yellow: chromosomes duplicated the control RCCC4222; Black: non-duplicated chromosomes; Grey: outlier chromosome C19. Data are provided in Table S6 and S7.

265 ***Searching for independent molecular signatures of translation modulation***

266 The observed dosage compensation at the translation level on chromosome C04
267 suggests that a post-transcriptional mechanism occurs on duplicated transcripts. As
268 the length of the poly(A) tail is a key feature of many cytoplasmic mRNAs and is known
269 to regulate translation (Subtelny et al. 2014; Eichhorn et al. 2016; Lim et al. 2016), we
270 tested if DC observed for the translation of transcripts linked to duplicated
271 chromosome C04 is associated with a variation of poly(A) tail length. Using 3'RACE
272 experiment, we measured the poly(A) tail length in Bp37B and control (Figure 4 and
273 S18) of the transcripts of three genes located on chromosome C04 (04g00840,
274 04g01730 and 04g04360), one gene located on chromosome C08 (08g03470) and
275 C01 gene located on chromosome C01a (01g01300). We systematically observed that
276 the poly(A) tail length is reduced in Bp37B line compared to control line for transcripts
277 located on chromosome C04 (Figure 4), whereas no difference is observed control
278 (Figure S18).

279

280



281

282 **Figure 4.** Transcripts from genes located on duplicated chromosome C04 present shorter poly(A) tail.
283 Poly(A) tail measurement was performed using modified 3'RACE. PCR amplification was performed
284 using primers flanking poly(A) tail (**A**) or primers anchored in 3'UTR just before poly(A) tail as internal
285 control (**B**). Experiments was performed using total RNA from RCC4222 line (control line) and BP37B
286 line (with duplicated chromosome C04). Illustrations representing PCR amplification are present on the
287 right panel. Black line represents the messenger RNA and the orange line represent the ligated adapter
288 used for reverse transcription and PCR amplification. FW. Forward primer. RV. reverse primer.

289

290

291

292

293 **DISCUSSION**

294 ***Rates of whole chromosome duplication in eukaryotes***

295 The rates of spontaneous whole chromosome duplication reported here vary between
296 $6x10^{-5}$ to $1x10^{-3}$ events per haploid genome per generation and are thus one to two
297 orders of magnitude lower than spontaneous point mutations per genome per
298 generation. Theoretically, there are two possible mechanisms at the origin of
299 chromosome duplication i) the supernumerary replication of one chromosome before
300 cell division, leading to one cell with one chromosome and one cell with two
301 chromosomes; or ii) the unequal segregation of chromosomes during cell division,
302 leading to one cell with two copies and one cell without any copy. The latter scenario
303 has been intensively experimentally studied in human cells (Ford and Correll 1992;
304 Cimini et al. 2001), and seems all the more likely in the extremely small-sized
305 Mamiellophyceae cells (1 µm cell diameter) as it has been reported that they contain
306 less kinetochore microtubules than chromosomes (Gan et al. 2011), which may induce
307 high error rate in the chromosome segregation process. However, the whole
308 chromosome duplication rates reported in Mamiellophyceae are in line with those
309 found in the diatom *P. tricornutum* and in yeast ($9.7x10^{-5}$) (Zhu et al. 2014), the latter
310 containing approximately as many kinetochore microtubules as chromosomes
311 (Peterson and Ris 1976). Altogether, the spontaneous whole chromosome duplication
312 rates reported previously in yeast (Zhu et al. 2014) and in the species from
313 evolutionary distant eukaryotic lineages reported here (Yoon et al. 2004) suggest a
314 high rate of spontaneous aneuploidy across the eukaryotic tree of life. To investigate
315 whether spontaneous whole chromosome duplication had been overlooked in other
316 lineages, we screened publicly available resequencing data of MA on the freshwater
317 green algae *Chlamydomonas reinhardtii* (Ness et al. 2015), and identified one event
318 for chromosome 14 in the strain CC1373. We also provided evidence of spontaneous
319 whole chromosome duplication (four chromosomes in a single individual from a
320 pedigree) in the brown algae *Ectocarpus* (Krasovec et al. 2022).

321 Interestingly, we were also able to estimate the rate of chromosome duplication loss
322 during one experiment to be about three orders of magnitude higher than the
323 spontaneous duplication rate. This is consistent with previous observation in *S.*
324 *cerevisiae*, and our observations that cells with duplicated chromosomes are
325 ephemeral and unlikely to be maintained for a long time in batch culture. Indeed, the
326 coverage of the large duplication events on C01b, C02b, C04 and C10 were 1.4,1.6,

327 1.8 and 1.4 times the coverage of single copy chromosomes respectively, suggesting
328 that the proportion of cells carrying a supernumerary chromosome within the
329 population is 0.4, 0.6 and 0.8. In addition, we observed the loss of the chromosome
330 C04 duplication in one of the Bp37B lines over 4 months of sub-culturing.
331 As opposed to the aneuploidy generated during meiotic divisions, which has been
332 intensively studied in mammalian cells (Nagaoka et al. 2012), the chromosomal
333 duplication events reported here are generated during mitotic division. Since the
334 molecular mechanisms involved in chromosomal segregation in mitosis, meiosis 1 and
335 meiosis 2, are different, the associated aneuploidy rates are not expected to be equal.
336 The aneuploidy rates in mitotically dividing human cells, such as HCT116 lines, has
337 been estimated to be $7x10^{-2}$ (Thompson and Compton 2008), more than one order of
338 magnitude higher than the highest estimation of $1.2x10^{-3}$ reported here in *B. prasinos*.
339 The observed difference of the spontaneous point mutation rate per genome per
340 generation is expected to vary between species as the consequence of different
341 genome sizes, effective population sizes (Lynch et al. 2016), and the difference
342 between the observed and expected GC content (Krasovec et al. 2017). However, the
343 only eight spontaneous aneuploidy rates yet available are not sufficient to investigate
344 the reasons of chromosome duplication rate variations, such as the impact of
345 chromosome number or effective population size.
346
347 ***Fitness effect of whole chromosome duplication and dosage compensation***
348 The high spontaneous rates of spontaneous chromosome duplication challenge the
349 common idea that whole chromosome duplications are highly deleterious for individual
350 cells. Using the number of cell divisions per days as a proxy of fitness during the MA
351 experiment (Krasovec et al. 2016), only one out of the 15 MA lines with a whole
352 chromosome duplication displayed a significant fitness decrease (Bp28b, Pearson
353 correlation, p-value=0.016, ρ=-0.846, Table S8). Recent studies in yeast are
354 inconsistent with a highly deleterious fitness effect of aneuploidy and point towards
355 slightly deleterious effects, as well as a significant effect of the genetic background on
356 aneuploidy tolerance (Scopel et al. 2021). Compensating point mutations have been
357 previously linked to aneuploidy tolerance in yeast (Torres et al. 2010), notably in the
358 gene encoding the deubiquitinating enzyme UBP6. Point mutations are unlikely to
359 impact the aneuploidy tolerance of the different MA lines reported here, as some MA

360 lines with chromosome duplication do not carry any point mutation (Mc3, Bp28b, Bp26
361 and Bp25), or carry only synonymous mutations and mutations located in intergenic
362 region (*i.e.* Om3 or Mc28, Table S9). The deleterious effect of aneuploidy could be
363 mitigated by a dosage compensation mechanism preserving a relative protein
364 homeostasis. Here in *B. prasinos*, despite the important variation in transcription and
365 translation between genes, we observed that transcription rates overwhelmingly
366 scaled up with chromosome copy number (Figure 3A) and that translation rates
367 differed significantly between chromosomes (Figure 3B). However, chromosome C19
368 displays a unique pattern. This chromosome is a small hypervariable outlier
369 chromosome found in all Mamiellophyceae species sequenced so far (Blanc-Mathieu
370 et al. 2017; Yau et al. 2020). Previous studies reported a strong variation in
371 transcription rates of genes on this chromosome, which is associated with resistance
372 to viruses (Yau et al. 2016) and characterized by lower GC content than other
373 chromosomes.
374 For the duplicated chromosomes or regions, we observed three different patterns of
375 translation efficiency. First, we observed no relative difference in translation rates for
376 genes on duplicated chromosome C10, which is duplicated in 40% percent of cells,
377 and region C01a, which is duplicated in 100% of cells. The inferred excess of protein
378 synthesis of duplicated genes on this chromosome is thus 40% (Figure 3C). Second,
379 we observed a decrease of the translation rate for genes linked to chromosome C04
380 and region C02b. More precisely, the estimated percent of mRNA linked to ribosomes
381 is inversely proportional to the excess of mRNA produced for duplicated genes on
382 these two chromosomes, predicting approximately the same protein production rate
383 as in cells without duplicated chromosomes (Figure 3C). Third, and very surprisingly,
384 we observed an increase of the translation level of genes on C01b, that are in two
385 copies in 40% of cells. As a consequence, the rate of protein synthesis on genes
386 located on region C01b reach the same relative protein synthesis than genes located
387 on region C01a, that is twice the protein synthesis predicted in cells with a single copy
388 of this chromosome. We speculate the absence of dosage compensation on region
389 C01a and the excess of translation on region C01b simulate the ancestral protein
390 homeostasis in the mother line, which contained two entirely duplicated chromosome
391 C01.
392

393 Chromosome-wide regulation of gene translation is relatively unexplored and poorly
394 understood. Indeed, while location on a chromosome may predict a gene's
395 transcription rate as a result of epigenetic marks on the DNA, epitranscriptomics
396 chromosome wide mRNA modifications mechanisms are yet to be discovered.
397 Notwithstanding, chromosome wide translational dosage compensation has been
398 previously observed in *Drosophila* (Zhang and Presgraves 2016), and at the gene
399 scale in several species (Zhang and Presgraves 2017; Chang and Liao 2020).
400
401 ***Poly(A) tail length***
402 The role of poly(A) tail length in the translation efficiency is starting to be better
403 understood (Weill et al. 2012; Subtelny et al. 2014), and alternative polyadenylation is
404 indeed implicated in several processes: transcription termination by RNAP II, mRNA
405 stability, mRNA export and translation efficiency (Zhang et al. 2010; Di Giammartino
406 et al. 2011). In the cytoplasm, the poly(A) tail plays important roles in mRNA translation
407 and stability and the modulation of its length has an important impact on translation
408 efficiency (Subtelny et al. 2014; Eichhorn et al. 2016; Lim et al. 2016). As example,
409 during oocyte maturation and early embryonic development, an increase of poly(A) tail
410 length occurs for particular mRNAs resulting in an increase of translation (Subtelny et
411 al. 2014; Eichhorn et al. 2016; Lim et al. 2016). This modulation of poly(A) tail length
412 has also been found to activate some neuronal transcripts (Udagawa et al. 2012).
413 Recently, it has been suggested that poly(A) tails can be modulated to balance mRNA
414 levels and adjust translation efficiency (Slobodin et al. 2020). In this last case, the
415 CCR4-Not complex shortened the poly(A) tails that reduces the stability of mRNAs.
416 Here, our data suggest that, in a context of chromosome duplication, modulation of
417 poly(A) tail length could be a key post-transcriptional mechanism necessary for
418 dosage compensation at translation level. Future work to explore genome wide
419 analysis of poly(A) tail variation and post-transcriptional RNA modifications such as
420 adenosine methylation (Miao et al. 2022) are poised to clarify the translation
421 compensation mechanisms.
422
423
424
425
426

CONCLUSION

In conclusion, the high prevalence of whole chromosome duplication in five unicellular photosynthetic eukaryotes suggest spontaneous whole chromosome duplication is pervasive in eukaryotes. In one species, *B. prasinos*, we provide evidence that a whole duplication chromosome event is associated with dosage compensation at the post-transcriptional level which might involve the adjustment of poly(A) tail length. These results point to yet unknown post-transcriptional regulation mechanisms in DC of aneuploid karyotypes.

461 **MATERIALS AND METHODS**

462 **Sequencing data from mutation accumulation (MA) experiments**

463 Mutation accumulation (MA) experiments of the six species *Picochlorum*
464 *costavermella* RCC4223 (Krasovec, Vancaester, et al. 2018), *Ostreococcus tauri*
465 RCC4221 (Blanc-Mathieu et al. 2014), *O. mediterraneus* RCC2590 (Subirana et al.
466 2013; Yau et al. 2020), *Bathycoccus prasinos* RCC1105 (synonym to RCC4222)
467 (Moreau et al. 2012), *Micromonas commoda* RCC299 (Worden et al. 2009) and
468 *Phaeodactylum tricornutum* RCC2967 (Bowler et al. 2008; Giguere et al. 2022) were
469 conducted with a flow cytometry protocol described previously for phytoplankton
470 species in liquid medium (Krasovec et al. 2016). Briefly, a mutation accumulation
471 experiment consists in following of MA lines that have evolved from a same cell (the
472 ancestral line) during hundreds of generations. Relaxed selection pressure on
473 spontaneous mutations is ensured by maintaining all MA lines at very low effective
474 population sizes ($6<Ne<8.5$) throughout the experiment (Krasovec et al. 2016). As a
475 consequence, MA experiments enable to estimate spontaneous mutation rates,
476 excluding lethal mutations. Coupled with whole genome sequencing of ancestral and
477 MA lines, this experiment enables the direct estimation of the spontaneous mutation
478 rate of a species. Here, MA lines came from a single cell obtained by dilution serving
479 as T_0 culture (named the mother line, ML) and were maintained in 24-wells plates in
480 L1 medium at 20 °C with a 16h-dark 8h-light life cycle. Single cell bottleneck by dilution
481 were done each 14 days to have a low effective population size and limit selection.
482 DNA of initial line (ML line) and final time of MA lines were extracted with chloroform
483 protocol and sequencing done with Illumina HiSeq or MiSeq by GATC biotech
484 (Germany). To detect duplications, raw reads were mapped against the reference
485 genomes for each strain with bwa mem v0.7.12 (Li and Durbin 2010). Then, bam files
486 were treated with samtools v1.3.1 (Li et al. 2009) and bedtools v 2-2.18.0 (Quinlan
487 and Hall 2010) to extract the coverage.

488

489 ***Statistical analyses***

490 The probability of drawing one chromosome twice out of k WCD events chosen from
491 n chromosomes, $P(j,n)$, is equal to the ratio of the number of combinations of (k-1) out
492 of n without replacement multiplied by k-1, which corresponds to the number of
493 possible k-1 different chromosomes with one in two copies, out of n, divided by the

494 number of combinations of *k* out of *n* with replacement (*k*-1)*C*(*k*-1,*n*)/(*n*+*k*-1), which
495 corresponds to the total number of possible subsamples of k chromosomes out of *n*.

496 $$P(k,n) = \frac{C(k-1,n) * (k-1)}{(n+k-1)!/k!\,(n-1)!}$$

497 All statistical analyses were performed with R (R Core Team 2022).
498

499 ***Dosage compensation analysis***

500 This study was started three years after the end of the MA experiments, for which
501 some MA lines had been cryopreserved which allowed us to re-start a culture of one
502 *B. prasinos* MA line with a duplicated chromosome C04 (Bp37) to investigate possible
503 dosage compensation. The restarted culture from the cryopreserved Bp37 was
504 renamed Bp37B. As a control, we used the reference culture of the strain *B. prasinos*
505 RCC4222, that is the derived from the RCC1105 used for the MA experiment. All
506 cultures (Bp37B and RCC4222) were maintained under a 12:12 h light:dark regime
507 under 50 µmol photon m^{-2} s^{-1} white light at 20 °C. The karyotype of the defrozen
508 cultures Bp37B and the control RCC4222 were checked by DNAseq resequencing.
509 The karyotype of Bp37B contained on additional chromosome copy number as
510 compared to Bp37, and the karyotype of RCC4222 contained one additional copy
511 numbers and one less chromosome copy number as compared to the mother line (ML)
512 (Figure S17). The relative transcription rate was corrected by the number of DNA copy
513 in each line for further analysis below.
514 For expression analysis, total RNA was extracted using the Direct-zol RNA MiniPrep
515 Kit (Zymo Research, Californie, USA) from pooling flasks of cultures (100 ml cultures
516 with 200 million cells per ml) taken 6h before and 1h before the light on, in triplicates
517 for the control RCC4222 and duplicates for Bp37B. Then for translation efficiency
518 analysis, polysome extraction was performed for Ribo-seq as described previously
519 (Carpentier et al. 2020) with few modifications. Briefly, 600 mL of *B. prasinos* culture
520 were centrifuged at 8,000 g for 20 minutes. After centrifugation, pellets were
521 resuspended in 2.4 mL of polysome extraction buffer. After 10 minutes of incubation
522 on ice and centrifugation, 2 mL of supernatant was loaded on a 9 mL 15–60% sucrose
523 gradient and centrifuged for 3h at 38 000 rpm with rotor SW41 Ti. Fractions
524 corresponding to polysomes were pooled and polysomal RNA was extracted as
525 previously described (Carpentier et al. 2020). RNA library preparation was performed
526 on total or polysomal RNA using a NEBNext Poly(A) mRNA Magnetic Isolation Module

527 and a NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs)
528 according to the manufacturer's instructions with 1 µg of RNA as a starting point.
529 Libraries were multiplexed and sequenced on a NextSeq 550. Raw reads were
530 mapped against the reference transcriptome of *B. prasinos* with RSEM with standard
531 parameters (Li and Dewey 2011). We obtained the TPM average of total RNA and
532 polysomes linked RNA that we compared between *B. prasinos* RCC4222 and the MA
533 line Bp37B. At the gene scale, we first used the total RNA TPM values to estimate the
534 expression difference ratio, r_{RNA} for each gene *i*, between two copies and single copy
535 genes as:

536 $$r_{RNA}(i) = \frac{TPM_{TWOCOPIES}\ (i)}{TPM_{SINGLECOPY}\ (i)}$$

537 These relative expression rates $r_{RNA}(i)$ were normalized by the expression rate median
538 of genes located on non-duplicated chromosomes in Bp37B (C03, C05, C06, C07,
539 C08, C09, C011, C012, C013, C014, C015, C016, C017 and C018), in order to
540 estimate the transcription rate *tr(i)* of duplicated chromosomes related to non-
541 duplicated chromosomes. Chromosome C19 was not considered because it is a
542 particularly outlier chromosome (see Discussion).
543 Second, we used the total RNA and polysomes linked RNA TPM values for each gene
544 *i* to calculate the translation rates of each gene, *r(i)*, as:

545 $$r\ (i) = \frac{PolysomeTPM\ (i)}{TotalTPM(i)}$$

546 The translation efficiency between a gene *i* in the two lines in single versus two copies,
547 *te(i)* was estimated as:

548 $$te(i) = \frac{r_{TWOCOPIES}\ (i)}{r_{SINGLECOPY}\ (i)}$$

549 The *te*(*i*) ratio was then normalized by the median translation efficiency of all genes
550 on non-duplicated chromosomes in strain BP37B (5,267 genes).
551 Then, to estimate the dosage compensation at the scale of the chromosomes, the
552 values were normalized by the average TPM of all single copy chromosomes prior to
553 the calculation of the ratios (Table S7).
554 Poly(A) tail analysis was performed as previously described with slight modifications
555 (Sement and Gagliardi 2014). PCR products were resolved on a 2.5% agarose gel.
556 Primers used in this study are available in Table S10.
557

558 ***Differential gene expression analyses***

559 The statistical significance of the genes differential expression levels was further
560 estimated with Deseq2 (Love et al. 2014). Transcriptional invariant genes were
561 defined as genes present on duplicated regions and for which the relative transcription
562 rate $r_{RNA}(i)$ was comprised between 0.9 and 1.1. The over-representation of a certain
563 GO term in the transcriptional invariant gene set was compared to the genome-wide
564 GO term background frequency using the GO enrichment analysis with default values
565 implemented in pico-PLAZA workbench (Vandepoele et al. 2013).

566

567 **AUTHOR CONTRIBUTIONS**

568 MK, SSB and GP performed the mutation accumulation experiments. MK performed
569 the bioinformatics analysis. FS performed the cell cultures, RNA extractions, PCR and
570 gel migration experiments. RM designed and performed the polysome analyses,
571 poly(A) tail tests and RNAseq preparation. GP performed statistical analyses and
572 coordinated the project. MK drafted the first version of the manuscript and all authors
573 participated to writing the final version.

574

575 **ACKNOWLEDGEMENTS**

576 We are grateful to Claire Hemon and Elodie Desgranges for technical assistance with
577 the MA experiments. We acknowledge the GenoToul Bioinformatics platform
578 (Toulouse, France) for bioinformatics analysis support and cluster availability, the
579 BIOPIC platform for support with the cytometry, and the sequencing facility of the
580 Université de Perpignan Via Domitia BioEnvironnement platform. This work was
581 funded by ANRJCJC-SVSE6-2013-0005 to GP and SSB and ANR PHYTOMICS
582 (ANR-21-CE02-0026). This study is set within the framework of the "Laboratoires
583 d'Excellences (LABEX)" TULIP (ANR-10-LABX-41) and of the "École Universitaire de
584 Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019).

585

586 **COMPETING INTERESTS**

587 The authors declare no conflicts of interest.

588

589 **DATA AVAILABILITY STATEMENT**

590 All genomic raw reads of ML and mutation accumulation lines are available under the
591 bioprojects PRJNA531882 (*Ostreoccocus tauri*, *O. mediterraneus*, *Micromonas*

592 *commoda*, *Bathycoccus prasinos*), PRJNA453760 and PRJNA389600 (*Picochlorum*
593 *costavermella*), and PRJNA478011 (*Phaeodactylum tricornutum*). Transcriptomic raw
594 reads of the *Bathycoccus prasinos* MA lines 37 (Bp37B) and control line RCC4222 are
595 available under the bioproject PRJNA715163. A summary is provided in Table S11.
596 **SUPPLEMENTARY MATERIAL**
597
598 **Figure S1 to S16.** Raw coverage by 1 kb windows of all mutation accumulation lines
599 from *Ostreoccocus tauri*, *O. mediterraneus*, *Bathycoccus prasinos*, *Micromonas*
600 *commoda*, *Picochlorum costavermella* and *Phaeodactylum tricornutum*. The
601 horizontal grey lines indicate the chromosome separation, and the duplicated
602 chromosome are in blue. ML is the T0 genome of mutation accumulation experiments.
603 In *M. commoda*, a duplication of a fraction of chromosome 2 in the mother line (ML) is
604 maintained in all MA lines, suggesting there might be a miss-assembly due to a
605 duplicated region at this location.
606 **Figure S17**. Coverage analysis of the Bp37B and control RCC4222.
607 **Figure S18**. Transcripts from duplicated chromosome C01 and non-duplicated
608 chromosome C08 present similar poly(A) tail.
609
610 **Table S1.** List of duplicated chromosomes in the MA lines.
611 **Table S2.** Limits of sub-regions of chromosomes C01 and C02.
612 **Table S3.** Raw transcript per million (TPM) table for all 10 samples.
613 **Table S4.** Estimation of the number of dosage insensitive genes per chromosome.
614 **Table S5.** List of the dosage insensitive genes with annotations.
615 **Table S6.** Relative copy numbers of chromosomes in the two lines of *Bathycoccus*
616 *prasinos*.
617 **Table S7.** Relative TPM average per chromosome used for Figure 2.
618 **Table S8.** Fitness of MA lines with duplicated chromosomes during MA experiments.
619 **Table S9.** Point mutations previously identified in the MA lines with whole
620 chromosome duplications.
621 **Table S10.** Sequences of primers used for poly(A) tail length analysis.
622 **Table S11.** Summary of all accessions of the data used in this study with Bioprojects
623 and biosamples.
624
625

626
627
628
629
630
631

632 **BIBLIOGRAPHY**

633 Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004.
634 Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat. Rev.*
635 *Genet.* 5:725–738.

636 Baker BS, Gorman M, Marín I. 1994. Dosage compensation in Drosophila. *Annu.*
637 *Rev. Genet.* 28:491–521.

638 Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, Martin J,
639 Schackwitz W, Kuo A, Salin G, Donnadieu C, et al. 2017. Population genomics of
640 picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* 3:e1700239.

641 Blanc-Mathieu R, Verhelst B, Derelle E, Rombauts S, Bouget F-Y, Carré I, Château
642 A, Eyre-Walker A, Grimsley N, Moreau H, et al. 2014. An improved genome of the
643 model marine alga Ostreococcus tauri unfolds by assessing Illumina de novo
644 assemblies. *BMC Genomics* 15:1103.

645 Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U,
646 Martens C, Maumus F, Otillar RP, et al. 2008. The Phaeodactylum genome reveals
647 the evolutionary history of diatom genomes. *Nature* 456:239–244.

648 Carpentier M-C, Deragon J-M, Jean V, Be SHV, Bousquet-Antonelli C, Merret R.
649 2020. Monitoring of XRN4 Targets Reveals the Importance of Cotranslational Decay
650 during Arabidopsis Development. *Plant Physiol.* 184:1251–1262.

651 Chang AY-F, Liao B-Y. 2020. Reduced Translational Efficiency of Eukaryotic Genes
652 after Duplication Events. *Mol. Biol. Evol.* 37:1452–1461.

653 Charlesworth D. 2019. Young sex chromosomes in plants and animals. *New Phytol.*
654 224:1095–1107.

655 Chen G, Bradford WD, Seidel CW, Li R. 2012. Hsp90 stress potentiates rapid
656 cellular adaptation through induction of aneuploidy. *Nature* 482:246–250.

657 Cimini D, Howell B, Maddox P, Khodjakov A, Degrassi F, Salmon ED. 2001.
658 Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic
659 Mammalian Tissue Cells. *J. Cell Biol.* 153:517–528.

660 Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, Torres EM.
661 2014. Quantitative proteomic analysis reveals posttranslational responses to
662 aneuploidy in yeast.Dikic I, editor. *eLife* 3:e03023.

663 Devlin RH, Holm DG, Grigliatti TA. 1988. The influence of whole-arm trisomy on
664 gene expression in Drosophila. *Genetics* 118:87–101.

665 Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of
666 alternative polyadenylation. *Mol. Cell* 43:853–866.

667 Disteche CM. 2012. Dosage Compensation of the Sex Chromosomes. *Annu. Rev.*
668 *Genet.* 46:537–560.

669 Eichhorn SW, Subtelny AO, Kronja I, Kwasnieski JC, Orr-Weaver TL, Bartel DP.
670 2016. mRNA poly(A)-tail changes specified by deadenylation broadly reshape
671 translation in Drosophila oocytes and early embryos.Izaurralde E, editor. *eLife*
672 5:e16955.

673 FitzPatrick DR, Ramsay J, McGill NI, Shade M, Carothers AD, Hastie ND. 2002.
674 Transcriptome analysis of human autosomal trisomy. *Hum. Mol. Genet.* 11:3249–
675 3256.

676 Ford JH, Correll AT. 1992. Chromosome errors at mitotic anaphase. *Genome*
677 35:702–705.

678 Gan L, Ladinsky MS, Jensen GJ. 2011. Organization of the Smallest Eukaryotic
679 Spindle. *Curr. Biol.* 21:1578–1583.

680 Gearhart JD, Oster-Granite ML, Reeves RH, Coyle JT. 1987. Developmental
681 consequences of autosomal aneuploidy in mammals. *Dev. Genet.* 8:249–265.

682 Giguere DJ, Bahcheli AT, Slattery SS, Patel RR, Browne TS, Flatley M, Karas BJ,
683 Edgell DR, Gloor GB. 2022. Telomere-to-telomere genome assembly of
684 Phaeodactylum tricornutum. *PeerJ* 10:e13607.

685 Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage
686 compensation. *Nat. Rev. Genet.* 17:33–46.

687 Halligan DL, Keightley PD. 2009. Spontaneous Mutation Accumulation Studies in
688 Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.* 40:151–172.

689 Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. *Annu.*
690 *Rev. Genet.* 31:571–610.

691 Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F,
692 Ruedi M, Kaessmann H, Reymond A. 2009. Segmental copy number variation
693 shapes tissue transcriptomes. *Nat. Genet.* 41:424–429.

694 Hose J, Escalante LE, Clowers KJ, Dutcher HA, Robinson D, Bouriakov V, Coon JJ,
695 Shishkova E, Gasch AP. 2020. The genetic basis of aneuploidy tolerance in wild
696 yeast.Klein H, Tyler JK, Gresham D, Argueso JL, Bloom K, editors. *eLife* 9:e52063.

697 Hou J, Shi X, Chen C, Islam MS, Johnson AF, Kanno T, Huettel B, Yen M-R, Hsu F-
698 M, Ji T, et al. 2018. Global impacts of chromosomal imbalance on gene expression
699 in Arabidopsis and other taxa. *Proc. Natl. Acad. Sci.* 115:E11321–E11330.

700 Kaya A, Mariotti M, Tyshkovskiy A, Zhou X, Hulke ML, Ma S, Gerashchenko MV,
701 Koren A, Gladyshev VN. 2020. Molecular signatures of aneuploidy-driven adaptive
702 evolution. *Nat. Commun.* 11:588.

703 Kojima S, Cimini D. 2019. Aneuploidy and gene expression: is there dosage
704 compensation? *Epigenomics* 11:1827–1837.

705 Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju
706 V. 2018. Mutational and transcriptional landscape of spontaneous gene duplications
707 and deletions in Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* 115:7386–7391.

708 Krasovec M, Eyre-Walker A, Grimsley N, Salmeron C, Pecqueur D, Piganeau G,
709 Sanchez-Ferandin S. 2016. Fitness Effects of Spontaneous Mutations in
710 Picoeukaryotic Marine Green Algae. *G3 Genes Genomes Genet.* 6:2063–2071.

711 Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G. 2017. Spontaneous

712 Mutation Rate in the Smallest Photosynthetic Eukaryotes. *Mol. Biol. Evol.* 34:1770–
713 1779.

714 Krasovec M, Lipinska AP, Coelho SM. 2022. Low spontaneous mutation rate in a
715 complex multicellular eukaryote with a haploid-diploid life cycle. :2022.05.13.491831.
716 Available from: https://www.biorxiv.org/content/10.1101/2022.05.13.491831v1

717 Krasovec M, Sanchez-Brosseau S, Grimsley N, Piganeau G. 2018. Spontaneous
718 mutation rate as a source of diversity for improving desirable traits in cultured
719 microalgae. *Algal Res.* 35:85–90.

720 Krasovec M, Sanchez-Brosseau S, Piganeau G. 2019. First Estimation of the
721 Spontaneous Mutation Rate in Diatoms. *Genome Biol. Evol.* 11:1829–1837.

722 Krasovec M, Vancaester E, Rombauts S, Bucchini F, Yau S, Hemon C,
723 Lebredonchel H, Grimsley N, Moreau H, Sanchez-Brosseau S, et al. 2018. Genome
724 Analyses of the Microalga Picochlorum Provide Insights into the Evolution of
725 Thermotolerance in the Green Lineage. *Genome Biol. Evol.* 10:2347–2365.

726 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data
727 with or without a reference genome. *BMC Bioinformatics* 12:323.

728 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler
729 transform. *Bioinformatics* 26:589–595.

730 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
731 Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence
732 Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25:2078–2079.

733 Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A)
734 tail regulation in oocyte-to-embryo development. *Genes Dev.* [Internet]. Available
735 from: http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116

736 Liu H, Zhang J. 2019. Yeast Spontaneous Mutation Rate and Spectrum Vary with
737 Environment. *Curr. Biol. CB* 29:1584-1591.e3.

738 Liu Y, Borel C, Li L, Müller T, Williams EG, Germain P-L, Buljan M, Sajic T,
739 Boersema PJ, Shao W, et al. 2017. Systematic proteome and proteostasis profiling
740 in human Trisomy 21 fibroblast cells. *Nat. Commun.* 8:1212.

741 Loane M, Morris JK, Addor M-C, Arriola L, Budd J, Doray B, Garne E, Gatt M,
742 Haeusler M, Khoshnood B, et al. 2013. Twenty-year trends in the prevalence of
743 Down syndrome and other trisomies in Europe: impact of maternal age and prenatal
744 screening. *Eur. J. Hum. Genet. EJHG* 21:27–33.

745 Loehlin DW, Carroll SB. 2016. Expression of tandem gene duplicates is often greater
746 than twofold. *Proc. Natl. Acad. Sci.* 113:5988–5992.

747 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
748 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

749 Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE. 2004. Gene
750 expression from the aneuploid chromosome in a trisomy mouse model of down
751 syndrome. *Genome Res.* 14:1268–1274.

752 Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016.
753 Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.*
754 17:704–714.

755 Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ,
756 Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum
757 of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci.* 105:9272–9277.

758 Miao Z, Zhang T, Xie B, Qi Y, Ma C. 2022. Evolutionary Implications of the RNA N6-
759 Methyladenosine Methylome in Plants. *Mol. Biol. Evol.* 39:msab299.

760 Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, Van Bel M,
761 Poulain J, Katinka M, Hohmann-Marriott MF, et al. 2012. Gene functionalities and
762 genome structure in Bathycoccus prasinos reflect cellular specializations at the base
763 of the green lineage. *Genome Biol.* 13:R74.

764 Muyle A, Shearn R, Marais GA. 2017. The Evolution of Sex Chromosomes and
765 Dosage Compensation in Plants. *Genome Biol. Evol.* 9:627–645.

766 Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid
767 De Novo Evolution of X Chromosome Dosage Compensation in Silene latifolia, a
768 Plant with Young Sex Chromosomes. *PLOS Biol.* 10:e1001308.

769 Nagaoka SI, Hassold TJ, Hunt PA. 2012. Human aneuploidy: mechanisms and new
770 insights into an age-old problem. *Nat. Rev. Genet.* 13:493–504.

771 Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. 2015.
772 Extensive de novo mutation rate variation between individuals and across the
773 genome of Chlamydomonas reinhardtii. *Genome Res.* 25:1739–1749.

774 Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A,
775 Barre B, Freel K, Llored A, et al. 2018. Genome evolution across 1,011
776 Saccharomyces cerevisiae isolates. *Nature* 556:339–344.

777 Peterson JB, Ris H. 1976. Electron-microscopic study of the spindle and
778 chromosome movement in the yeast Saccharomyces cerevisiae. *J. Cell Sci.* 22:219–
779 242.

780 Qian W, Liao B-Y, Chang AY-F, Zhang J. 2010. Maintenance of duplicate genes and
781 their functional redundancy by reduced expression. *Trends Genet. TIG* 26:425–430.

782 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing
783 genomic features. *Bioinformatics* 26:841–842.

784 R Core Team (2022). R: A language and environment for statistical computing. R
785 Foundation for Statistical Computing, Vienna, Austria. URL https://www.R project.org/.

786 Rajagopalan H, Lengauer C. 2004. Aneuploidy and cancer. *Nature* 432:338–341.

787 Scopel EFC, Hose J, Bensasson D, Gasch AP. 2021. Genetic variation in aneuploidy
788 prevalence and tolerance across Saccharomyces cerevisiae lineages. *Genetics*
789 217:iyab015.

790 Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in
791 drug-resistant Candida albicans. *Science* 313:367–370.

792 Sement FM, Gagliardi D. 2014. Detection of uridylated mRNAs. *Methods Mol. Biol.*
793 *Clifton NJ* 1125:43–51.

794 Slobodin B, Bahat A, Sehrawat U, Becker-Herman S, Zuckerman B, Weiss AN, Han
795 R, Elkon R, Agami R, Ulitsky I, et al. 2020. Transcription Dynamics Regulate Poly(A)
796 Tails and Expression of the RNA Degradation Machinery to Balance mRNA Levels.
797 *Mol. Cell* 78:434-444.e5.

798 Song MJ, Potter BI, Doyle JJ, Coate JE. 2020. Gene Balance Predicts
799 Transcriptional Responses Immediately Following Ploidy Change in Arabidopsis
800 thaliana. *Plant Cell* 32:1434–1448.

801 Stenberg P, Lundberg LE, Johansson A-M, Rydén P, Svensson MJ, Larsson J.
802 2009. Buffering of Segmental and Chromosomal Aneuploidies in Drosophila
803 melanogaster. *PLOS Genet.* 5:e1000465.

804 Subirana L, Péquin B, Michely S, Escande M-L, Meilland J, Derelle E, Marin B,
805 Piganeau G, Desdevises Y, Moreau H, et al. 2013. Morphology, Genome Plasticity,
806 and Phylogeny in the Genus Ostreococcus Reveal a Cryptic Species, O.
807 mediterraneus sp. nov. (Mamiellales, Mamiellophyceae). *Protist* 164:643–659.

808 Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling
809 reveals an embryonic switch in translational control. *Nature* 508:66–71.

810 Thompson SL, Compton DA. 2008. Examining the link between chromosomal
811 instability and aneuploidy in human cells. *J. Cell Biol.* 180:665–672.

812 Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP,
813 Dunham MJ, Amon A. 2010. Identification of aneuploidy-tolerating mutations. *Cell*
814 143:71–83.

815 Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A.
816 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast.
817 *Science* 317:916–924.

818 Udagawa T, Swanger SA, Takeuchi K, Kim JH, Nalavadi V, Shin J, Lorenz LJ, Zukin
819 RS, Bassell GJ, Richter JD. 2012. Bidirectional control of mRNA translation and
820 synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol. Cell* 47:253–
821 266.

822 Vandepoele K, Van Bel M, Richard G, Van Landeghem S, Verhelst B, Moreau H,
823 Van de Peer Y, Grimsley N, Piganeau G. 2013. pico-PLAZA, a genome database of
824 microbial photosynthetic eukaryotes. *Environ. Microbiol.* 15:2147–2153.

825 de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le
826 Bescot N, Probert I, et al. 2015. Ocean plankton. Eukaryotic plankton diversity in the
827 sunlit ocean. *Science* 348:1261605.

828 Veitia RA, Potier MC. 2015. Gene dosage imbalances: action, reaction, and models.
829 *Trends Biochem. Sci.* 40:309–317.

830 Weill L, Belloc E, Bava F-A, Méndez R. 2012. Translational control by changes in
831 poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.* 19:577–585.

832 Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE,
833 Amon A. 2008. Aneuploidy affects proliferation and spontaneous immortalization in
834 mammalian cells. *Science* 322:703–709.

835 Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier
836 ML, Derelle E, Everett MV, et al. 2009. Green evolution and dynamic adaptations
837 revealed by genomes of the marine picoeukaryotes Micromonas. *Science* 324:268–
838 272.

839 Yau S, Hemon C, Derelle E, Moreau H, Piganeau G, Grimsley N. 2016. A Viral
840 Immunity Chromosome in the Marine Picoeukaryote, Ostreococcus tauri. *PLoS*
841 *Pathog.* 12:e1005965.

Yau S, Krasovec M, Benites LF, Rombauts S, Groussin M, Vancaester E, Aury J-M, Derelle E, Desdevises Y, Escande M-L, et al. 2020. Virus-host coexistence in phytoplankton through the genomic lens. *Sci. Adv.* 6:eaay2587.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21:809–818.

Zhang X, Virtanen A, Kleiman FE. 2010. To polyadenylate or to deadenylate: that is the question. *Cell Cycle Georget. Tex* 9:4437–4449.

Zhang Z, Presgraves DC. 2016. Drosophila X-Linked Genes Have Lower Translation Rates than Autosomal Genes. *Mol. Biol. Evol.* 33:413–428.

Zhang Z, Presgraves DC. 2017. Translational compensation of gene copy number alterations by aneuploidy in Drosophila melanogaster. *Nucleic Acids Res.* 45:2986–2993.

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 111:E2310-2318.

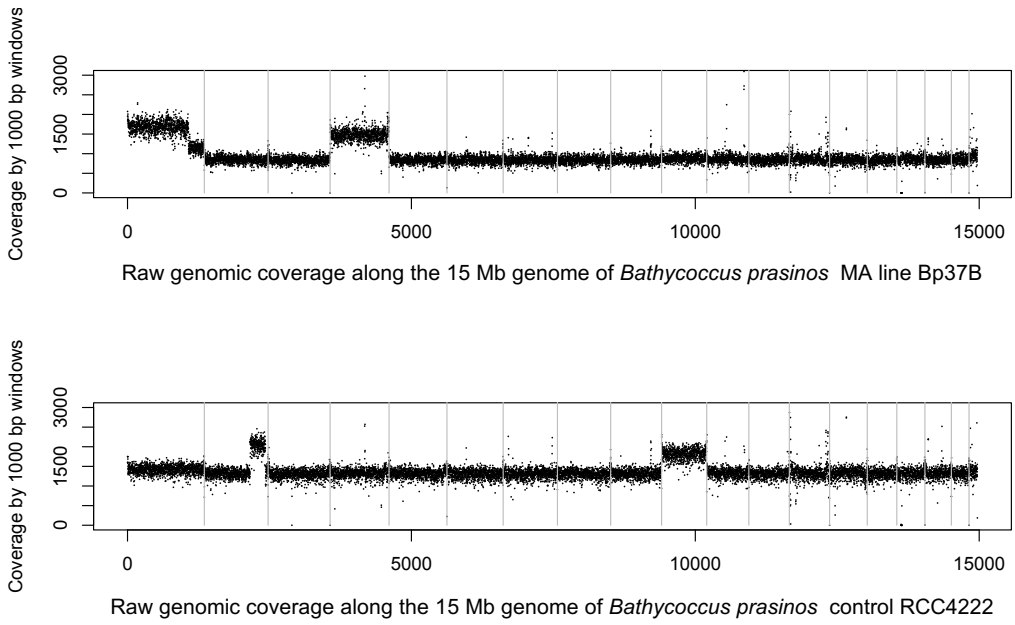


Figure S17. Coverage by 1 kb windows of the *Bathycoccus prasinos* MA line Bp37 and control RCC4222.

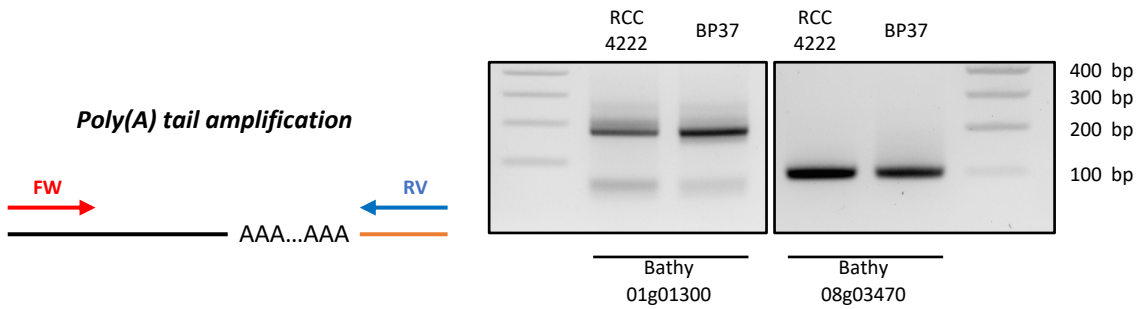


Figure S18. Transcripts from duplicated chromosome C01 and non-duplicated chromosome C08 present similar poly(A) tail. Poly(A) tail measurement was performed using modified 3'RACE. PCR amplification was performed using primers flanking poly(A) tail. Experiments was performed using total RNA from RCC4222 line (control line) and Bp37B line (chromosome C04 duplicated). Illustrations representing PCR amplification are present on the left panel. Black line represents the mRNA and the orange line represent the ligated adapter used for reverse transcription and PCR amplification.