



A high frequency of chromosomal duplications in unicellular algae is compensated by translational regulation

Marc Krasovec, Rémy Merret, Frédéric Sanchez, Sophie Sanchez-Brosseau,
Gwenaël Piganeau

► To cite this version:

Marc Krasovec, Rémy Merret, Frédéric Sanchez, Sophie Sanchez-Brosseau, Gwenaël Piganeau. A high frequency of chromosomal duplications in unicellular algae is compensated by translational regulation. *Genome Biology and Evolution*, 2023, 15 (6), 10.1093/gbe/evad086 . hal-04104946

HAL Id: hal-04104946

<https://univ-perp.hal.science/hal-04104946>

Submitted on 5 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A high frequency of chromosomal duplications in unicellular algae is compensated by translational regulation

Marc Krasovec^{1*}, Rémy Merret^{2,3,*}, Frédéric Sanchez¹, Sophie Sanchez-Brosseau¹, Gwenaél 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

While duplications have long been recognized as a fundamental process driving major evolutionary innovations, direct estimates of spontaneous chromosome duplication rates, leading to aneuploid karyotypes, are scarce. Here, from mutation accumulation (MA) experiments, we provide the first estimates of spontaneous chromosome duplication rates in six unicellular eukaryotic species, which range from 1×10^{-4} to 1×10^{-3} per genome per generation. Although this is ~ 4 to ~ 50 times less frequent than spontaneous point mutations per genome, chromosome duplication events can affect 1 to 7% of the total genome size. In duplicated chromosomes, mRNA levels reflected gene copy numbers, but the level of translation estimated by polysome profiling revealed that dosage compensation must be occurring. In particular, one duplicated chromosome showed a 2.1-fold increase of mRNA but translation rates were decreased to 0.7-fold. Altogether, our results support previous observations of chromosome-dependent dosage compensation effects, providing evidence that compensation occurs during translation. We hypothesize that an unknown post-transcriptional mechanism modulates the translation of hundreds of transcripts from genes located on duplicated regions in eukaryotes.

SIGNIFICANCE STATEMENT

Aneuploid karyotypes cause genetic disease and decrease fitness due to gene dosage imbalance. The deleterious effects of aneuploidy can lead to the evolution of dosage compensation, a mechanism that restores ancestral ploidy and usually affects the transcription rate. Here, we estimated the rate of spontaneous aneuploidy in six eukaryotic phytoplankton species from mutation accumulation experiments and explored transcriptional and translational responses following whole chromosome duplications. We provide evidence for chromosome-specific post-transcriptional dosage compensation.

INTRODUCTION

Complete or partial chromosome duplications leading to aneuploid karyotypes are known to contribute to genetic diseases, such as Trisomy 21 in humans or cancer (Rajagopalan and Lengauer 2004). Duplication of a single chromosome leads to an immediate imbalance in cellular metabolism because of the increased levels of transcripts and proteins encoded by the doubled chromosome. This is expected to have deleterious effects because it is costly and may disrupt the function of biochemical pathways 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 an excess of transcripts, as compared to gene duplications associated with invariant transcript levels (Konrad et al. 2018). Moreover, the consequences of aneuploidy on gene transcription are complex and while gene transcription may increase with chromosome copy number, gene transcription 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).

In contrast, in some circumstances, aneuploidy may confer a selective advantage such as in resistance to drugs in *Candida albicans* (Selmecki et al. 2006) or *Saccharomyces cerevisiae* (Chen et al. 2012). Aneuploidy was also prevalent and tolerated across

1 *Saccharomyces cerevisiae* lineages (Scopel et al. 2021), where a fifth of the sequenced
 2 strains harboured atypical aneuploidy karyotypes (Peter et al. 2018), witnessing a lack
 3 of strong deleterious effects. Indeed, different dosage compensation mechanisms have
 4 evolved and may restore the ancestral gene dose leading to aneuploidy tolerance.
 5 Three possible targets of dosage compensation mechanisms for duplicated genes
 6 include the modification of the transcription rate (e.g the amount of transcript products),
 7 the translation rate (e.g the number of translated transcripts compared to the number of
 8 total produced transcripts), or the protein degradation rate (the number of degraded
 9 proteins compared to the total number of produced proteins). The most studied
 10 mechanisms are those involved in modifying the transcription rate during the evolution
 11 of heterogametic sex chromosomes, known both in plants (Muyle et al. 2012; Muyle et
 12 al. 2017; Charlesworth 2019) and animals (Disteche 2012; Graves 2016). Different
 13 mechanisms evolved either by simulating the ancestral ploidy by doubling the
 14 transcription of the genes of the single copy male X (Baker et al. 1994); or by equalizing
 15 the ploidy between the two sexes by halving the transcription – silencing of one X in
 16 female (Heard et al. 1997). Although dosage compensation is well studied in sex
 17 chromosome evolution, whether and how it evolves after a chromosome ploidy variation
 18 is unclear, particularly for autosomes (Kojima and Cimini 2019). Previous studies have
 19 reported inconsistent results; on the one hand, a significant increase of transcription for
 20 the duplicated genes has been observed in *C. albicans* (Selmecki et al. 2006),
 21 *Drosophila* (Loehlin and Carroll 2016), *S. cerevisiae* (Torres et al. 2007) and
 22 mammalian cells (Williams et al. 2008); on the other hand, decreased transcription of
 23 the duplicated genes has been reported in yeast and mammals, suggesting a
 24 compensation at the transcriptional level (Henrichsen et al. 2009; Qian et al. 2010). In
 25 the case of chromosome duplication, evidence for dosage compensation at the
 26 transcriptional level is scarce (Stenberg et al. 2009; Hose et al. 2020), and scaling of
 27 gene transcription with gene copy number seems more prevalent and has been
 28 reported in disomic yeasts (Kaya et al. 2020), *Candida albicans* and human cells
 29 (Kojima and Cimini 2019). Moreover, dosage compensation has been reported to occur
 30 at the post-transcriptional level, via the modification of translation efficiency, in
 31 *Drosophila* (Zhang and Presgraves 2017) or at the post-translational level, via an

1 increased degradation rate of proteins, for 20% of the proteome in yeast (Dephoure et
2 al. 2014). Increased protein degradation is involved in the compensation mechanism in
3 human Down's syndrome for proteins encoded on triplicated chromosome 21 (Liu et al.
4 2017).

5 Although the consequences of chromosome duplication on transcription rates - and to a
6 lesser extent on translation rates and protein abundance - have been studied in many
7 model organisms, our knowledge of chromosome duplication rates in eukaryotes is
8 currently limited to *Saccharomyces cerevisiae* (Lynch et al. 2008; Zhu et al. 2014; Liu
9 and Zhang 2019) and humans (Nagaoka et al. 2012; Loane et al. 2013). Here, we
10 investigate the spontaneous chromosome duplication rate in six unicellular species by
11 analysing sequence data from a mutation accumulation experiment. The principle of
12 mutation accumulation experiments is to follow the descendants originated from a single
13 cell under minimal selection, ensured by serial bottlenecks during dozens to thousands
14 cell divisions (Halligan and Keightley 2009). The six species include five green algae
15 and one diatom, all ecological relevant primary producers in the sunlit ocean (de Vargas
16 et al. 2015), with a large phylogenetic spread encompassing 1.5 billion years of
17 divergence (Yoon et al. 2004). We also recovered cryopreserved mutation accumulation
18 lines to investigate the consequence of chromosome duplication on the transcription
19 and translation rates.

20

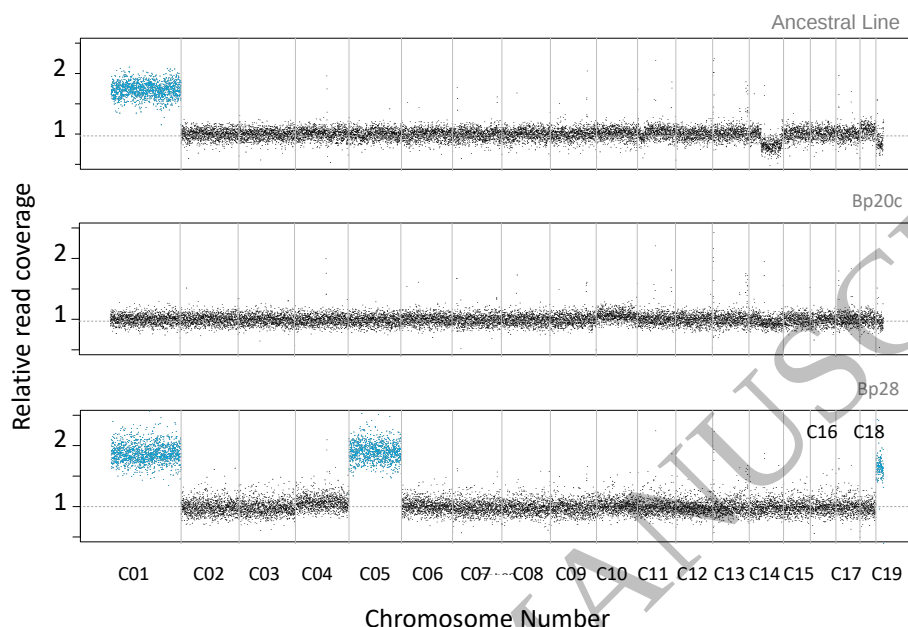
ACCEPTED MANUSCRIPT

RESULTS

Whole chromosome duplication rate

In this study, we analysed whole genome resequencing data from previous mutation accumulation experiments (Krasovec et al. 2016; Krasovec et al. 2017; Krasovec, Sanchez-Brosseau, et al. 2018) in five haploid green algae species : *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 one diploid diatom, *Phaeodactylum tricornutum* RCC2967 (Bowler et al. 2008; Giguere et al. 2022). We maintained 12 to 40 mutation accumulation lines for a total of 1,595 to 17,250 generations depending of species (Table S1). 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 S2). 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 twice over the course of the experiment: C19 in lines Bp25c and Bp28b; C17 in Mc08 and Mc09; C23 in Pt11 and Pt10c. The probability of observing two independent whole chromosome duplications of the same chromosome in *B. prasinos*, *M. commoda* and *P. tricornutum* are 0.46, 0.47 and 0.34 respectively (see methods), and these probabilities are thus consistent with the null hypothesis of an equal probability of duplication across chromosomes. All independent duplication events inferred from coverage analysis and genealogies are summarized in Table 1 and S2. Unexpectedly, the analyses revealed that the ancestral line of the MA experiment in *B. prasinos* carried two copies of chromosome C01. One copy was subsequently lost 11 times independently over 4,145 generations, corresponding to a spontaneous duplicated chromosome loss 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

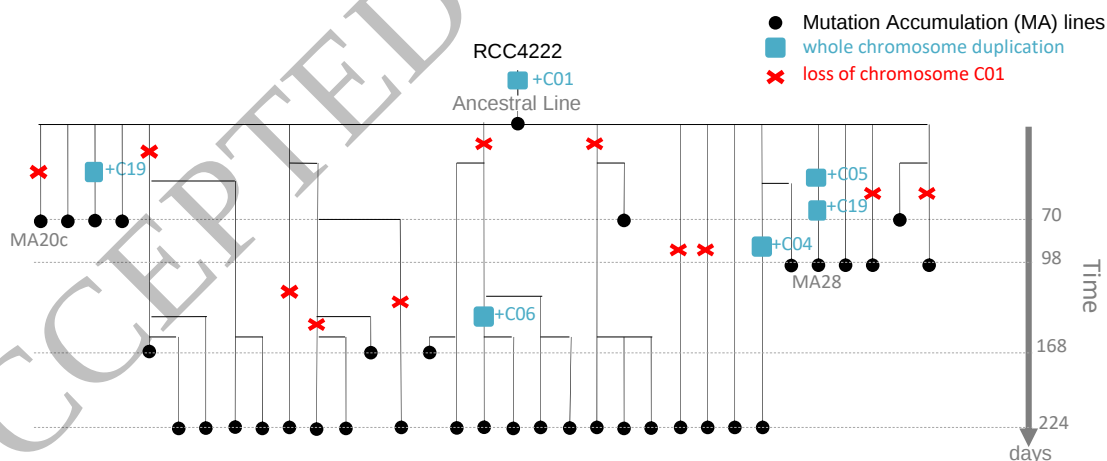


Figure 1. A. Normalized raw genomic coverage of the ancestral line (T0 of mutation accumulation experiment), Bp20c and Bp28 mutation accumulation lines of *Bathycoccus prasinos*. Vertical grey lines are chromosome separators. Read coverage highlighted in blue shows chromosomes in double copies. Raw coverage 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 ancestral line) of the experiment. This duplication is then lost several times, and five other

1 duplications of chromosomes C04, C05, C06 and C19 occurred. In case of shared ancestry between
 2 lineages, the number of shared mutations and generations were counted only once, in order to include
 3 only independent generations and mutations.

4
 5 **Table 1.** Spontaneous whole chromosome duplication rate in six species. N_{lines} : number of mutation
 6 accumulation lines, Gen : average number of generations per MA line, Tot_{Gen} : total number of
 7 generations, N_{Chrom} : number of chromosomes in the ancestral karyotype, N_{WCD} is the number of
 8 independent whole chromosome duplications, U_{WCD} is the whole chromosome duplication rate per
 9 chromosome per cell division, and U_{Cell} is the whole chromosome duplication rate per cell division, U_{bs} is
 10 the base substitution mutation rate per cell division. The estimation of the upper limit of N_{WCD} and U_{Cell} in
 11 *O. tauri* and *P. costavermella* 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}
<i>Bathycoccus prasinos</i>	35	265	4,145	19	5	C04, C05, C06, C19	0.000063	0.00121	0.0046
<i>Micromonas commoda</i>	37	272	4,994	17	5	C05, C12, C16, C17	0.000059	0.00100	0.0171
<i>Ostreococcus tauri</i>	40	512	17,250	20	0	-	<0.000003	<0.00006	0.0054
<i>Ostreococcus mediterraneus</i>	33	272	8,380	19	1	C14	0.000006	0.00012	0.0064
<i>Picochlorum costavermella</i>	12	133	1,596	10	0	-	<0.00006	<0.00063	0.0119
<i>Phaeodactylum tricornutum</i>	36	181	6,516	25	4	C02, C14, C23	0.000025	0.00061	0.0132

12 13 14 **Consequences of chromosome duplication on transcription**

15 To explore the consequences of whole or partial chromosome duplication on
 16 transcription, we compared transcription rates in a control strain (*B. prasinos* RCC4222)
 17 and one mutation accumulation line of *B. prasinos*, which was revived after 4 years of
 18 cryopreservation. This cryopreserved culture originated from line Bp37, and the
 19 recovered culture is hereafter referred to as Bp37B. Whole genome coverage was used
 20 as a proxy for chromosome copy number in both lines. This confirmed the presence of
 21 two copies of chromosome C04 in Bp37B as in the original Bp37 (Figure 2A and S6),
 22 and also revealed an additional chromosome C01 in Bp37B (Figure 2A, Figure S17).
 23 The heterogeneous coverage of chromosome C01 led us to divide it into two regions for
 24 subsequent analyses; region C01a (1.98 fold coverage) and C01b (1.35 fold coverage).
 25 The control line also contained duplicated regions (Figure 2A, Figure S17),
 26 chromosome C10 and a region of chromosome C02 named C02b, while C02a was
 27 single copy. The boundaries of C01 and C02 subregions are provided in Table S3. We
 28 interpreted coverage values less than 2 (Figure S17), e.g. C01b in Bp37B, C10 and

C02b as duplications that were carried by a subpopulation of cells. Cultures were grown up to ~10 million cells per ml prior to extraction so that polymorphism is not unexpected. Comparative transcriptome analyses of Bp37B and the control line revealed that there were on average twice the number of transcripts for genes located on the duplicated chromosomes as compared to the control line - C04 (transcription rate $tr(i)$ average=2.13, median=1.95, estimated from TPM, Figure 2B, raw data available in Table S4) and C01a (transcription rate $tr(i)$ average=2.29, median=2.02). The transcription rate of genes on chromosomes C10 and C02b was also affected. Altogether, the transcript production was scaled up with the chromosome copy number in the Bp37B and control lines (Pearson correlation, $\rho=0.87$, $p\text{-value}<0.001$, Figure 3A).

Gene-by-gene variation of the consequences of DNA copy number

Dosage invariant genes (Antonarakis et al. 2004; Lyle et al. 2004) are genes for which the transcription is not affected by copy number, and they may be involved in aneuploidy tolerance. We investigated differential gene transcription at the individual gene scale with Deseq2 (Love et al. 2014) to identify candidate invariant genes. We found that the transcription rate of 158 out of the 1776 genes in two copies (9%) located on duplicated chromosomes or regions were not significantly higher than for genes in single copy (Table 2). Two Gene Ontology categories (RNA processing, GO:0006396 and DNA metabolic process GO:0006259) were over-represented in this transcriptomic invariant gene set. First, genes involved in RNA processing (3.8 times more frequent, $p\text{-value}<0.01$) including members of heteromeric protein complexes such as subunits 7 and 2 of the U6 snRNA-associated Sm-like protein (Table S5), invariant 158 annotations). Second, genes involved in DNA metabolic process (2.2 times more frequent in the invariant gene subset, $p\text{-value}<0.04$) including the DNA Polymerase A and the DNA primase large subunit. Notable protein complex members of the invariant data set are Histone 3 and the subunit E of the translation initiation factor 3. There are in total 12 genes annotated as subunits in the subset of invariant genes, this is significantly more than the frequency of protein coding genes annotated as subunits in the complete proteome (Fisher exact test, $p\text{-value}=0.0006$). This suggests that a higher

- 1 proportion of genes coding for protein forming complexes have a gene-specific
- 2 regulated transcription that is not affected by gene copy number.

3

ACCEPTED MANUSCRIPT

Table 2: Estimation of number of dosage insensitive genes (duplicated genes for which transcription is not changing). C14 and C19 are outliers in terms of gene transcription as they encode the mating type locus (C14) and a hypervariable chromosome involved in immunity against viruses (C19). DE₁ is the number of differentially expressed genes between the two lines (adjusted p-value <0.01), DE₂ is the subset of genes of DE₁ for which the transcription rate is higher in the duplicated gene.

	Genes		Chromosomes																				duplicated in strains 37B										duplicated in strains WT									
	Two copies	Single copy	01 A	01 b	02 a	02 b	02 c	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19																		
Total *genes	1767	5234	575	136	400	159	24	550	488	482	507	475	488	449	409	365	366	279	340	256	242	207	144	66																		
Transcription rates: number of Differentially Expressed (DE) genes and relative transcription rates																																										
DE ₁ : p _{adj} <0.01	407 (23%)	504 (10%)	188	10	40	24	0	70	114	40	49	42	60	39	71	37	26	20	18	29	20	25	7	33																		
r **	2.7	1.2	2.7	1.5	1.5	2.5	-	0.9	2.6	1.2	1.0	1.3	1.1	1.4	2.8	1.3	1.1	1.6	0.8	1.6	1.5	1.1	1.0	0.3																		
DE ₂ :	385 (95%)	250 (50%)	187	6	-	21	-	-	112	-	-	-	-	-	59	-	-	-	-	-	-	-	-	0																		
DI _{RNA} ***	158 (9%)	-	58	26	-	9	-	-	38	-	-	-	-	-	27	-	-	-	-	-	-	-	-	-																		

*total gene number is 7530 after exclusion of 280 genes with Transcript per Million TPM=0 in one replicate.

** relative transcription rate, r, is obtained by dividing the average TPM in the strain with two copies by the TPM in the strain with one copy, whereas for single copy genes, it is the relative TPM_{37B}/TPM₄₂₂₂.

*** Dosage Insensitive genes were defined by genes with a relative transcription ratio, r, 0.9<r<1.1.

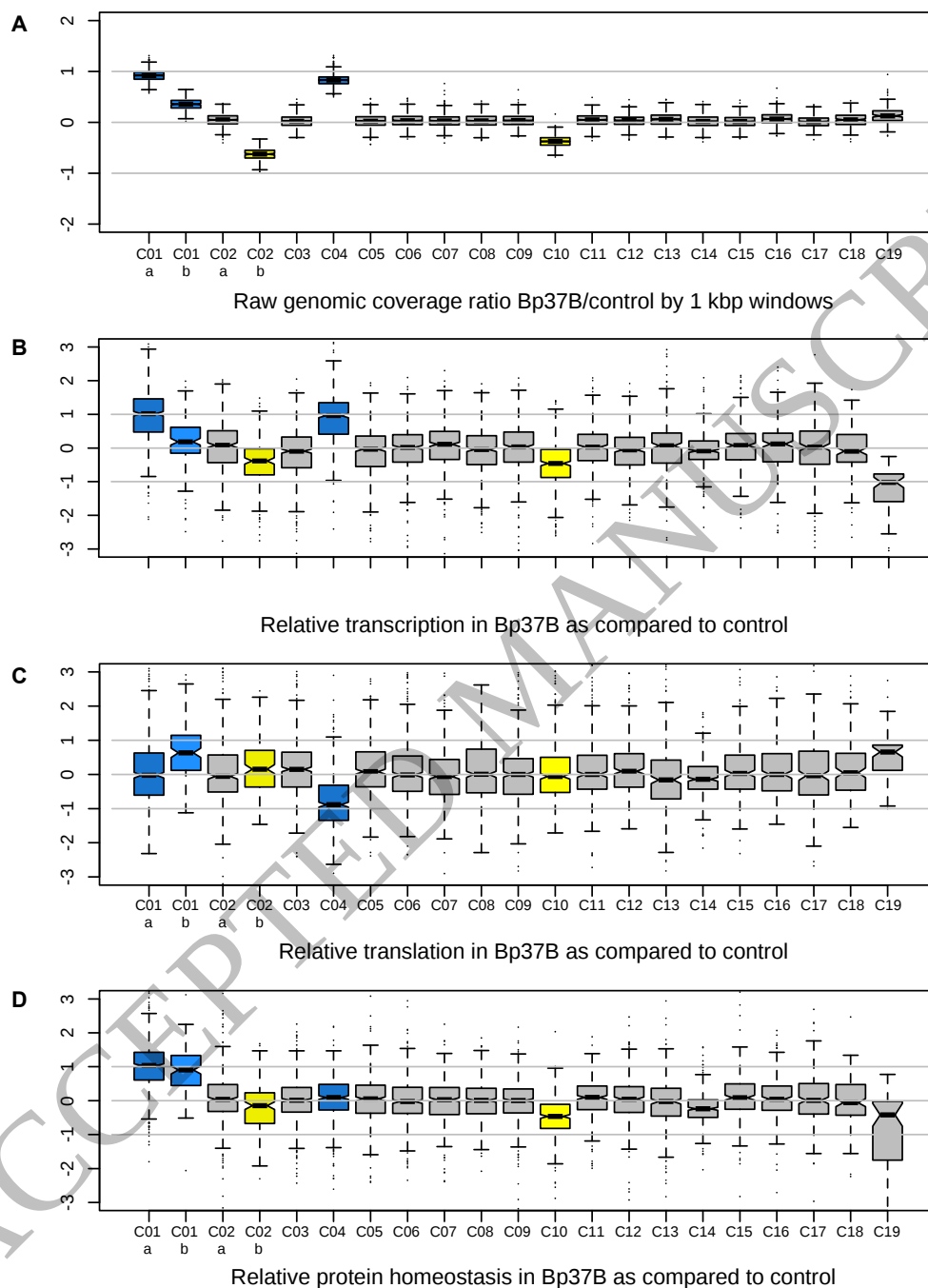


Figure 2. A. Genomic coverage of 1 kbp windows along the genome of *Bathycoccus prasinos* Bp37B. **B.** Distribution of gene transcription rates $tr(i)$ between chromosomes. **C.** Distribution of gene translation efficiencies $te(i)$ between chromosomes. **B and C.** Both the transcription and translation efficiencies ratios are significantly different between chromosomes (ANOVA, $p\text{-value} < 10^{-15}$). **D.** Distribution of protein

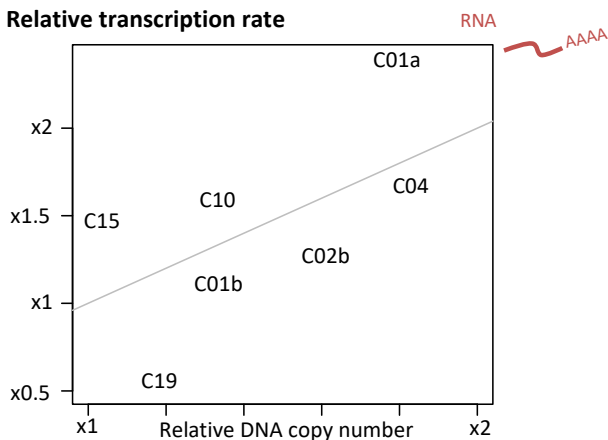
homeostasis (transcription rate x translation rate) between chromosomes. Y axes are in log2. Blue: duplicated chromosome in Bp37B. Yellow: duplicated chromosome in control RCC4222.

Consequences of chromosome duplication on translation

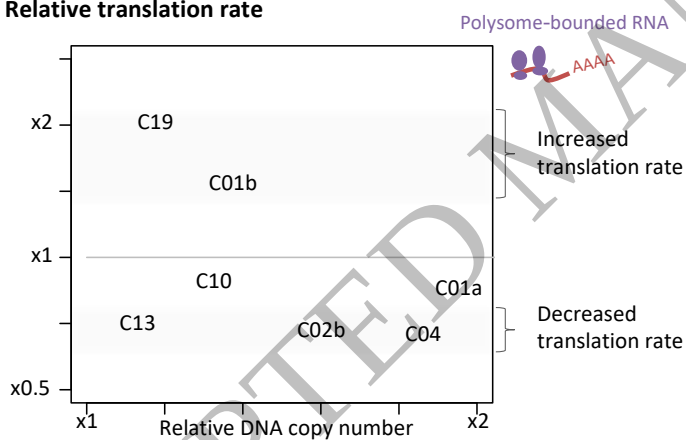
The compelling evidence of overtranscription of the majority of duplicated genes prompted us to investigate whether post-transcriptional processes may temperate this excess of transcripts. This hypothesis was tested in the same *B. prasinus* line by sequencing mRNAs associated with ribosomes (polysomes) in order to compare the translation efficiency $te(i)$ of genes located on duplicated and non-duplicated lines, by estimating the relative proportion of ribosome-bound mRNA in single versus duplicated genes. We found that the average translation efficiency of genes located on duplicated chromosome C04 was 0.71 (median=0.54) as compared to the genes on this chromosome in the control line (Figure 2C). However, the translation efficiency of genes on region C01a was 0.93, while it was 1.54 for genes on region C01b, 1.04 and 0.76 for C02a and C02b, respectively, and 0.99 for chromosome C10. Translation rate modification is thus chromosome dependent (Figure 3): genes located on the C04 and C02b show a significant decrease in translation efficiency in the line with the duplicated regions, while genes located on duplicated region C01b show a significant increase in translation efficiency in the line with the duplicated region. Last, we estimated the expected protein production in Bp37B as compared to the control line by using to the absolute translation rate for each gene: that is the ratio of mRNA in polysomes in Bp37B as compared to the control (Figure 2D). This predicted that genes on C04 have a similar protein production rate in Bp37B and in the control line, despite a higher transcription rate as a consequence of the chromosome duplication, as well for the two parts of the chromosome C02 (Figure 2D). However, genes located on chromosome C01, despite different transcription and translation efficiency rates, produce twice the level of protein compared to the single copy genes in the control line. Likewise, genes on chromosome C10 produce an excess of proteins compared to the single copy genes in Bp37B. Altogether, we observed no dosage compensation at the transcription level for any of the duplicated regions, whereas dosage compensation occurs at the translational level in 2 out of 5 duplicated regions and is thus chromosome dependent (Figure 3). In

1 addition, both transcription and translation efficiency averages are significantly different
2 (ANOVA, $p\text{-value} < 10^{-15}$) among chromosomes or chromosomal regions.

A. Relative transcription rate



B. Relative translation rate



C. Relative protein synthesis

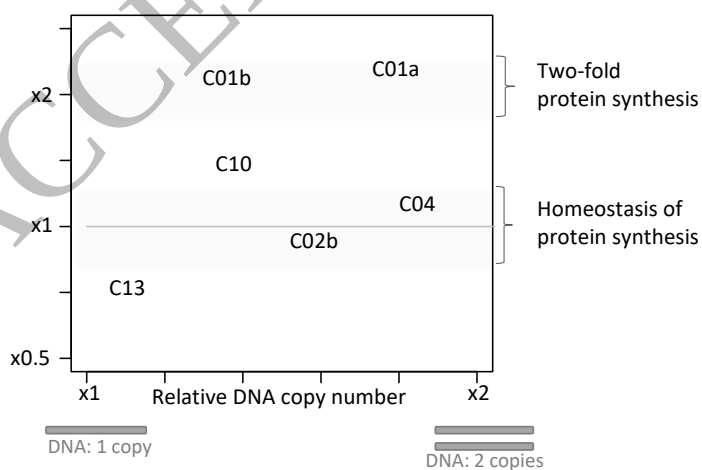


Figure 3. Raw relative genomic coverage average (Bp37B/Control) related to average transcription rate $tr(i)$ (A), translation efficiency $te(i)$ (B), and average protein synthesis (average transcription rate x average translation efficiency) per chromosome (C). Blue: chromosomes duplicated in Bp37B;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

Searching for independent molecular signatures of translation modulation

The observed dosage compensation at the translational level on chromosome C04 suggests that post-transcriptional regulation occurs on transcripts from duplicated genes. As the length of the poly(A) tail is a key feature of many cytoplasmic mRNAs and is known to regulate translation (Subtelny et al. 2014; Eichhorn et al. 2016; Lim et al. 2016), we tested whether the translation of transcripts linked to duplicated chromosome C04 might be associated with a variation of poly(A) tail length. Using 3'RACE experiments, we measured the poly(A) tail length in Bp37B and RCC4222 (Figure 4 and S18) transcripts from three genes located on chromosome C04 (duplicated in line Bp37B: 04g00840, 04g01730 and 04g04360), one gene duplicated in the control line located on chromosome C01a (01g01300), and one control gene located on chromosome C08 (08g03470, single copy in both lines). Raw transcription TPM data of these five genes are provided in Table S8. We first ligated an RNA adapter (RA3) to the 3'end of RNAs before reverse-transcription. Then, the poly(A) tail was PCR amplified with a gene-specific primer and the RA3 anchor primers. Finally, the PCR products were analysed by gel electrophoresis. Short poly(A) tails tend to generate a band. However, in the case of long poly(A) tails, amplicons present diverse poly(A) tail lengths and create a smear on the gel (Figure 4A). Additionally, an internal control PCR was performed using a gene-specific primer in the 3'UTR to determine the size without a poly(A) tail (Figure 4B). For genes located on chromosome C04, a smear is observed in Bp37B and RCC4222 (Figure 4A). However, the smear range size decreases in Bp37B compared to RCC4222 for the three transcripts tested suggesting the polyA tail size is shorter in Bp37B. No difference is observed for the two other genes located on a non-duplicated chromosome (Figure S18).

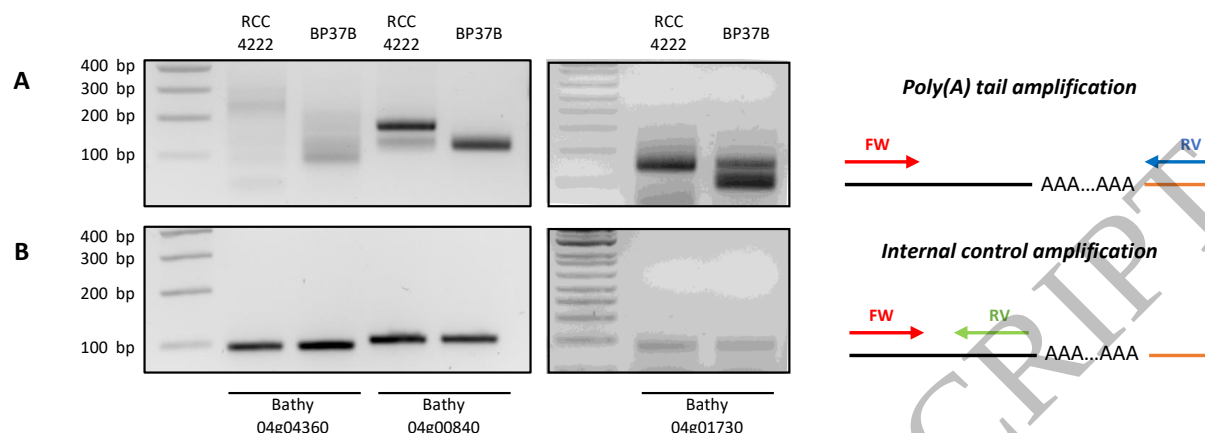


Figure 4. Transcripts from genes located on *B. prasinos* duplicated chromosome C04 present reduced poly(A) tail lengths. Poly(A) tail measurement was performed using modified 3'RACE. PCR amplification was performed using primers flanking poly(A) tail (A) or primers anchored in 3'UTR just before poly(A)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

DISCUSSION

Rates of whole chromosome duplication in unicellular eukaryotes

The rates of spontaneous whole chromosome duplication reported here in four out of six species vary between 6×10^{-4} to 1×10^{-3} events per haploid genome per generation and are thus one to two orders of magnitude lower than spontaneous point mutations per genome per generation. Theoretically, there are two possible mechanisms at the origin of chromosome duplication i) the supernumerary replication of one chromosome before cell division, leading to one cell with one chromosome and one cell with two chromosomes; or ii) the unequal segregation of chromosomes during cell division, leading to one cell with two copies and one cell without any copy. The latter scenario has been intensively experimentally studied in human cells (Ford and Correll 1992; Cimini et al. 2001), and seems all the more likely in the extremely small-sized Mamiellophyceae cells (1 μ m cell diameter) as they have been reported to contain less kinetochore microtubules than chromosomes (Gan et al. 2011), which may induce a high error rate in the chromosome segregation process. However, the whole

1 chromosome duplication rates reported here in the Mamiellophyceae are in line with
2 those found in the diatom *P. tricornutum* and in yeast (9.7×10^{-5}) (Zhu et al. 2014), yeast
3 containing approximately as many kinetochore microtubules as chromosomes
4 (Peterson and Ris 1976). Altogether, the spontaneous whole chromosome duplication
5 rates reported previously in yeast (Zhu et al. 2014) and in the species from evolutionary
6 distant eukaryotic lineages reported here (Yoon et al. 2004) suggest a high rate of
7 spontaneous aneuploidy across unicellular eukaryotes. To investigate whether
8 spontaneous whole chromosome duplication had been overlooked in other lineages, we
9 screened publicly available resequencing data of mutation accumulation lines from the
10 freshwater green alga *Chlamydomonas reinhardtii* (Ness et al. 2015), and identified one
11 event for chromosome 14 in the strain CC1373. Evidence of spontaneous whole
12 chromosome duplication (four chromosomes in a single individual from a pedigree) has
13 also been recently observed in the brown alga *Ectocarpus* (Krasovec et al. 2022).

14 Interestingly, we were able to estimate the rate of chromosome duplication loss during
15 one experiment, and it is about three orders of magnitude higher than the spontaneous
16 duplication rate. This is consistent with previous observations in *S. cerevisiae*, and our
17 observations that cells with duplicated chromosomes are ephemeral and unlikely to be
18 maintained for a long time in batch culture. Indeed, the coverage of the large duplication
19 events on C01b, C02b, C04 and C10 were 1.4, 1.6, 1.8 and 1.4 times the coverage of
20 single copy chromosomes respectively, suggesting that the proportion of cells carrying a
21 supernumerary chromosome within the population is 0.4, 0.6 and 0.8. Consistent with
22 this, we observed the loss of the duplication of chromosome C04 in one of the Bp37B
23 lines over 4 months of sub-culturing.

24 As opposed to the aneuploidy generated during meiotic divisions, which has been
25 intensively studied in mammalian cells (Nagaoka et al. 2012), the chromosomal
26 duplication events reported here are generated during mitotic division. Since the
27 molecular mechanisms involved in chromosomal segregation in mitosis, meiosis 1 and
28 meiosis 2, are different, the associated aneuploidy rates are not expected to be equal.
29 The aneuploidy rates in mitotically dividing human cells, such as HCT116 lines, has
30 been estimated to be 7×10^{-2} (Thompson and Compton 2008), more than one order of
31 magnitude higher than the highest estimation of 1.2×10^{-3} reported here in *B. prasinos*.

The observed difference of the spontaneous point mutation rate per genome per generation is expected to vary between species as the consequence of different genome sizes, effective population sizes (Lynch et al. 2016), and the difference between the observed and expected GC content (Krasovec et al. 2017). However, the eight spontaneous aneuploidy rates currently available are not yet sufficient to investigate the origin of chromosome duplication rate variations, such as the effect of chromosome number or effective population size.

Fitness effect of whole chromosome duplication and dosage compensation

The high spontaneous rates of spontaneous chromosome duplication challenge the common idea that whole chromosome duplications are highly deleterious for individual cells, at least in mitotically dividing cells. Deleterious effects of chromosome duplication may be higher in meiosis than in mitosis as the molecular mechanisms involved in meiosis comprise additional checkpoints (Roeder and Bailis 2000). Using the number of cell divisions per days as a proxy of fitness during the mutation accumulation experiment (Krasovec et al. 2016), only one out of the 15 mutation accumulation lines with a whole chromosome duplication displayed a significant fitness decrease with generation time (Bp28b, Pearson correlation, p -value=0.016, ρ =-0.846, Table S9). Recent studies in yeast point towards slightly deleterious effects of aneuploidy, as well as a significant effect of the genetic background on aneuploidy tolerance (Scopel et al. 2021). Compensating point mutations have been previously linked to aneuploidy tolerance in yeast (Torres et al. 2010), notably in the gene encoding the deubiquitinating enzyme UBP6. Point mutations are unlikely to impact aneuploidy tolerance in the different mutation accumulation lines reported here, because some lines with chromosome duplications do not carry any point mutation (Mc3, Bp28b, Bp26 and Bp25), or only carry synonymous mutations and mutations located in intergenic regions (i.e. Om3 or Mc28, Table S10). The deleterious effect of aneuploidy could be mitigated by a dosage compensation mechanism preserving relative protein homeostasis. In *B. prasinos*, despite the huge gene-to-gene variation in transcription and translation rates, transcription rates were overwhelmingly scaled up with chromosome copy number (Figure 3A), whereas translation rates differed significantly between chromosomes

(Figure 3B). Nevertheless, genes on chromosome C19 displayed a unique pattern of DNA copy number-independent transcription and translation variation between the two lines. C19 is a small idiosyncratic chromosome found in all Mamiellophyceae species sequenced so far (Blanc-Mathieu et al. 2017; Yau et al. 2020) and is characterized by a lower GC content. Previous studies reported a strong variation in transcription rates of genes on this chromosome, which is associated with resistance to viruses (Yau et al. 2016).

In *B. prasinos*, genes located on duplicated regions fell into three different categories based on their translation efficiency. First, there were genes with no relative difference in translation rates, such as genes located on duplicated chromosome C10, which is duplicated in 40% percent of cells, and on region C01a, which is duplicated in 100% of cells. Proteins encoded by these genes were inferred to be overabundant in the cell (Figure 3C). Second, there were genes with a decreased translation rate on chromosome C04 and region C02b. More precisely, the estimated frequency of mRNA linked to ribosomes was inversely proportional to the excess of mRNA produced for duplicated genes on these two chromosomes, predicting approximately the same protein production rate as in cells without duplicated chromosomes (Figure 3C). Third, and very surprisingly, there were genes with higher translation rates associated with higher DNA copy numbers and higher transcription rates, such as genes located in the C01b region. As a consequence, the rate of protein synthesis on genes located on region C01b reached the same level of relative protein synthesis as genes located on region C01a, that is twice the protein synthesis predicted in cells with a single copy of this chromosome. We speculate that the absence of dosage compensation on region C01a and the excess of translation on region C01b simulate the ancestral protein homeostasis in the ancestral line, which contained two complete chromosome C01s.

Chromosome-wide regulation of gene translation is relatively unexplored and poorly understood. Indeed, while the position of a gene on a chromosome may allow its transcription rate to be predicted from epigenetic marks on the DNA, chromosome-wide epitranscriptomic mRNA modification mechanisms are yet to be discovered. Notwithstanding, chromosome-wide translational dosage compensation has been

previously observed in *Drosophila* (Zhang and Presgraves 2016), and at the gene scale in several species (Zhang and Presgraves 2017; Chang and Liao 2020).

Is poly(A) tail length a potential post-transcriptional mechanism involved in dosage compensation?

The role of poly(A) tail length in translation efficiency is starting to be better understood (Weill et al. 2012; Subtelny et al. 2014), and alternative polyadenylation is indeed implicated in several processes: transcription termination by RNAP II, mRNA stability, mRNA export and translation efficiency (Zhang et al. 2010; Di Giammartino et al. 2011). In the cytoplasm, the poly(A) tail plays important roles in mRNA translation and stability and the modulation of its length has an important impact on translation efficiency (Subtelny et al. 2014; Eichhorn et al. 2016; Lim et al. 2016). For example, during oocyte maturation and early embryonic development, an increase in poly(A) tail length occurs for particular mRNAs resulting in an increase of translation (Subtelny et al. 2014; Eichhorn et al. 2016; Lim et al. 2016). This modulation of poly(A) tail length has also been found to activate some neuronal transcripts (Udagawa et al. 2012). Slobadin et al (Slobadin et al. 2020), suggested that poly(A) tails can be modulated to balance mRNA levels and adjust translation efficiency since they observed that the CCR4-Not complex shortened the poly(A) tails, which reduced the stability of mRNAs. Here, our analysis of a subset of three genes suggests that, in a context of chromosome duplication, modulation of poly(A) tail length could be one of the post-transcriptional mechanisms involved in dosage compensation. However, the limited sample size of our observations does not allow us to draw any conclusion. Future genome-wide analyses will be necessary to validate this modulation and investigate the role of other mechanisms, such as adenosine methylation, in chromosome-wide translation compensation (Miao et al. 2022).

Transcription factories, nuclear structures where co-expressed or physically close genes are transcribed, have also been proposed to explain chromosome-dependent dosage compensation (Sutherland and Bickmore 2009). Our observation of a chromosome-scale similar translational behaviour of transcripts could be due to the action of a single factory bound to that chromosome, while other chromosomes are

under the control of their own transcription factories. The existence of such factories in unicellular eukaryotic species is unknown and needs further future investigations.

CONCLUSION

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

MATERIALS AND METHODS

Sequencing data from mutation accumulation (MA) experiments

Mutation accumulation (MA) experiments of six species; *Picochlorum costavermella* RCC4223 (Krasovec, Vancaester, et al. 2018), *Ostreococcus tauri* RCC4221 (Blanc-Mathieu et al. 2014), *O. mediterraneus* RCC2590 (Subirana et al. 2013; Yau et al. 2020), *Bathycoccus prasinos* RCC1105 (synonym to RCC4222) (Moreau et al. 2012), *Micromonas commoda* RCC299 (Worden et al. 2009) and *Phaeodactylum tricornutum* RCC2967 (Bowler et al. 2008; Giguere et al. 2022) were conducted with a flow cytometry protocol described previously for phytoplankton species in liquid medium (Krasovec et al. 2016). Briefly, a mutation accumulation experiment consists in monitoring MA lines that have evolved from a same cell (the ancestral line) during hundreds of generations. Relaxed selection pressure on spontaneous mutations is ensured by maintaining all MA lines at very low effective population sizes ($6 < N_e < 8.5$) throughout the experiment (Krasovec et al. 2016). As a consequence, comparison of the complete genome sequence of the ancestral and MA lines enables the direct estimation of spontaneous mutation rates, excluding lethal mutations. Here, MA lines came from a single cell obtained by dilution serving as T_0 culture (named the ancestral line, AL) and were maintained in 24-wells plates (with one well per MA line) in L1 medium at 20 °C

with a 16h-dark 8h-light life cycle. One-cell bottlenecks were performed by dilution every 14 days to keep a low effective population size to limit selection. Effective population size was calculated by the harmonic mean of cell number over the 14-day period (Krasovec et al. 2016). The number of cell divisions per day between bottlenecks was calculated from the total number of cells obtained by flow cytometry for each MA line (Krasovec et al. 2016), knowing that inoculation is done with one cell. DNA of initial line (ancestral line) and final time of MA lines were extracted with chloroform protocol and sequencing done with Illumina HiSeq or MiSeq by GATC biotech (Germany). To detect duplications, raw reads were mapped against the reference genomes for each strain with bwa mem v0.7.12 (Li and Durbin 2010). Next, bam files were treated with samtools v1.3.1 (Li et al. 2009) and bedtools v 2-2.18.0 (Quinlan and Hall 2010) to extract the coverage.

Probability calculations

To test whether our observations were compatible with the null hypothesis of “equal duplication rates between chromosomes” in each species, we calculated the probability of observing two independent duplications in the same chromosome for a given duplication events. The probability of drawing one chromosome twice out of k WCD events chosen from n chromosomes, $P(k,n)$, is equal to the ratio of the number of combinations of $(k-1)$ out of n without replacement multiplied by $k-1$, which corresponds to the number of possible $k-1$ different chromosomes with one in two copies, out of n , divided by the number of combinations of k out of n with replacement $(k-1)*C(k-1,n)/(n+k-1)$, which corresponds to the total number of possible subsamples of k chromosomes out of n .

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

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

Dosage compensation analysis

This study was started three years after the end of the MA experiments, for which some MA lines had been cryopreserved which allowed us to re-start a culture for one *B.*

prasinus MA line with a duplicated chromosome C04 (Bp37) to investigate the effect of the duplication on transcription and translation rates. The restarted culture from the cryopreserved Bp37 was renamed Bp37B. As a control, we used the reference culture of the strain *B. prasinus* RCC4222, that is the derived from the RCC1105 used for the MA experiment. All cultures (Bp37B and RCC4222) were maintained under a 12:12 h light:dark regime under 50 $\mu\text{mol photon m}^{-2} \text{s}^{-1}$ white light at 20 °C. The karyotype of the defrozen cultures Bp37B and the control RCC4222 were checked by DNaseq resequencing at the Bioenvironment platform (UPVD, Perpignan). The karyotype of Bp37B contained on additional chromosome copy number as compared to Bp37, and the karyotype of RCC4222 contained one additional copy numbers and one less chromosome copy number as compared to the ancestral line (AL) (Figure S17). The relative transcription rate was corrected by the number of DNA copy in each line for further analysis below.

For transcription analyses, total RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo Research, Californie, USA) from pooling flasks of cultures (100 ml cultures with 200 million cells per ml) taken 6h before and 1h before the light on, in triplicates for the control RCC4222 and duplicates for Bp37B. Next, polysome extraction was performed for Ribo-seq as described previously (Carpentier et al. 2020) with few modifications. Briefly, 600 mL of *B. prasinus* culture were centrifuged at 8,000 g for 20 minutes. After centrifugation, pellets were resuspended in 2.4 mL of polysome extraction buffer. After 10 minutes of incubation on ice and centrifugation, 2 mL of supernatant was loaded on a 9 mL 15–60% sucrose gradient and centrifuged for 3h at 38 000 rpm with rotor SW41 Ti. Fractions corresponding to polysomes were pooled and polysomal RNA was extracted as previously described (Carpentier et al. 2020). RNA library preparation was performed on total or polysomal RNA using a NEBNext Poly(A) mRNA Magnetic Isolation Module and a NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs) according to the manufacturer's instructions with 1 μg of RNA as a starting point. Libraries were multiplexed and sequenced on a NextSeq 550. Raw reads were mapped against the reference transcriptome of *B. prasinus* with RSEM with standard parameters (Li and Dewey 2011). We obtained the TPM average of total RNA and polysomes linked RNA that we compared between *B. prasinus* RCC4222 and

the MA line Bp37B. TPM stand for transcripts per kilobase million, and the sum of all TPM values is the same in all samples, such that a TPM value represents a relative transcription level that is comparable between samples. At the gene scale, we first used the total RNA TPM values to estimate the transcription difference ratio, r_{RNA} for each gene i , between two copies and single copy genes as:

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

These relative transcription rates $r_{RNA}(i)$ were normalized by the transcription rate median of genes located on non-duplicated chromosomes in Bp37B (C03, C05, C06, C07, C08, C09, C011, C012, C013, C014, C015, C016, C017 and C018), in order to estimate the transcription rate $tr(i)$ of duplicated chromosomes related to non-duplicated chromosomes. Chromosome C19 was not considered because it is an idiosyncratic chromosome (see Discussion).

Second, we used the total RNA and polysomes linked RNA TPM values for each gene i to calculate the translation rates of each gene, $r(i)$, as:

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

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

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

The $te(i)$ ratio was then normalized by the median translation efficiency of all genes on non-duplicated chromosomes in strain BP37B (5,267 genes).

To estimate the dosage compensation at the scale of the chromosomes, the values were normalized by the average TPM of all single copy chromosomes prior to the calculation of the ratios (Table S7).

Poly(A) tail analysis was performed as previously described with slight modifications (Sement and Gagliardi 2014). PCR products were resolved on a 2.5% agarose gel. Primers used in this study are available in Table S11.

Differential gene transcription analyses

The statistical significance of the genes differential transcription levels was further estimated with Deseq2 (Love et al. 2014). Transcriptional invariant genes were defined as genes present on duplicated regions and for which the relative transcription rate $r_{RNA}(i)$ was comprised between 0.9 and 1.1. Genes with no transcription were removed from this analysis. The over-representation of a certain GO term in the transcriptional invariant gene set was compared to the genome-wide GO term background frequency using the GO enrichment analysis with default values implemented in pico-PLAZA workbench (Vandepoele et al. 2013).

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

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

COMPETING INTERESTS

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Genomic raw reads are available under the bioprojects PRJNA531882 (*Ostreococcus tauri*, *O. mediterraneus*, *Micromonas commoda*, *Bathycoccus prasinus*), PRJNA453760 and PRJNA389600 (*Picochlorum costaverdella*), and PRJNA478011 (*Phaeodactylum tricornutum*). Transcriptomic raw reads of the *Bathycoccus prasinus* MA lines 37 (Bp37B) and control line RCC4222 are available under the bioproject PRJNA715163. A summary is provided in Table S12.

BIBLIOGRAPHY

- Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat. Rev. Genet.* 5:725–738.
- Baker BS, Gorman M, Marín I. 1994. Dosage compensation in *Drosophila*. *Annu. Rev. Genet.* 28:491–521.
- Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, Martin J, Schackwitz W, Kuo A, Salin G, Donnadieu C, et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* 3:e1700239.
- Blanc-Mathieu R, Verhelst B, Derelle E, Rombauts S, Bouget F-Y, Carré I, Château A, Eyre-Walker A, Grimsley N, Moreau H, et al. 2014. An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics* 15:1103.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–244.
- Carpentier M-C, Deragon J-M, Jean V, Be SHV, Bousquet-Antonelli C, Merret R. 2020. Monitoring of XRN4 Targets Reveals the Importance of Cotranslational Decay during *Arabidopsis* Development. *Plant Physiol.* 184:1251–1262.
- Chang AY-F, Liao B-Y. 2020. Reduced Translational Efficiency of Eukaryotic Genes after Duplication Events. *Mol. Biol. Evol.* 37:1452–1461.
- Charlesworth D. 2019. Young sex chromosomes in plants and animals. *New Phytol.* 224:1095–1107.
- Chen G, Bradford WD, Seidel CW, Li R. 2012. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* 482:246–250.
- Cimini D, Howell B, Maddox P, Khodjakov A, Degraffi F, Salmon ED. 2001. Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue Cells. *J. Cell Biol.* 153:517–528.

- 1 Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, Torres EM. 2014.
- 2 Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Dikic
- 3 I, editor. *eLife* 3:e03023.
- 4 Devlin RH, Holm DG, Grigliatti TA. 1988. The influence of whole-arm trisomy on gene
- 5 expression in *Drosophila*. *Genetics* 118:87–101.
- 6 Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative
- 7 polyadenylation. *Mol. Cell* 43:853–866.
- 8 Distech CM. 2012. Dosage Compensation of the Sex Chromosomes. *Annu. Rev. Genet.* 46:537–
- 9 560.
- 10 Eichhorn SW, Subtelny AO, Kronja I, Kwasniewski JC, Orr-Weaver TL, Bartel DP. 2016. mRNA
- 11 poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila*
- 12 oocytes and early embryos. Izaurralde E, editor. *eLife* 5:e16955.
- 13 FitzPatrick DR, Ramsay J, McGill NI, Shade M, Carothers AD, Hastie ND. 2002. Transcriptome
- 14 analysis of human autosomal trisomy. *Hum. Mol. Genet.* 11:3249–3256.
- 15 Ford JH, Correll AT. 1992. Chromosome errors at mitotic anaphase. *Genome* 35:702–705.
- 16 Gan L, Ladinsky MS, Jensen GJ. 2011. Organization of the Smallest Eukaryotic Spindle. *Curr.*
- 17 *Biol.* 21:1578–1583.
- 18 Gearhart JD, Oster-Granite ML, Reeves RH, Coyle JT. 1987. Developmental consequences of
- 19 autosomal aneuploidy in mammals. *Dev. Genet.* 8:249–265.
- 20 Giguere DJ, Bahcheli AT, Slattery SS, Patel RR, Browne TS, Flatley M, Karas BJ, Edgell DR,
- 21 Gloor GB. 2022. Telomere-to-telomere genome assembly of *Phaeodactylum tricornutum*. *PeerJ*
- 22 10:e13607.
- 23 Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat.*
- 24 *Rev. Genet.* 17:33–46.
- 25 Halligan DL, Keightley PD. 2009. Spontaneous Mutation Accumulation Studies in Evolutionary
- 26 Genetics. *Annu. Rev. Ecol. Evol. Syst.* 40:151–172.
- 27 Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. *Annu. Rev. Genet.*
- 28 31:571–610.
- 29 Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M,
- 30 Kaessmann H, Reymond A. 2009. Segmental copy number variation shapes tissue
- 31 transcriptomes. *Nat. Genet.* 41:424–429.
- 32 Hose J, Escalante LE, Clowers KJ, Dutcher HA, Robinson D, Bouriakov V, Coon JJ, Shishkova
- 33 E, Gasch AP. 2020. The genetic basis of aneuploidy tolerance in wild yeast. Klein H, Tyler JK,
- 34 Gresham D, Argueso JL, Bloom K, editors. *eLife* 9:e52063.
- 35 Hou J, Shi X, Chen C, Islam MS, Johnson AF, Kanno T, Huettel B, Yen M-R, Hsu F-M, Ji T, et
- 36 al. 2018. Global impacts of chromosomal imbalance on gene expression in *Arabidopsis* and other
- 37 taxa. *Proc. Natl. Acad. Sci.* 115:E11321–E11330.
- 38 Kaya A, Mariotti M, Tyshkovskiy A, Zhou X, Hulke ML, Ma S, Gerashchenko MV, Koren A,
- 39 Gladyshev VN. 2020. Molecular signatures of aneuploidy-driven adaptive evolution. *Nat.*
- 40 *Commun.* 11:588.

- 1 Kojima S, Cimini D. 2019. Aneuploidy and gene expression: is there dosage compensation?
2 *Epigenomics* 11:1827–1837.
- 3 Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju V. 2018.
4 Mutational and transcriptional landscape of spontaneous gene duplications and deletions in
5 *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* 115:7386–7391.
- 6 Krasovec M, Eyre-Walker A, Grimsley N, Salmeron C, Pecqueur D, Piganeau G, Sanchez-
7 Ferandin S. 2016. Fitness Effects of Spontaneous Mutations in Picoeukaryotic Marine Green
8 Algae. *G3 Genes Genomes Genet.* 6:2063–2071.
- 9 Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G. 2017. Spontaneous Mutation
10 Rate in the Smallest Photosynthetic Eukaryotes. *Mol. Biol. Evol.* 34:1770–1779.
- 11 Krasovec M, Lipinska AP, Coelho SM. 2022. Low spontaneous mutation rate in a complex
12 multicellular eukaryote with a haploid-diploid life cycle. :2022.05.13.491831. Available from:
13 <https://www.biorxiv.org/content/10.1101/2022.05.13.491831v1>
- 14 Krasovec M, Sanchez-Brosseau S, Grimsley N, Piganeau G. 2018. Spontaneous mutation rate as
15 a source of diversity for improving desirable traits in cultured microalgae. *Algal Res.* 35:85–90.
- 16 Krasovec M, Vancaester E, Rombauts S, Bucchini F, Yau S, Hemon C, Lebredonchel H,
17 Grimsley N, Moreau H, Sanchez-Brosseau S, et al. 2018. Genome Analyses of the Microalga
18 *Picochlorum* Provide Insights into the Evolution of Thermotolerance in the Green Lineage.
19 *Genome Biol. Evol.* 10:2347–2365.
- 20 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
21 without a reference genome. *BMC Bioinformatics* 12:323.
- 22 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform.
23 *Bioinformatics* 26:589–595.
- 24 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
25 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format
26 and SAMtools. *Bioinforma. Oxf. Engl.* 25:2078–2079.
- 27 Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A) tail
28 regulation in oocyte-to-embryo development. *Genes Dev.* [Internet]. Available from:
29 <http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116>
- 30 Liu H, Zhang J. 2019. Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment.
31 *Curr. Biol. CB* 29:1584-1591.e3.
- 32 Liu Y, Borel C, Li L, Müller T, Williams EG, Germain P-L, Buljan M, Sajic T, Boersema PJ,
33 Shao W, et al. 2017. Systematic proteome and proteostasis profiling in human Trisomy 21
34 fibroblast cells. *Nat. Commun.* 8:1212.
- 35 Loane M, Morris JK, Addor M-C, Arriola L, Budd J, Doray B, Garne E, Gatt M, Haeusler M,
36 Khoshnood B, et al. 2013. Twenty-year trends in the prevalence of Down syndrome and other
37 trisomies in Europe: impact of maternal age and prenatal screening. *Eur. J. Hum. Genet. EJHG*
38 21:27–33.
- 39 Loehlin DW, Carroll SB. 2016. Expression of tandem gene duplicates is often greater than
40 twofold. *Proc. Natl. Acad. Sci.* 113:5988–5992.

- 1 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
2 RNA-seq data with DESeq2. *Genome Biol.* 15:550.
- 3 Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE. 2004. Gene expression
4 from the aneuploid chromosome in a trisomy mouse model of down syndrome. *Genome Res.*
5 14:1268–1274.
- 6 Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic
7 drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* 17:704–714.
- 8 Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K,
9 Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous
10 mutations in yeast. *Proc. Natl. Acad. Sci.* 105:9272–9277.
- 11 Miao Z, Zhang T, Xie B, Qi Y, Ma C. 2022. Evolutionary Implications of the RNA N6-
12 Methyladenosine Methylome in Plants. *Mol. Biol. Evol.* 39:msab299.
- 13 Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, Van Bel M, Poulain J,
14 Katinka M, Hohmann-Marriott MF, et al. 2012. Gene functionalities and genome structure in
15 *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome*
16 *Biol.* 13:R74.
- 17 Muyle A, Shearn R, Marais GA. 2017. The Evolution of Sex Chromosomes and Dosage
18 Compensation in Plants. *Genome Biol. Evol.* 9:627–645.
- 19 Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid De Novo
20 Evolution of X Chromosome Dosage Compensation in *Silene latifolia*, a Plant with Young Sex
21 Chromosomes. *PLOS Biol.* 10:e1001308.
- 22 Nagaoka SI, Hassold TJ, Hunt PA. 2012. Human aneuploidy: mechanisms and new insights into
23 an age-old problem. *Nat. Rev. Genet.* 13:493–504.
- 24 Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. 2015. Extensive de
25 novo mutation rate variation between individuals and across the genome of *Chlamydomonas*
26 *reinhardtii*. *Genome Res.* 25:1739–1749.
- 27 Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel
28 K, Llored A, et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates.
29 *Nature* 556:339–344.
- 30 Peterson JB, Ris H. 1976. Electron-microscopic study of the spindle and chromosome movement
31 in the yeast *Saccharomyces cerevisiae*. *J. Cell Sci.* 22:219–242.
- 32 Qian W, Liao B-Y, Chang AY-F, Zhang J. 2010. Maintenance of duplicate genes and their
33 functional redundancy by reduced expression. *Trends Genet. TIG* 26:425–430.
- 34 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
35 features. *Bioinformatics* 26:841–842.
- 36 R Core Team (2022). R: A language and environment for statistical computing. R Foundation for
37 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 38 Rajagopalan H, Lengauer C. 2004. Aneuploidy and cancer. *Nature* 432:338–341.
- 39 Roeder GS, Bailis JM. 2000. The pachytene checkpoint. *Trends Genet. TIG* 16:395–403.

- Scopel EFC, Hose J, Bensasson D, Gasch AP. 2021. Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics* 217:iyab015.
- Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* 313:367–370.
- Sement FM, Gagliardi D. 2014. Detection of uridylated mRNAs. *Methods Mol. Biol. Clifton NJ* 1125:43–51.
- Slobodin B, Bahat A, Sehwat U, Becker-Herman S, Zuckerman B, Weiss AN, Han R, Elkon R, Agami R, Ulitsky I, et al. 2020. Transcription Dynamics Regulate Poly(A) Tails and Expression of the RNA Degradation Machinery to Balance mRNA Levels. *Mol. Cell* 78:434–444.e5.
- Song MJ, Potter BI, Doyle JJ, Coate JE. 2020. Gene Balance Predicts Transcriptional Responses Immediately Following Ploidy Change in *Arabidopsis thaliana*. *Plant Cell* 32:1434–1448.
- Stenberg P, Lundberg LE, Johansson A-M, Rydén P, Svensson MJ, Larsson J. 2009. Buffering of Segmental and Chromosomal Aneuploidies in *Drosophila melanogaster*. *PLOS Genet.* 5:e1000465.
- Subirana L, Péquin B, Michely S, Escande M-L, Meilland J, Derelle E, Marin B, Piganeau G, Desdevises Y, Moreau H, et al. 2013. Morphology, Genome Plasticity, and Phylogeny in the Genus *Ostreococcus* Reveal a Cryptic Species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist* 164:643–659.
- Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508:66–71.
- Sutherland H, Bickmore WA. 2009. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* 10:457–466.
- Thompson SL, Compton DA. 2008. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* 180:665–672.
- Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, Dunham MJ, Amon A. 2010. Identification of aneuploidy-tolerating mutations. *Cell* 143:71–83.
- Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A. 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317:916–924.
- Udagawa T, Swanger SA, Takeuchi K, Kim JH, Nalavadi V, Shin J, Lorenz LJ, Zukin RS, Bassell GJ, Richter JD. 2012. Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol. Cell* 47:253–266.
- Vandepoele K, Van Bel M, Richard G, Van Landeghem S, Verhelst B, Moreau H, Van de Peer Y, Grimsley N, Piganeau G. 2013. pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environ. Microbiol.* 15:2147–2153.
- de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, et al. 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605.
- Veitia RA, Potier MC. 2015. Gene dosage imbalances: action, reaction, and models. *Trends Biochem. Sci.* 40:309–317.
- Weill L, Belloc E, Bava F-A, Méndez R. 2012. Translational control by changes in poly(A) tail

- length: recycling mRNAs. *Nat. Struct. Mol. Biol.* 19:577–585.
- Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, Amon A. 2008. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* 322:703–709.
- Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324:268–272.
- Yau S, Hemon C, Derelle E, Moreau H, Piganeau G, Grimsley N. 2016. A Viral Immunity Chromosome in the Marine Picoeukaryote, *Ostreococcus tauri*. *PLoS Pathog.* 12:e1005965.
- Yau S, Krasovec M, Benites LF, Rombauts S, Groussin M, Vancaester E, Aury J-M, Derelle E, Desdevise 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.