High rates of spontaneous chromosomal duplications are compensated by translational regulation in a photosynthetic unicellular eukaryote

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

High rates of spontaneous chromosomal duplications are compensated by translational regulation in a photosynthetic unicellular eukaryote

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 author: krasovec@obs-banyuls.fr ; remy.merret@univ-perp.fr

Keywords: chromosome duplication, mutation accumulation, aneuploidy, dosage compensation, translation efficiency, poly(A) tail length, phytoplankton.

Running title: Spontaneous chromosome duplication rates.

ABSTRACT

While duplications have long been recognized as a fundamental process driving major evolutionary innovations, direct estimations of spontaneous chromosome duplication rates are scarce. Here, we provide the first estimations of spontaneous whole chromosome duplication rates in four green algae and in one diatom species from mutation accumulation (MA) experiments. The spontaneous whole duplication events are 7 to 70 times less frequent than spontaneous point mutations per cell division. Comparative transcriptome analyses between the control versus a MA line with a whole chromosome duplication indicate a ~1.94-fold higher relative level of mRNAs in genes located within the duplicated regions. However, comparative analyses of the translation rate of mRNAs demonstrate that the excess of mRNAs is compensated by a proportional decrease to ~0.67 in translation rates of genes located on duplicated chromosomes. This provides evidence that translation regulation mechanisms orchestrate dosage compensation in duplicated chromosomes. We further investigate the role of the poly(A) tail length in tempering mRNA translation and suggest that poly(A) tail length is regulated and shorter for transcripts linked to the chromosome duplication. These results point out the existence of a post-transcriptional mechanism co-ordinating the translation of hundreds of transcripts from genes located on duplicated regions in eukaryotes.

INTRODUCTION

Copy number variants or genome-chromosome duplications have huge implications in evolution and can be involved in various processes such as speciation and the evolution of novel functions (Ohno 1970; Van de Peer et al. 2009; Flagel and Wendel 2009). Well documented cases include whole genome duplication in ancestral vertebrate (Dehal and Boore 2005) or multiple whole genome duplications in plants (Vanneste et al. 2014; Scott et al. 2016; Segraves 2017; Smith et al. 2018). However, when a single chromosome is duplicated, an immediate issue arises: the imbalance of ploidy and gene dose within a same karyotype. This is deleterious because it creates an imbalance of the transcript production, which is costly and can disrupt the function of a pathway and gene interactions (Veitia and Potier 2015). In *Caenorhabditis elegans*, mutation accumulations at different effective population sizes showed purifying selection against duplications causing excess of transcripts as compared to transcript levels in the ancestral state (Konrad et al. 2018). Aneuploid karyotypes are also known to be at the origin of genetic diseases, such as Trisomy 21 in humans. The effects of aneuploidy lead to the evolution of mechanisms that restore the ancestral gene dose by modifying the transcription level, such as the dosage compensation (DC) during the evolution of heterogametic sex chromosomes, known both in plants (Muyle et al. 2012, 2017; Charlesworth 2019) and animals (Disteche 2012; Graves 2016). As a consequence of the degeneration of the non-recombining sex chromosome, such as the Y in human or *Drosophila*, the X-linked genes are haploid in males. Different DC mechanisms evolved, either by simulating the ancestral ploidy (by doubling the expression of the genes of the single copy male X) or by equalizing the ploidy between the two sexes (by halving the expression – silencing of one X in female). Although DC is well studied in sex chromosome evolution, how DC evolves shortly after a gene dose – chromosome ploidy variation is unclear, particularly for autosomes. In *Drosophila*, it has been reported that gene duplication induces a significant increase of expression for the duplicated genes (Loehlin and Carroll 2016). In contrary, in yeast and mammals, gene duplication has been associated with a decrease of the expression of the two copies, suggesting immediate DC at the transcriptional level (Qian et al. 2010; Henrichsen et al. 2009). In the case of whole chromosome duplication, disomic yeasts show a dramatic increase in the expression of genes located in the duplicated chromosome (Kaya et al. 2020). In *Arabidopsis*, the consequences of aneuploidy on gene expression are complex and

highly variable as gene expression changes spread outside the duplicated regions (Song et al. 2020; Hou et al. 2018). These previous studies show that the transcriptional response and immediate DC following a duplication are variable between species and genotypes.

Here, we investigate the spontaneous whole chromosome duplication rate in six photosynthetic unicellular eukaryotic species from distant lineages (Trebouxiophyceae, Mamiellophyceae, Bacillaryophyceae) with mutation accumulation (MA) experiments. Mutation accumulation experiment provides the most accurate method to directly estimate the spontaneous mutation rate. The principle is to follow mutation accumulation lines originated from an inbred couple (such as in *Drosophila*) or a clone (such as in yeast) under minimal selection obtain by serial bottlenecks during dozen to thousands generations (Halligan and Keightley 2009). Mutation accumulation experiments were done previously by maintaining 12 to 40 mutation accumulations line which accumulated 1,595 to 17,250 generations (Krasovec et al. 2017, 2018a, 2019). Whole chromosome duplications detected here are part of structural mutations, such as large insertions-deletions, inversions or chromosome rearrangements and may have huge phenotypical effect, much more than a nucleotide or short insertion-deletion mutations because they impact a larger proportion of the genome. However, studies on the structural mutation rate are very limited (Press et al. 2019; Ho and Schaack 2021) making this study a significant step in our understanding of structural mutation rate in unicellular eukaryote. We then estimate the effect of such chromosome duplication on the transcription and translation rates by coupling transcriptomics with translatomics sequencing in one cryopreserved mutation accumulation line of *Bathycoccus prasinos*.

RESULTS

Whole chromosome duplication rate

In this study, we leveraged sequencing data from mutation accumulation (MA) experiments (Krasovec et al. 2017, 2018a, 2016) to estimate the whole chromosome duplication rate in five haploid green algae species (Chlorophyta) and one diploid diatom (Bacillariophyta) : *Picochlorum costavermella* RCC4223 (Krasovec et al. 2018b), *Ostreococcus tauri* RCC4221 (Blanc-Mathieu et al. 2014), *O. mediterraneus* RCC2590 (Subirana et al. 2013) , *Bathycoccus prasinos* RCC1105 (Moreau et al. 2012), *Micromonas pusilla* RCC299 (Worden et al. 2009) and *Phaeodactylum*

tricornutum RCC2967 (Krasovec et al. 2019). Sequence coverage was used as a proxy for copy number (Figure 1A and Supplemental figures S1 to S16), and analyses unveiled four independent whole chromosome duplications in the species *M. pusilla* (chromosomes 5, 12, 16 and 17), four in *B. prasinos* (chromosomes 4, 5, 6 and 19), one in *O. mediterraneus* (chromosome 14) and 6 in the species *P. tricornutum* (chromosomes 2, 12, 15, 30 and 31). The chromosome duplication events can be mapped onto the genealogies of the MA lines to identify the independent whole chromosome duplication events (Figure 1B for *B. prasinos*). Several chromosomes were independently duplicated two times: Bp25c and Bp28b carry two independent duplications of chromosome 19; Mp08 and Mp09 carry two independent duplications of chromosome 19; and Pt11 and Pt10c carry two independent duplications of chromosomes 30 and 31. All independent duplication events inferred from coverage analyses and genealogies are summarised in Table 1 and S1. In addition to chromosome duplications, there have been 11 independent whole chromosome duplication losses in *B. prasinos*. The ancestral line of *B. prasinos* carried two copies of chromosome 1, and one of the copies had subsequently been lost in 23 daughter lines over 4145 generations, corresponding to a spontaneous chromosome loss of 0.006 per duplicated chromosome per generation.

Table 1. Spontaneous whole chromosome duplication rate in six species. N_{lines} is the number of MA lines, Gen the average number of generations per MA line, Tot_{Gen} the total number of generations, N_{Chrom} is the number of chromosomes in the ancestral karyotype, N_{WCD} is the number of independent whole chromosome duplications, U_{WCDl} is the whole chromosome duplication rate per chromosome per cell division, and U_{Cell} is the whole chromosome duplication rate per cell division, U_{bs} is the base substitution mutation rate per cell division. N_{WCD} *and* U_{cell} of *O. tauri* and *P. costavermella* are calculated assuming one duplication event. The whole chromosome duplication rate per cell division represents 5% of the rate of spontaneous point mutations in these species in average.

Species	N_{lines}	**Gen**	Tot_{Gen}	N_{Chrom}	N_{WCD}	U_{WCD}	U_{cell}	U_{bs}
Bathycoccus prasinos RCC1105	35	265	4,145	19	4	0.000051	0.00097	0.0066
Micromonas pusilla RCC299	37	272	4,994	17	5	0.000059	0.00100	0.0205
Ostreococcus tauri RCC4221	40	512	17,250	20	0	<0.000003	<0.00006	0.0062
Ostreococcus mediterraneus RCC2590	33	272	8,380	19	1	0.000006	0.00012	0.0081
Picochlorum costavermella RCC4223	12	133	1,596	10	0	<0.00006	<0.0006	0.0132
Phaeodactylum tricornutum RCC2967	36	181	6,516	33	7	0.000033	0.00107	0.0125

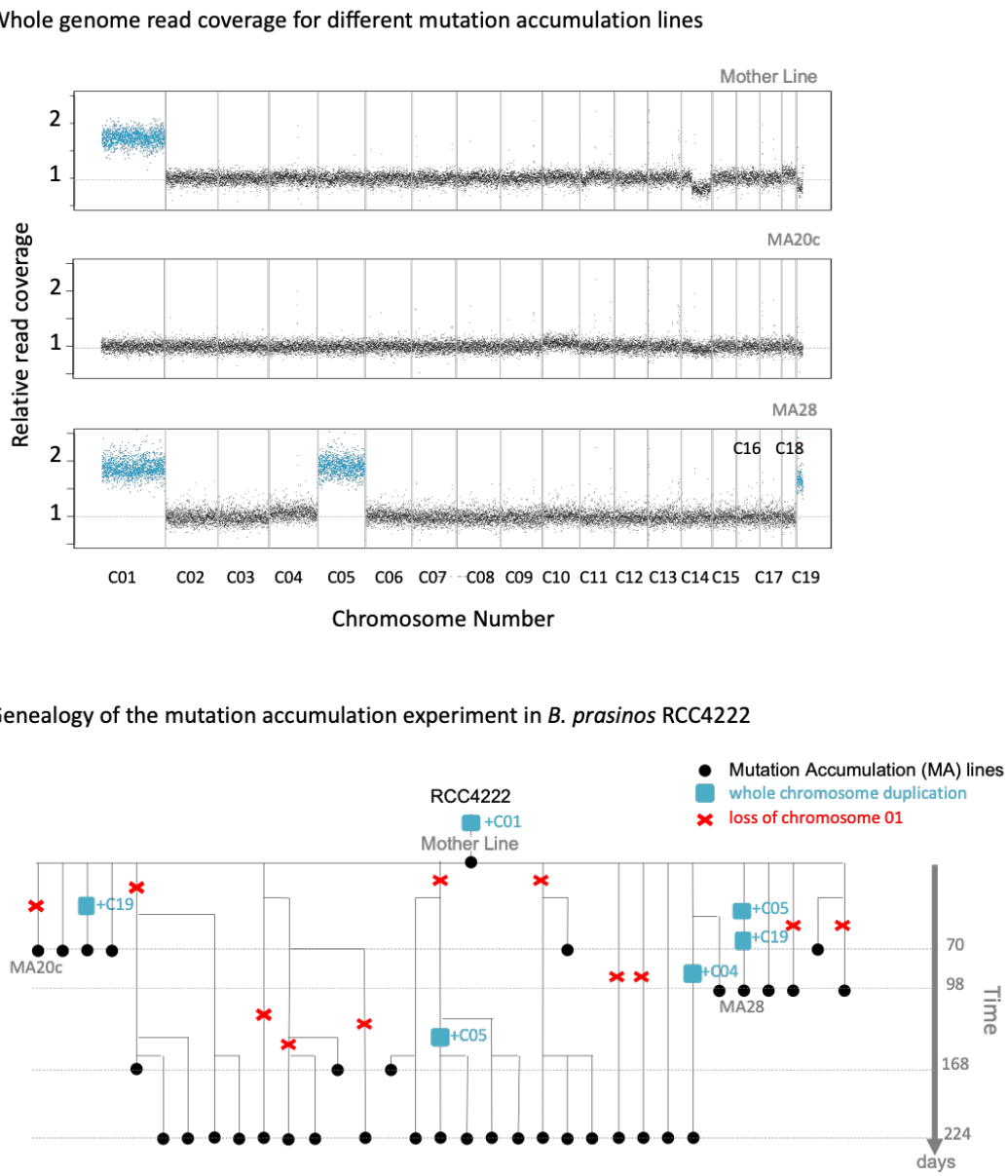

Figure 1. A. Normalised raw genomic coverage of the mother line, Bp20c and Bp28 mutation accumulation lines of *Bathycoccus prasinos*. In blue are the chromosomes in double copies. Raw coverage of all lines from all species are provided in Figures S1 to S16. **B.** Pedigree of the mutation accumulation lines from the *Bathycoccus prasinos* experiment. The chromosome 01 is duplicated in the T0 line (named mother line) of the experiment. This duplication is then lost several times, and 5 independent chromosome other duplications of chromosome 4, 5 and 19 occurred.

Compensation of Whole Chromosome Duplication (WCD)

To explore the effect of the WCD on transcript production, we estimated the level of transcription of a control line (*B. prasinos* RCC4222) and one MA line of *B. prasinos* by recovering one 4-year-old cryopreserved mutation accumulation lines. This

cryopreserved culture originated from the MA line Bp37, and the recovered culture was hereafter named Bp37B. Chromosome copy number of Bp37B and the control line were estimated by whole genome resequencing. This confirmed the two copies of chromosome 4 in Bp37B as in the original Bp37 (Figure 2A and S7), but also revealed additional changes: an increase in chromosome 1 copy number (Figure 2A) in Bp37B, as well as a duplicated copy of chromosome 10 in the control line (Figure S18). Transcriptomic analyses of Bp37B and the control line revealed that there are in average twice the number of transcripts for genes located on chromosome 4 (Transcription Rate (TR) average=1.94, median=1.87, estimated in TPM, Figure 2B, raw data provided in Table S3) compared to the control line. There is thus evidence of overexpression of the duplicated genes. As no DC could be observed at the transcriptional level, we investigated whether post-transcriptional processes may temperate this twofold excess of transcripts. This second hypothesis was tested in the same *B. prasinos* MA line by sequencing mRNAs associated with ribosomes (polysomes) in order to compare the translation efficiency (TE) of genes located on duplicated and non-duplicated between lines.

The translation efficiency of each transcript was estimated by the ratio of mRNAs in polysomes versus total RNA between Bp37B and control. We found that the average TE for the genes located on the duplicated chromosome 4 was 0.67 (median=0.50) compared to the genes on this chromosome in the control line (Figure 2 C). Last, we estimated the expected protein production in Bp37B as compared to the control line by going back 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 showed that genes on chromosome 4 have a similar protein production in Bp37B and in the control line, despite a 1.94 higher transcription rate as a consequence of the WCD. Interestingly, the duplications who arose more recently as part of the strain recovery from cryopreserved lines were not compensated neither at the transcription, nor at the translation level (chromosome 1 and chromosome 10, Figure 2).

The observed compensation at the translation level on chromosome 4 suggests that a post-transcriptional mechanism occurs on duplicated transcripts to allow DC. Poly(A) tail is a key feature of many cytoplasmic mRNAs and is known to regulate positively or negatively translation via size variation (Subtelny et al. 2014; Lim et al. 2016; Eichhorn et al. 2016). Here we tested if translation compensation observed for transcripts linked to duplicated chromosome 4 is associated with a variation of poly(A)

tail length. Using 3'RACE experiment, we measured the poly(A) tail length of three duplicated transcripts on chromosome 4 (04g00840, 04g01730 and 04g04360) compared to two transcripts on chromosome 8 (non-duplicated) and chromosome 1 (duplicated but without translation regulation) as control in Bp37B and control lines (Figure 3 and S18). For the 3 transcripts tested, we systematically observed that the poly(A) tail length is significantly reduced in Bp37B line compared to control line for transcripts located on chromosome 4 (Figure 3), whereas no difference is observed for transcripts of non-duplicated chromosome (Figure S18).

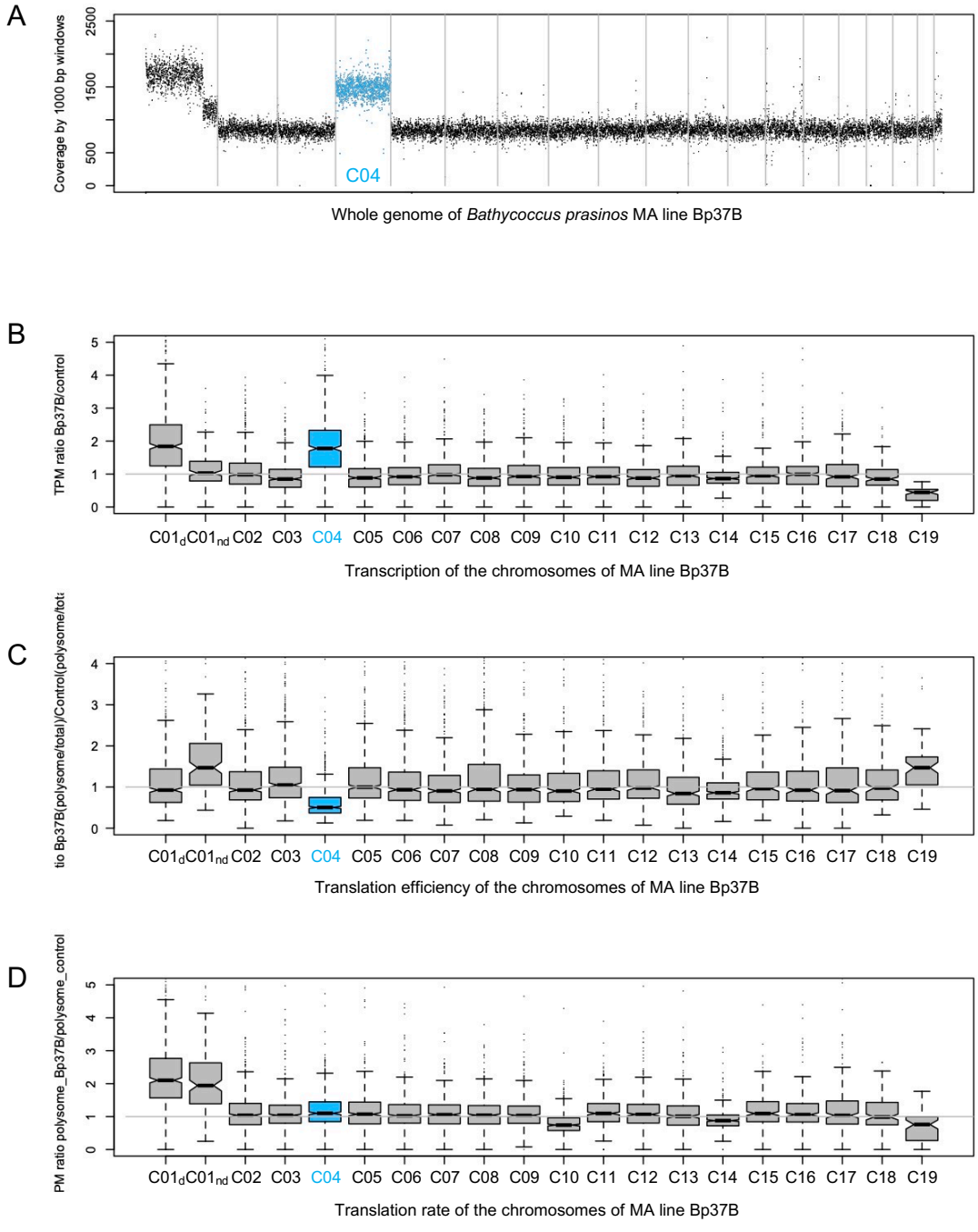

Figure 2. A. Raw coverage by 1000 bp windows along the genome of the MA line *Bathycoccus prasinos* Bp37B (Bp37B is the culture started in 2019 from the MA Bp37 culture cryopreserved in 2016). Grey lines are chromosome separators. $C01_d$: duplicated part of the chromosome 1. $C01_{nd}$: non duplicated part of the chromosome 1. **B.** Distribution of genes expression ratio of Bp37B over control TPM of the 19 chromosomes. **C.** Distribution of genes translation efficiency of the 19 chromosomes. **B** and **C**. The expression and translation efficiency ratio averages are different between chromosomes (ANOVA, p-value < 0.001). The expression is different between chromosome 4 and all other chromosomes and the non-duplicated part of the chromosome 1 (Student test, adjusted p-value for multiple tests <0.0001), excepted with the duplicated part chromosome 1 (Student test, adjusted p-value for multiple tests =0.101); and translation efficiency is different between chromosome 4 (highlighted in blue) and all other chromosomes (Student test, adjusted p-value for multiple tests <0.001). C01 to C19 are the chromosome numbers. **D.** Distribution of genes translation rate ratio of the 19 chromosomes.

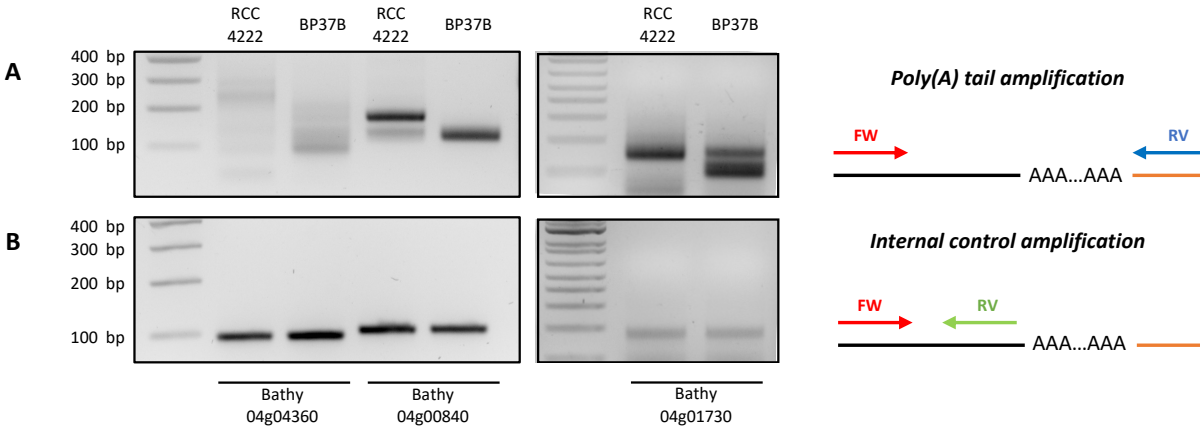

Figure 3. Transcripts from genes located on duplicated chromosome 04 present shorter poly(A) tail. Poly(A) tail measurement was performed using modified 3'RACE. PCR amplification was performed using primers flanking poly(A) tail (A) or primers anchored in 3'UTR just before poly(A) tail as internal control (B). Experiments was performed using total RNA from RCC4222 line (control line) and BP37B line (with duplicated chromosome 4). Illustrations representing PCR amplification are present on the right panel. Black line represents the messenger RNA and the orange line represent the ligated adapter used for reverse transcription and PCR amplification. FW. Forward primer. RV. reverse primer.

DISCUSSION

Mechanism of WCD in unicellular eukaryotes

There are two possible mechanisms at the origin of whole chromosome duplication: i) the additional replication of a chromosome before mitotic cell division, leading to one cell with one chromosome and one cell with two copies; or ii) the unequal segregation of chromosomes during cell division, leading to one cell with two copies and one cell without any copy of the chromosome. The latter scenario seems more likely in extremely small-sized Mamiellophyceae (1µm cell diameter) cells as they contain fewer kinetochore microtubules than chromosomes (Gan et al. 2011), which could lead to a higher error rate in chromosome segregation. However, yeast cells contain approximatively as many kinetochore microtubules as chromosomes (Peterson and Ris 1976) and whole-chromosome duplication and loss estimates are $9.7x10^{-5}$ and $7.0x10^{-6}$ events per diploid genome per generation, respectively (Zhu et al. 2014), which are values close to the rates of whole chromosome duplication reported here. Furthermore, these rates are expected to vary with environmental conditions, as previous mutation accumulation studies in seven different environments reported significant variation in the rates of chromosome gain and loss (Liu and Zhang 2019). Interestingly, the rate of chromosome duplication loss was estimated to be about three orders of magnitude higher than the whole chromosome duplication rate. This suggests that cells with duplicated chromosomes are ephemeral and are unlikely to be maintained for a long time in batch culture.

Fitness effect of WCD and dosage compensation.

The disruption of chromosome numbers is expected to incur a significant fitness cost, in contrast to point mutations whose fitness effects cover the deleterious to advantageous range (Eyre-Walker and Keightley 2007). Using the number of cell divisions per days as a proxy of fitness during the mutation accumulation experiment (Krasovec et al. 2016), we found that only one out of the 15 MA lines with a WGD displayed a significant fitness decrease along the experiment (Bp28b, Pearson correlation, p-value = 0.015, rho = -0.850). A possible cause of aneuploidy tolerance, or at least of the non-detection of fitness decrease in most of MA lines, may be due to a compensating point mutation (Torres et al. 2010): in yeast, point mutations leading to non-synonymous changes in specific genes (particularly the deubiquitinating enzyme involved in proteins degradation) have been found in several disomic lines

with higher growth rates than their parents. However, these mutations did not lead to a fitness increase in all disomic strains, and a few disomic strains with higher growth rates did not harbour any non-synonymous mutations. The previously identified point mutations from each MA line carrying a WCD in this study (Table S2) (Krasovec et al. 2017, 2018a, 2019) are unlikely to impact the aneuploidy tolerance for two reasons. First, some lines such as Mp03, Bp28b, Bp26 or Bp25 did not carry any additional point mutation as compared to MA lines without WCD. Second, the number of mutations per line was limited and, in some cases, appears only in intergenic region or leads to synonymous codon change (*i.e.* Om3 or Mp28).

The more likely process limiting the deleterious effect of aneuploidy is a DC mechanism that prevents the disruption caused by the excess of transcripts. Here in *B. prasinos*, there is evidence for DC at the translation level for the duplicated chromosome 4 only. However, more recent WCD in cryopreserved lines like chromosome 1 led to an excess of transcripts from this chromosome which was not tempered by a lower translation rate. On the contrary, it seems that the part of the chromosome with a lower copy number experienced an increased translation rate (Figure 2C) as if to compensate with the lower transcription level. Another chromosome, chromosome 19, also shows a particular pattern with lower transcription and higher translational rates in the BP37B strain. This chromosome is an outlier chromosome typical of Mamiellophyceae species (Yau et al. 2020; Blanc-Mathieu et al. 2017) with lower GC content than the rest of the genome, likely involved in antiviral immunity. Compensation at the level of gene translation such as observed here in chromosome 4 is relatively unexplored and poorly understood. In *Drosophila*, a compensation of aneuploid genes has been reported and seems to depend on both the genotypic and the developmental context (Zhang and Presgraves 2017). In yeast, a previous study suggested compensation at the post-translational level by protein degradation but failed to detect any compensation during translation (Dephoure et al. 2014).

The role of poly(A) tail length

The role of poly(A) tail length in the 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 (Di Giammartino et al. 2011; Zhang

et al. 2010). 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; Lim et al. 2016; Eichhorn et al. 2016). As an example, during oocyte maturation and early embryonic development, an increase of poly(A) tail length occurs for particular mRNAs resulting in an increase of translation (Subtelny et al. 2014; Lim et al. 2016; Eichhorn et al. 2016). This modulation of poly(A) tail length has also been found to activate some neuronal transcripts (Udagawa et al. 2012). This modulation seems to be specific of a particular developmental context. Recently, it has been also suggested that poly(A) tails can be modulated to balance mRNA levels and adjust TE (Slobodin et al. 2020). In this last case, the CCR4-Not complex shortened the poly(A) tails that reduce the stability of mRNAs. Here, our data suggest that, in a context of chromosome duplication, modulation of poly(A) tail length appears as a key post-transcriptional mechanism necessary for DC at translation level.

CONCLUSION

In conclusion, we estimated an unexpected high rate of whole chromosome duplication in five unicellular photosynthetic eukaryotes. In one species, *B. prasinos*, we identified that a whole duplication chromosome event is associated with a chromosome wide post-transcriptional regulation which might involve the adjustment of poly(A) tail length. These results stress the importance of post-transcriptional regulation mechanisms in DC of aneuploid karyotypes.

MATERIALS AND METHODS

Sequencing data from mutation accumulation experiments

Mutation accumulation (MA) experiments of the 6 species *Picochlorum costavermella* RCC4223, *Ostreococcus tauri* RCC4221, *O. mediterraneus* RCC2590, *Bathycoccus prasinos* RCC1105 (synonym to RCC4222), *Micromonas pusilla* RCC299 and *Phaeodactylum tricornutum* RCC2967 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 following of mutation accumulation lines which evolved from the 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 ($6<Ne<8.5$ (Krasovec et al. 2016)) throughout the experiments. Coupled with whole genome

sequencing of ancestral and MA lines, this experiment enables the direct estimation of the spontaneous mutation rate of a species. Here, MA lines came from a single cell obtained by dilution serving as T_0 culture (named the mother line, ML) and were maintained in 24-wells plates in L1 medium at 20 °C with a 16h-dark 8h-light life cycle. Single cell bottleneck by dilution were done each 14 days to have a low effective population size and limit selection. DNA of initial line (ML 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 with bwa mem v0.7.12 (Li and Durbin 2010). Then, 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.

Dosage compensation analysis

This study was done three to four years after the end of MA experiments, but some MA lines had been cryopreserved which allowed us to re-start a culture of one *B. prasinos* MA line with a duplicated chromosome 4 (Bp37) to investigate dosage compensation. The restarted culture from the cryopreserved Bp37 was renamed Bp37B. As control, we used the reference culture of the strain *B. prasinos* 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 µmol photon m^{-2} s^{-1} white light at 20 ˚C. The karyotype of the defrozen cultures Bp37B and the control RCC4222 were checked by DNAseq resequencing. There was a change in the karyotype of Bp37B as compared to Bp37 and in the karyotype of RCC4222 as compared to the mother line (ML) (Figure S18). As a consequence, the relative transcription rate was corrected by the number of DNA copy in the control RCC4222 line for further analysis below.

For expression analysis, 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. Three extractions were done for the control culture RCC4222 and two for Bp37B. Then for translation efficiency analysis, polysome extraction was performed for Ribo-seq as described previously (Carpentier et al. 2020) with few modifications. Briefly, 600 mL of *B. prasinos* 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. prasinos* with RSEM with standard parameters (Li and Dewey 2011). We obtained the TPM average that we compared between *B. prasinos* RCC4222 and the MA line Bp37B. To be able to compare the expression and translation rate values, TPM were first normalised within each replicate by the gene Bathy11g03260. Then, expression difference between control and Bp37B was estimated as: normalised_TPM_Bp37B/ normalised_TPM_RCC4222. Last, the expression rate obtain for Bp37B were normalised again for each gene by the expression rate average of non-duplicated chromosome. The polysome RNA TPM values were normalised identically, and the translation efficiency rates were calculated as: (polysome_TPM_Bp37B/total_TPM_Bp37B)/(TPM_RCC4222/total_TPM_RCC4222) . 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 present in Table S2.

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 check. RM designed and performed the polysome analyses, poly(A) tail tests and RNAseq preparation. GP coordinated the project. SSB and GP supervised the project and generated the data, MK drafted 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 ANRJCJC-SVSE6-2013-0005 to GP and SSB. 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

All genomic raw reads of ML and mutation accumulation lines are available under the bioprojects PRJNA531882 (*Ostreoccocus tauri*, *O. mediterraneus*, *Micromonas pusilla*, *Bathycoccus prasinos*), PRJNA453760 and PRJNA389600 (*Picochlorum costavermella*), and PRJNA478011 (*Phaeodactylum tricornutum*). Transcriptomic raw reads of the *Bathycoccus prasinos* MA lines 37 (Bp37B) and control line RCC4222 are available under the bioproject PRJNA715163. A summary of all data is provided in Table S5.

SUPPLEMENTARY MATERIAL

Figure S1 to S16. Raw coverage by 1 kb windows of all mutation accumulation lines from *Ostreoccocus tauri*, *O. mediterraneus*, *Bathycoccus prasinos*, *Micromonas pusilla*, *Picochlorum costavermella* and *Phaeodactylum tricornutum*. The horizontal grey lines indicate the chromosome separation, and the duplicated chromosome are in blue. ML is the T0 genome of mutation accumulation experiments. In *M. pusilla*, a duplication of a fraction of chromosome 2 in the mother line (ML) is maintained in all MA lines, suggesting there might be a miss-assembly due to a duplicated region at this location.

Figure S17. Coverage analysis of the control RCC4222.

Figure S18. Transcripts from duplicated chromosome 1 and non-duplicated chromosome 8 present similar poly(A) tail.

Table S1. List of duplicated chromosomes in the MA lines.

Table S2. Point mutations previously identified in the MA lines with whole chromosome duplications.

Table S3. Raw TPM data from expression and translation analysis of Bp37B.

Table S4. Primers used for poly(A) tail length analysis.

Table S5. Summary of all data used in this study with bioproject and biosample accessions.

BIBLIOGRAPHY

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.

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.

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

Dehal P, Boore JL. 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLOS Biol* **3**: e314.

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

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

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

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

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

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol* **183**: 557–564.

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

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

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

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

Ho EKH, Schaack S. 2021. Intraspecific variation in the rates of mutations causing structural variation in Daphnia magna. *Genome Biol Evol* evab241.

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

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

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

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

Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G. 2017. Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes. *Mol Biol Evol* **34**: 1770–1779.

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

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

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

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

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

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

Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev*.

http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116 (Accessed October 1, 2021).

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

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

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

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

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

Ohno S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, Berlin Heidelberg https://www.springer.com/la/book/9783642866616 (Accessed April 8, 2019).

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

Press MO, Hall AN, Morton EA, Queitsch C. 2019. Substitutions Are Boring: Some Arguments about Parallel Mutations and High Mutation Rates. *Trends Genet TIG* **35**: 253–264.

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

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

Scott AD, Stenz NWM, Ingvarsson PK, Baum DA. 2016. Whole genome duplication in coast redwood (Sequoia sempervirens) and its implications for explaining the rarity of polyploidy in conifers. *New Phytol* **211**: 186–193.

Segraves KA. 2017. The effects of genome duplications in a community context. *New Phytol* **215**: 57–69.

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

Slobodin B, Bahat A, Sehrawat 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.

Smith SA, Brown JW, Yang Y, Bruenn R, Drummond CP, Brockington SF, Walker JF, Last N, Douglas NA, Moore MJ. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol* **217**: 836–854.

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.

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.

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.

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.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* **10**: 725–732.

Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res* **24**: 1334–1347.

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.

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, 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.

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. 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.

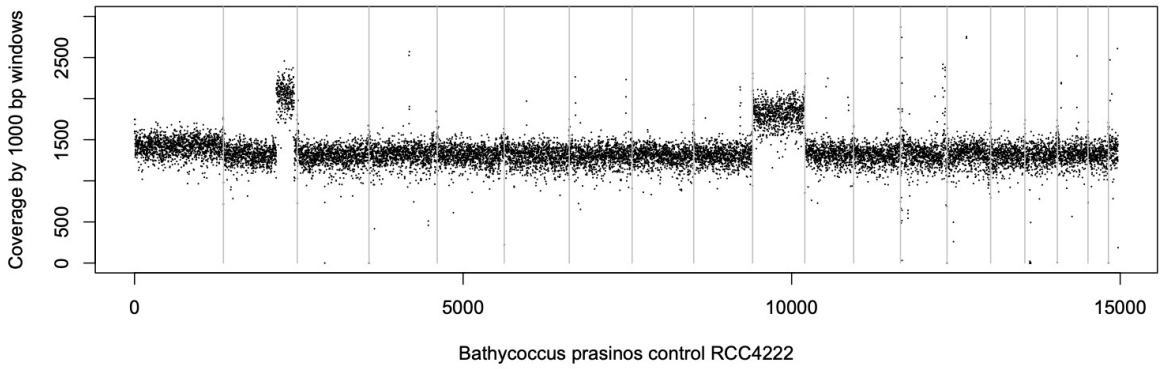

Figure S17. Coverage by 1 kb windows of the control *Bathycoccus prasinos* RCC4222.

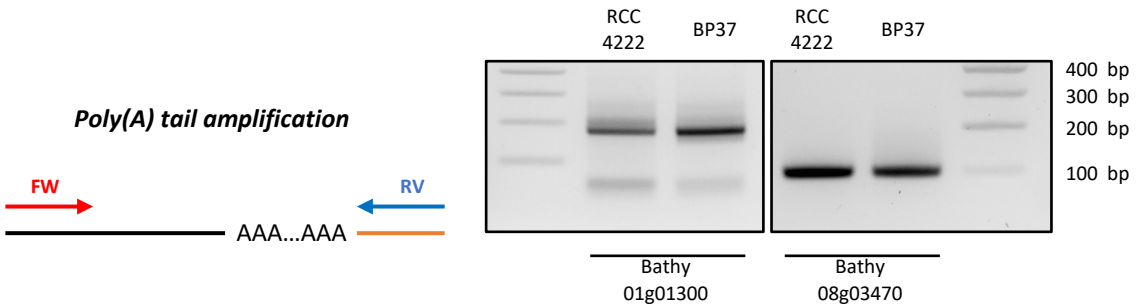

Figure S18. Transcripts from duplicated chromosome 1 and non-duplicated chromosome 8 present similar poly(A) tail. Poly(A) tail measurement was performed using modified 3'RACE. PCR amplification was performed using primers flanking poly(A) tail. Experiments was performed using total RNA from RCC4222 line (control line) and BP37 line (Chromosome 4 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.