Duplicate Gene Expression and Possible Mechanisms of Paralog Retention During Bacterial Genome Expansion

Arkadiy I. Garber[1], Emiko B. Sano[1], Amy L. Gallagher[1], and Scott R. Miller [1,*]

[1]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

*Corresponding author: E-mail: scott.miller@umontana.edu

Accepted: April 22, 2024

Abstract

Gene duplication contributes to the evolution of expression and the origin of new genes, but the relative importance of different patterns of duplicate gene expression and mechanisms of retention remains debated and particularly poorly understood in bacteria. Here, we investigated gene expression patterns for two lab strains of the cyanobacterium *Acaryochloris marina* with expanding genomes that contain about 10-fold more gene duplicates compared with most bacteria. Strikingly, we observed a generally stoichiometric pattern of greater combined duplicate transcript dosage with increased gene copy number, in contrast to the prevalence of expression reduction reported for many eukaryotes. We conclude that increased transcript dosage is likely an important mechanism of initial duplicate retention in these bacteria and may persist over long periods of evolutionary time. However, we also observed that paralog expression can diverge rapidly, including possible functional partitioning, for which different copies were respectively more highly expressed in at least one condition. Divergence may be promoted by the physical separation of most *Acaryochloris* duplicates on different genetic elements. In addition, expression pattern for ancestrally shared duplicates could differ between strains, emphasizing that duplicate expression fate need not be deterministic. We further observed evidence for context-dependent transcript dosage, where the aggregate expression of duplicates was either greater or lower than their single-copy homolog depending on physiological state. Finally, we illustrate how these different expression patterns of duplicated genes impact *Acaryochloris* biology for the innovation of a novel light-harvesting apparatus and for the regulation of *recA* paralogs in response to environmental change.

Key words: gene duplication, positive dosage, neofunctionalization, bacteria.

Significance

Case studies of both lab-evolved and naturally occurring bacteria have highlighted the adaptive potential of an increase in expression as a mechanism for the initial retention of duplicated genes, but its general importance for long-term bacterial genome evolution has remained a puzzle. For cyanobacteria with expanding genomes and a large number of recently duplicated genes, we generally observed a strong positive relationship between duplicate gene copy number and expression level that can persist in some cases for tens of millions of years. We conclude that increased dosage can play an important role for gene duplicate maintenance and for the long-term genome evolution of bacteria. This contrasts with the greater role for expression reduction of duplicates reported for several eukaryotes.

Introduction

Gene duplication is an important mechanism of genome evolution (Andersson and Hughes 2009; Kondrashov 2012) and a major source of new genes (Ohno 1970). Although most duplicates are rapidly lost from genomes (Lynch and Conery 2000), gene duplicates may be retained by several mechanisms. For functionally redundant gene copies, this may involve either a beneficial increase in the number of transcripts (i.e. increased transcript dosage; Ohno 1970) or in the reduction of expression of individual duplicates to recover the ancestral dosage level (i.e. dosage-sharing; Papp et al. 2003; Qian et al. 2010; Birchler and Yang 2022). Alternatively, functional partitioning of duplicates may occur through either the evolution of new functions or expression patterns (neofunctionalization; Ohno 1970) or by the dividing of different ancestral functions between duplicates (subfunctionalization; Force et al. 1999).

The relative importance of these different mechanisms of duplicate retention is still debated. This is particularly the case for duplicates that are not the product of whole-genome duplication, which can lead to transcript dosage imbalances that generally favor duplicate inactivation and loss (Papp et al. 2003). In mammals and yeast, paralogs are often retained through the sharing of ancestral levels of dosage in response to selection to avoid maladaptive stoichiometry following duplication (Qian et al. 2010; Lan and Pritchard 2016). For coregulated tandem gene duplicates, for example, this reduction of gene expression can arise by increased promoter methylation (Rodin and Riggs 2003; Weber et al. 2007; Keller and Yi 2014). While functional partitioning of paralogs is not common in mammals, it is a more likely outcome for duplicates that do not occur in tandem, which can promote the independent evolution of duplicates (Lan and Pritchard 2016).

In bacteria, increased transcript dosage of functionally redundant gene duplicates can contribute to adaptation to stressful environments (Sandegren and Andersson 2009; Kondrashov 2012). Still, most duplicates are quickly purged from bacterial genomes in the absence of selection to maintain them (Romero and Palacios 1997; Reams and Neidle 2003; Reams et al. 2010), despite a high frequency of gene duplication (Anderson and Roth 1977; Haack and Roth 1995; Reams et al. 2010). Consequently, bacterial genomes typically contain far fewer gene duplicates compared with those of eukaryotes (Lynch and Conery 2000, 2003; Hooper and Berg 2003). As a result of this limited sample size, understanding the general importance of increased dosage for the long-term evolution of individual bacterial genomes compared with other mechanisms of duplicate retention has been elusive.

To address this issue, we have taken an integrative approach to investigate gene expression and the potential mechanisms of duplicate retention in the genomes of two strains of the cyanobacterium *Acaryochloris marina* (Miyashita et al. 1996; Wood et al. 2002; Miller et al. 2005), which are notable for both their production of the far-red light absorbing Chlorophyll *d* as primary photosynthetic pigment (Swingley et al. 2008; Miller et al. 2011). The genomes of *A. marina* strains MBIC11017 and CCMEE 5410 are ~35% larger than more basal *A. marina* lineages (8.4 and 8.1 Mb, respectively, vs. 5.8 to 6.1 Mb in *A. marina* strains MU03 and WB-4; Miller et al. 2022), with a similarly greater number of genes (e.g. 8,528 in MBIC11017 vs. 6,366 in MU03); this is a consequence of recent genome expansion due in large part to gene duplication (Ulrich et al. 2021). MBIC11017 and CCMEE 5410 have 796 and 730 duplicate pairs with $d_S < 5$, respectively; by comparison, most bacterial genomes have far fewer duplicates (mean = 102 pairs for a random sample of ~2,400 bacterial genomes; median = 58; supplementary fig. S1, Supplementary Material online).

Duplicated genes in *A. marina* tend to be comparatively recent: frequency distributions based on synonymous nucleotide divergence (d_S) are skewed toward the youngest age classes, with a majority of duplicate pairs having $d_S < 1$ (supplementary fig. S2, Supplementary Material online; Miller et al. 2011). This skewed distribution more closely resembles what has been observed for duplicates in eukaryote genomes (Lynch and Conery 2000, 2003), rather than the uniform distribution of comparatively fewer duplicates reported for other bacteria (Hooper and Berg 2003). *Acaryochloris marina* genomes consist of a chromosome and a variable number of extrachromosomal plasmids (e.g. the MBIC11017 genome has nine ranging in size from ~2 to 375 kb; Swingley et al. 2008). These are low-copy number plasmids for which replication is tightly regulated during the bacterial cell cycle to maintain a characteristic copy number (~1:1 equivalency with the chromosome for most *A. marina* plasmids), and faithful plasmid segregation into daughter cells during cell division relies on specific partition mechanisms (Scott 1984). Most *A. marina* paralogs reside on different genetic elements, either on different plasmids or on the chromosome and a plasmid, respectively (Miller et al. 2011). In addition, except for a minority of the most recent duplicates, paralogs are experiencing strong purifying selection against protein change (Miller et al. 2011). Although the molecular mechanism(s) that underlie the gene duplication process in *A. marina* is unknown, the large number of transposable elements (specifically, insertion sequence [IS] elements) in these genomes suggests that transposition may be involved (Miller et al. 2021). *Acaryochloris* genomes and transcriptomes therefore provide the power to resolve the respective contributions of different mechanisms to duplicate retention in bacteria, as well as whether these roles tend to change over time as paralogs age.

Results and Discussion

Increased Transcript Dosage More Common Than Expression Reduction Following Gene Duplication

The genomes of *A. marina* strains MBIC11017 and CCMEE 5410 vary in copy number for many genes due to both the differential retention of ancestral duplicates and their idiosyncratic histories of duplication events following divergence (Miller et al. 2011). While the *A. marina* MBIC11017 genome is closed, the *A. marina* CCMEE 5410 genome is a high-quality draft assembly of 23 contigs that appears to be complete with respect to gene content (Ulrich et al. 2021); however, there is still the potential for missed duplicates at contig breaks. To investigate the expression of duplicates in the respective genomes, we used RNAseq data collected for both strains in three physiological states: exponential growth, starvation, and recovery (Gallagher and Miller 2018). Depending on strain and condition, recent duplicates ($d_S < 2$) accounted for ~5% to 17% of protein-coding gene transcripts (supplementary table S1, Supplementary Material online). To account for ambiguously mapping reads (reads that map with identical matching scores on closely related paralogs), we developed a custom read-mapping pipeline (see Materials and Methods).

We observed a strongly positive, generally stoichiometric relationship (i.e. slope of ~1) between the ratio of gene copy number between strains and its expression level in a given genome (Fig. 1a; slope = 1.03, 95% CI = (0.94, 1.13); $R^2 = 0.27$; $P < 0.0001$; $N = 1{,}194$ for duplicate genes with $d_S < 5$). For the most common gene copy ratio class (a duplicate pair in one genome and a single copy in the other), mean combined expression of duplicates closely matched the 2:1 ratio predicted for a doubling of expression (Fig. 1b). We conclude that increased transcript dosage of gene duplicates is more common in *A. marina* than has been observed in some eukaryotes, for which expression reduction is a more likely outcome (Qian et al. 2010; Lan and Pritchard 2016).

Still, there was great variation in the expression response of individual duplicate pairs (Fig. 1a and b), and duplicates with different expression fates can be found within blocks of functionally related genes. For example, iron acquisition gene duplicates and novel gene content that are physically clustered on plasmid pREB1 of *A. marina* MBIC11017 (supplementary fig. S3, Supplementary Material online) are associated with faster iron uptake and growth under conditions of low iron availability (Gallagher and Miller 2018). The present analysis revealed cases of both increased transcript dosage and expression reduction among duplicated iron transporter and siderophore genes (supplementary fig. S3, Supplementary Material online). Moreover, recently duplicated *feoAB* paralogs ($d_S = 0.11$), involved in the transport of ferrous iron, exhibited increased transcript dosage under some conditions but expression reduction following iron addition, compared with single-copy expression in strain CCMEE 5410 (supplementary fig. S3, Supplementary Material online).

This result emphasizes the potential context-dependence of dosage benefits in different physiological states (or tissues) following duplication. Few *A. marina* duplicates are in tandem and most are on different genetic elements (~3%; Miller et al. 2011); this physical separation may promote the rapid evolution of such differential regulation, compared with tandem duplicates that physically share *cis*-regulatory machinery.

Variation in expression responses among duplicate pairs may in part reflect duplicate age. Because *A. marina* duplicates are born with identical flanking DNA (or nearly so) to the parental copy (Miller et al. 2011; unpublished data), we may expect recent paralogs to be more likely to exhibit similar expression levels. This was indeed the case, particularly for duplicates with $d_S < 0.2$ in MBIC11017 and $d_S < 0.1$ in CCMEE 5410 (Fig. 2a). More recent duplicates also tended to be more lowly expressed (supplementary fig. S4, Supplementary Material online). This observed trend of lower expression of recent duplicates may generally reflect reduced selection for removal from the genome due to a low metabolic burden; however, in some cases, it also could be a selectively favored mechanism for overcoming the potentially deleterious stochastic fluctuations in the cellular levels of lowly expressed gene products by increasing average expression (Bar-Even et al. 2006). We also observed that equal expression of duplicates can persist over long periods of time (Fig. 2a), as has also been observed in other organisms (Lan and Pritchard 2016).

Asymmetric expression of duplicates (i.e. for which one copy was the major expressed copy in at least two conditions and was never the minor copy) could evolve rapidly but was generally more common for more divergent paralogs (Fig. 2a). Increased transcript dosage could also involve the asymmetric expression of duplicates (16 duplicate pairs with $d_S < 2$ in MBIC11017 and 14 pairs in CCMEE 5410), indicating that the regulation of expression resulting in a dosage increase could be more complicated than simply the equal expression of duplicates. In some cases (particularly among younger duplicates for CCMEE 5410), the minor copies of a duplicate pair were not expressed and are likely destined to be purged from the genome (supplementary fig. S5, Supplementary Material online); however, similar to what has been observed for humans (Lan and Pritchard 2016), on average the minor copy makes a meaningful contribution to expression in both strains (57% of major copy expression for MBIC11017% vs. 27% for CCMEE 5410 for duplicates with $d_S < 0.5$). Furthermore, possible functional partitioning of duplicates (i.e. different copies were the major copy in at least one condition each) in response to these conditions was rare

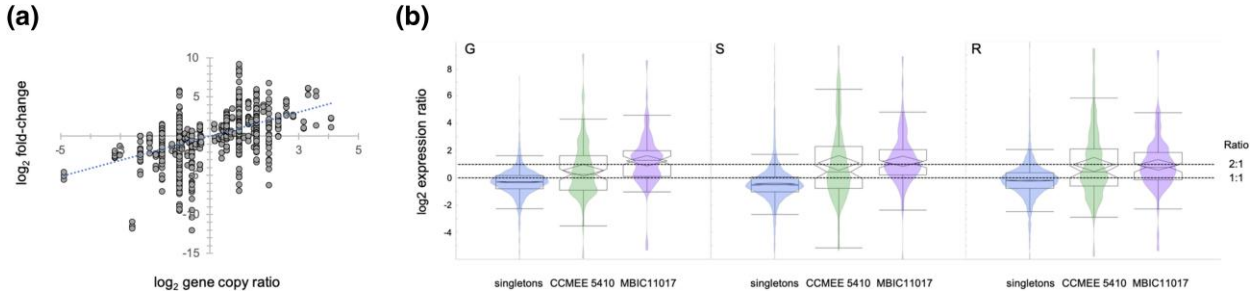

Fig. 1.—Increased transcript dosage predominates for *A. marina* gene duplicates ($d_S < 5$ in both panels). a) Stoichiometric relationship between gene copy number in *A. marina* genomes and expression pooled for three experimental conditions (ratios are CCMEE 5410/MBIC11017). b) Gene expression ratios for singletons in both genomes (ratio is CCMEE 5410/MBIC11017), duplicate pairs in CCMEE 5410 and duplicate pairs in MBIC11017, respectively, during growth (G), starvation (S), and recovery (R).

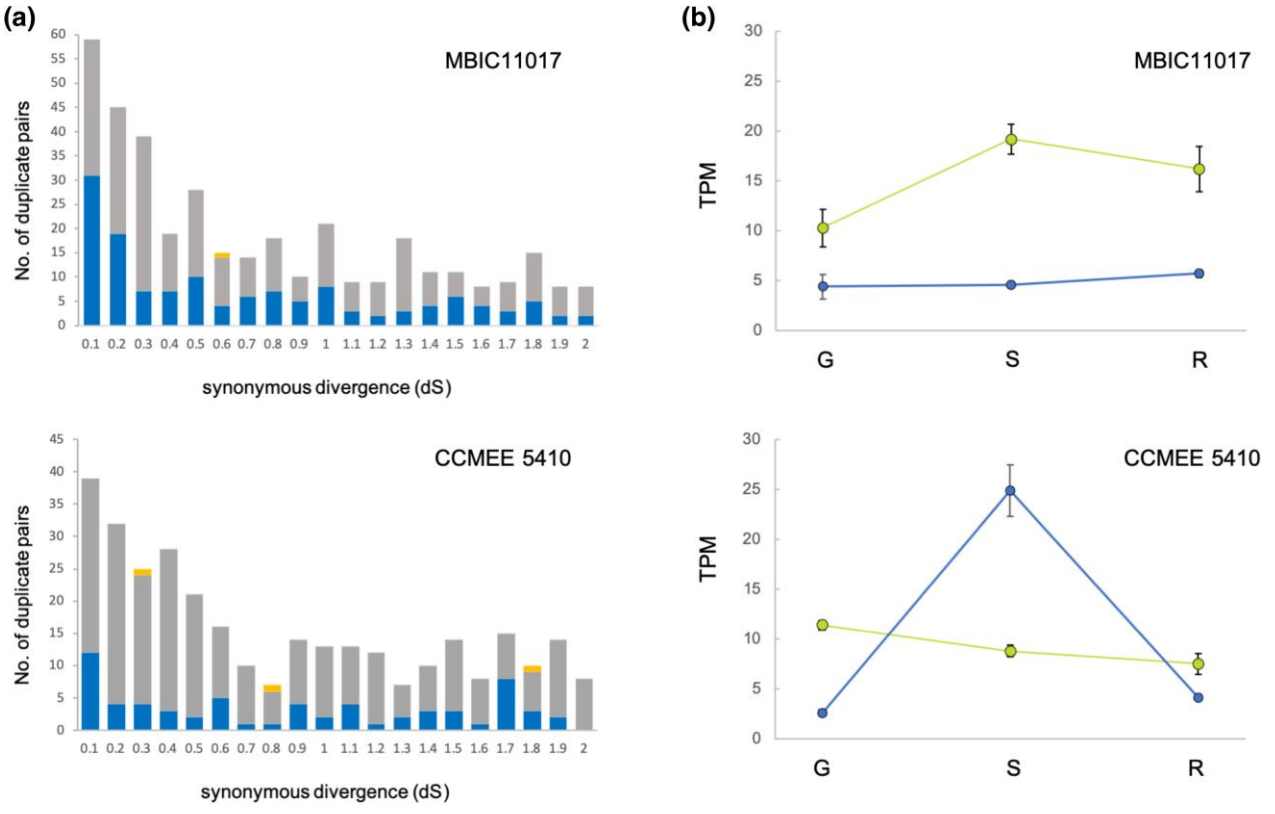

Fig. 2.—Expression responses of duplicate pairs. a) Number of duplicate pairs assigned to different expression categories as a function of duplicate age estimated as synonymous divergence: No difference in expression (blue), asymmetric expression (gray) or possible functional partitioning (yellow). b) The fates of ancestrally shared duplicates are not necessarily deterministic: PEP synthase expression is asymmetrically expressed in MBIC11017 but functionally partitioned in CCMEE 5410 during growth (G), starvation (S) and recovery (R). Parental chromosomal copy (green); daughter plasmid copy (blue). The respective pairs have experienced similar selective constraints: d_N/d_S of 0.054 (CCMEE 5410) and 0.060 (MBIC11017).

over this time scale of divergence (Fig. 2a). Physical separation of duplicates on different genetic elements, therefore, did not appear to make neofunctionalization or subfunctionalization an intrinsically more likely outcome, as has been proposed for mammals (Lan and Pritchard 2016).

Although this overall pattern is broadly similar for both strains, at the level of the individual duplicate pair, we observed that the expression fates of ancestrally shared duplicates are not necessarily deterministic. Excluding transposable elements, there are 70 ancestrally shared duplicate pairs with $d_S < 5$ that have been retained by both

strains. Although most exhibit similar expression patterns between strains, there are several ($N = 7$) examples of differential regulation. For example, in CCMEE 5410, paralogs of the gluconeogenesis enzyme phosphoenolpyruvate synthase are potentially functionally partitioned, with parental copy transcription predominating during growth and recovery and the daughter plasmid copy more highly expressed during starvation; in MBIC11017, by contrast, the parental copy is the major copy under all conditions (Fig. 2b). The other six cases involve asymmetric expression in one strain and equal expression in the other. Further, for duplicates that are asymmetrically expressed in both strains, in five cases the identity of the major copy (e.g. chromosomal parental copy vs. plasmid daughter copy) differed between strains (supplementary table S2, Supplementary Material online).

Below, we consider two cases of how the evolution of duplicate gene expression impacts *Acaryochloris* biology through its contributions to the origin of a new trait and to organismal response to environmental change, respectively.

Evolution of a Novel *A. marina* Light-harvesting Apparatus

Acaryochloris marina MBIC11017 plasmid pREB3 possesses *cpc* genes required to synthesize and assemble phycocyanin (PC), an accessory light-harvesting phycobiliprotein (Swingley et al. 2008). These genes were recently acquired by horizontal transfer, and many were subsequently duplicated (Fig. 3a; Ulrich et al. 2021). In addition, pREB3 has a 3 to 4× higher copy number compared with the chromosome and other plasmids of the *A. marina* MBIC11017 genome (Fig. 3b), indicative of more frequent replication during the bacterial cell cycle. We made a similar observation for both copy number and expression of plasmid p6 in CCMEE 5410 (supplementary fig. S6, Supplementary Material online). An increase in plasmid cellular copy number therefore represents an alternative mechanism of increasing gene copy dosage and expression in bacteria, along with gene duplication.

PC is composed of heterodimers of α and β peptides (encoded by *cpcAB*) that aggregate to form a rod of hexamers (i.e. a trimer of heterodimers) in association with linker proteins CpcC and CpcD (MacColl 1998). *Acaryochloris marina* MBIC11017 produces a novel four-hexamer PC rod (Chen et al. 2009; Bar-Zvi et al. 2018; Liu et al. 2019) that efficiently transfers energy to photosystem II (Hu et al. 1999) and is anchored to the thylakoid membrane or the photosynthetic reaction center itself by CpcL via a C-terminal hydrophobic segment (Watanabe et al. 2014). *Acaryochloris marina* acquired two divergent copies each of *cpcA* and *cpcB*, all subsequently duplicated (Ulrich et al. 2021); divergent CpcA and CpcB paralogs can co-occur in a rod, and this structural heterogeneity may be responsible for its red-shifted fluorescence emission that facilitates energy transfer to Chl *d* (Bar-Zvi et al. 2018).

Several PC genes exhibited asymmetric expression: *cpcB*1, *cpcB*2, and *cpcD* duplicate pair copies were regulated similarly (Fig. 3c; supplementary fig. S7, Supplementary Material online), with respective major copies located in close physical proximity (Fig. 3a). As expected, these genes exhibited peak expression during growth. In addition, *ycf27* paralogs C0086 and C0101 exhibited similar expression levels during growth, but C0086 is strongly induced during starvation and recovery (Fig. 3d). Ycf27 proteins are OmpR-family DNA binding response regulators; in *Synechocystis* PCC 6803, one of the functions of Ycf27 homologs RpaA and RpaB is to regulate the coupling and relative energy transfer between phycobiliproteins and the two photosystems (Ashby and Mullineaux 1999). The *ycf27* paralogs on pREB3 belong to a gene family including *rpaB* (supplementary fig. S8, Supplementary Material online), and C0101 appears to be a recombinant between a chromosomal copy and a plasmid copy following duplication.

Finally, we observed possible functional partitioning for paralogs of *cpcL*, two of which (C0092 and C0102) are respectively co-transcribed with *ycf27* duplicates C0086 and C0101. While two of the copies exhibited similar expression and were responsible for the majority of transcripts during growth, the third (C0092) was the majority copy during starvation and recovery (Fig. 3e). While the latter copy more closely resembles other cyanobacterial CpcL proteins in both length and hydropathy, the former are distinguished by a hydrophilic, serine-rich insertion of more than 30 amino acids between the linker domain and the C-terminal hydrophobic tail (supplementary fig. S9, Supplementary Material online). Linker proteins impact both the structure and spectral properties of the light-harvesting apparatus (David et al. 2011); consequently, this shift in expression in response to changes in physiological state potentially alters the nature of the interaction between PC rods and the photosynthetic apparatus. Rods are physically attached to PSII in growing cells of *A. marina* MBIC11017 (Hu et al. 1999), whereas they preferentially associate with PSI in other CpcL-producing cyanobacteria (Kondo et al. 2005, 2007; Watanabe et al. 2014). Therefore, one possibility is that the production of different CpcL proteins influences the tendency of PC to associate with different photosystems. Future work will seek to identify whether the observed divergence in expression of these *cpcL* genes, together with co-transcribed *rpaB* paralogs, impacts the distribution of energy transfer from PC to the different photosystems in different physiological states.

Expression Divergence of *recA* Duplicates

The bacterial recombinase RecA is a multifunctional protein involved in homologous recombination, DNA damage

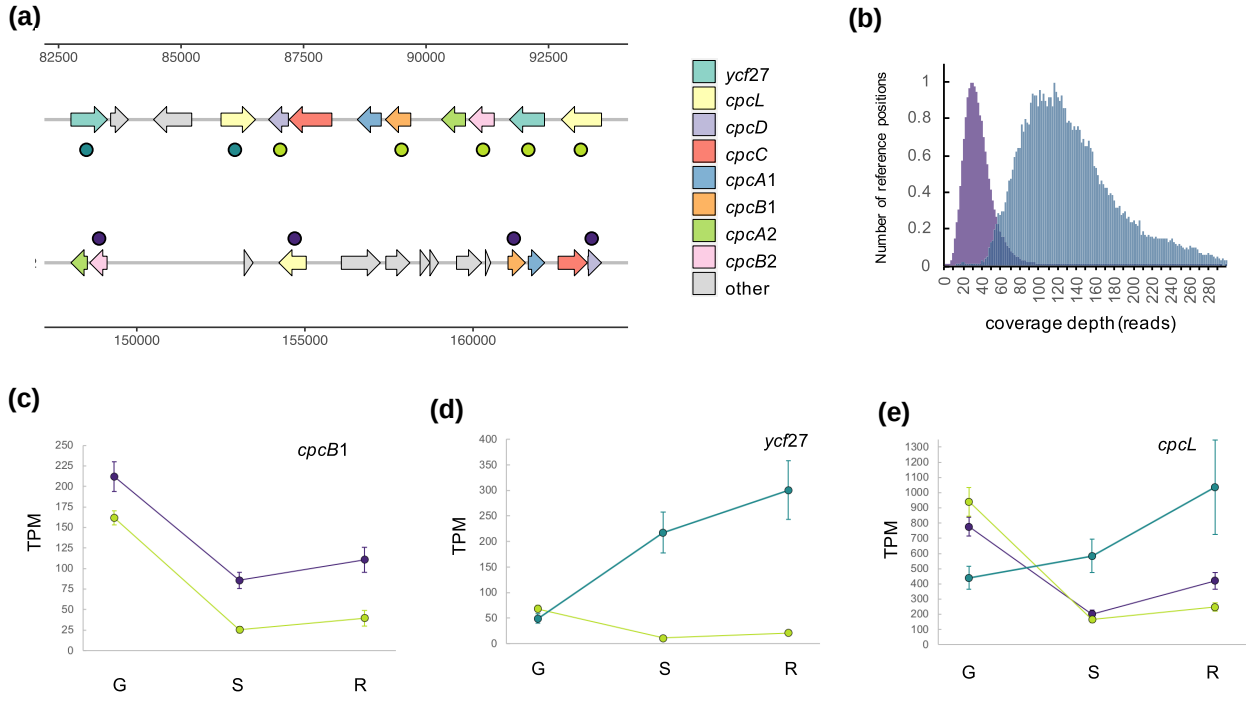

Fig. 3.—Asymmetric gene expression and functional partitioning of duplicates for the novel *A. marina* MBIC11017 phycobilisome. a) Gene maps of *A. marina* MBIC11017 plasmid pREB3 regions containing duplicated genes involved in phycobiliprotein synthesis and its regulation. Filled circles next to selected genes are color-coded to indicate expression values in panels c-e and supplementary fig. S7, Supplementary Material online. b) Distribution of coverage depth for Illumina reads that uniquely map to the strain MBIC11017 reference chromosome (purple) or plasmid pREB3 (blue), respectively. Gene expression (uniquely mapped reads) during growth (G), starvation (S) and recovery (R) for duplicated copies of c) *cpcB*1, d) *ycf*27, and e) *cpcL*.

repair, activation of error-prone DNA polymerase activity, and the regulation of gene expression through its coprotease activity (Miller and Kokjohn 1990). Members of *Acaryochloris* are extraordinary for their number of paralogs of this archetypal "single-copy" gene (Swingley et al. 2008; Miller et al. 2011). Evolution of these proteins has been marked by bursts of positively selected amino acid substitutions (Miller et al. 2011), which suggests that some copies may have diverged in function. Some of these predate the split between *A. marina* and sister taxon *A. thomasi* RCC1774, which does not produce Chl *d*; by contrast, other, plasmid-borne duplicates are more recent and idiosyncratic to individual *A. marina* strains (Fig. 4a).

We first addressed whether *recA* paralogs have retained recombinase activity. *recA* deletion mutants of *E. coli* exhibit a growth rate defect and chromosomal loss (Capaldo et al. 1974; Skarstad and Boye 1993), which stems from the loss of recombinase activity required to repair stalled or collapsed replication forks that can arise during DNA replication (Cox et al. 2000). We introduced four MBIC11017 *recA* genes with CCMEE 5410 orthologs (the three chromosomal copies and plasmid copy B0414; Fig. 4a) into an *E. coli* strain with a *recA* deletion via a plasmid carrying a rhamnose-inducible promoter. In the presence of rhamnose, these either partially or fully complemented the *recA* deletion (supplementary fig. S10, Supplementary Material online), indicating that these paralogs have recombinase activity.

Next, to investigate whether these *recA* paralogs have diverged in expression, we first examined their transcription patterns in our RNAseq data set. Comparing expression for starvation and recovery conditions with that of growing cells, we observed that orthologs of most copies were down-regulated in MBIC11017 during starvation and recovery and, conversely, up-regulated in CCMEE 5410 (Fig. 4b). By contrast, the basal copies in the phylogeny (Fig. 4a; orthologs *recA* 3550 in MBIC11017 and *recA* 4441 in CCMEE 5410) were more similarly expressed (Fig. 4b). The resulting differences between strains in the relative transcript abundance of paralogs may indicate divergence in paralog function and/or subtle differences in physiological state between strains.

Our analyses of publicly available RNAseq data for strain MBIC11017 (Hernández-Prieto et al. 2016, 2018) corroborate expression divergence of the basal copy and the other paralogs. With the exception of *recA* 3550, copies were strongly up-regulated by hypoxia (Fig. 4c), which induces DNA damage, replication arrest and *recA* expression in mycobacteria (Gill et al. 2009; Gorna et al. 2010; Prasad et al. 2019); in addition, *recA* 3550 was uniquely down-

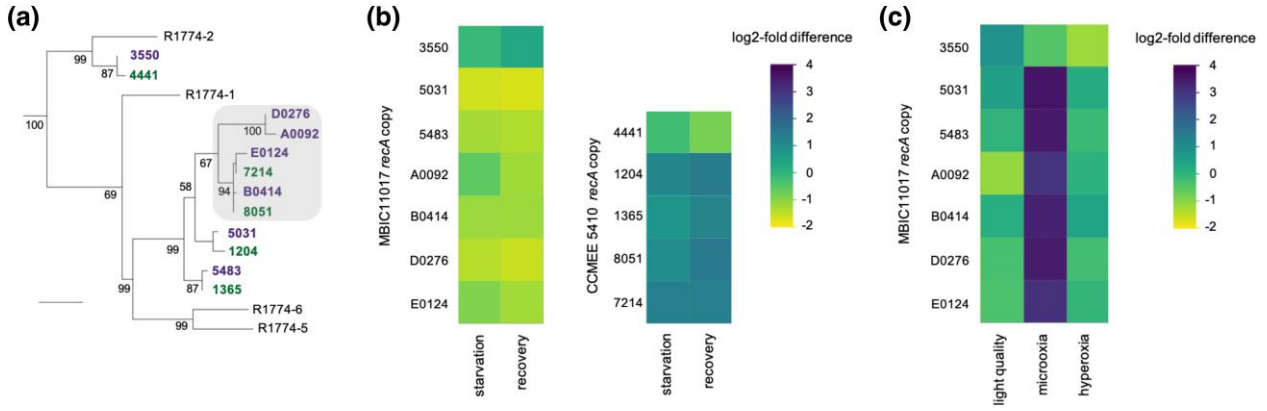

Fig. 4.—Differential expression of *recA* paralogs. a) RecA amino acid phylogeny reconstructed with IQtree by maximum likelihood using a LG + R3 model of sequence evolution. MBIC11017 and CCMEE 5410 sequences are in purple and green, respectively, and plasmid copies are in the gray box. R-1774 sequences are from the *A. thomasi* RCC1774 genome. Ultrafast bootstrap support greater than 50% for 1,000 bootstrap replicates is indicated at bifurcations. The tree was outgroup-rooted with sequences for *Cyanothece* PCC 7425, *Geitlerinema* PCC 7407 and *Microcoleus* FACHB-672. Scale bar is 0.05 amino acid substitutions per site. CCMEE 5410 copy 7214 is interrupted by a IS256 family transposase; although unresolved in the phylogeny, it appears to be recent duplicate (following the split with MBIC11017, rather than an ancestrally shared ortholog of MBIC11017 copy E0124) based on its gene order conservation with CCMEE 5410 copy 8051 and MBIC11017 copy B0414. b) Differential expression heat map of *A. marina* MBIC11017 and CCMEE 5410 *recA* paralogs for starvation and recovery conditions compared with growing cells; c) Differential expression heat map of *A. marina* MBIC11017 *recA* paralogs for differences in light quality and oxygen availability. Light quality is the difference in expression in far-red light versus white light (data from Hernández-Prieto et al. 2018); microoxia and hyperoxia are compared with normoxia (data from Hernández-Prieto et al. 2016).

regulated under hyperoxia, a condition expected to produce high levels of reactive oxygen species (ROS). Expression profiles of the other *recA* paralogs were more similar overall (Fig. 4b and c), but *recA* A0092 alone exhibited decreased expression during a shift in light quality from white light (absorbed primarily by PC) to far-red light, which is absorbed directly by Chl *d*.

Finally, we conducted qPCR assays for representative MBIC11017 *recA* paralogs in cells exposed to either UV radiation or hydrogen peroxide. In *E. coli*, induction of *recA* expression is a signature of the SOS response to DNA damage (Casaregola et al. 1982); however, *recA* is downregulated by UV radiation in the cyanobacteria that have been studied (Domain et al. 2004; Kolowrat et al. 2010). However, we found that *recA* B04014 was the only one of the tested copies with reduced expression in response to UV radiation (supplementary fig. S11, Supplementary Material online). We predicted that the ROS hydrogen peroxide would elicit a specific decline in expression of *recA* 3550, as observed for hyperoxia (Fig. 4c), which was the case (supplementary fig. S11, Supplementary Material online).

Together, these results for several environmental conditions show that divergence of gene expression among both ancient (e.g. basal vs. other copies) and recent duplicates (A0092 and D0276) contribute to *recA* expression patterns in *A. marina*. Consequently, although all *recA* copies are constitutively expressed, the stoichiometry of different *recA* transcripts is highly dynamic in response to environmental change, as expected during potential specialization on different sub-functions. Future studies will aim to use in vitro assays with purified *A. marina* RecA proteins to better resolve the nature of possible functional divergence among paralogs.

Concluding Remarks

Case studies of both lab-evolved and naturally occurring bacteria have highlighted the adaptive potential of increased transcript dosage following gene duplication, but its general importance compared with other mechanisms of duplicate retention has remained unclear. For two strains of *Acaryochloris* with high loads of recent duplicates, we showed that increased duplicate transcript dosage is more prevalent than what has been observed in examined eukaryotic genomes, for which expression reduction appears to be the primary mechanism of initial duplicate retention. Many of these duplicates are ultimately purged from the genome (Miller et al. 2011); this could be for several reasons, including the transcript dosage imbalances that duplication can create, or changes in whether selection favors maintenance of more than a single copy of a gene. However, increased transcript dosage can persist for long periods of time in *A. marina*. Mean d_S of orthologs between the two strains is ~0.3; using Bayesian relaxed clock analyses, Sánchez-Baracaldo (2015) estimated this split to have occurred ~46 MYA. Therefore, most duplicates in our data set have persisted for millions of years without deletion. By contrast, deletion rates for gene duplicates in bacteria have been estimated to be high in the absence of

selection to maintain them (Reams and Neidle 2003; Reams et al. 2010). In addition, even the most recent *A. marina* paralogs experience strong purifying selection (Miller et al. 2011). We consequently propose that increased transcript dosage may be an important mechanism of initial duplicate retention in these bacteria. Nonetheless, expression divergence of paralogs can also evolve quickly, including the emergence of possible functional partitioning through changes in the regulation of expression. Although rare, the latter can play an important role in *Acaryochloris* diversification, as illustrated by both the regulation of genes involved in the production of the light-harvesting phycobiliprotein phycocyanin and the differential expression of *recA* paralogs.

Materials and Methods

Data Acquisition

To measure expression of gene duplicates, we used RNAseq data collected for *A. marina* strains MBIC11017 (GCA_000018105.1) and CCMEE 5410 (GCA_000238775.3) under three different culture conditions: exponential growth, starvation, and recovery phases (Gallagher and Miller 2018). Data were downloaded from NCBI (BioProject: PRJNA681975) using the SRA-Toolkit (https://hpc.nih.gov/apps/sratoolkit). Genomes were downloaded from GenBank using the Bit software toolkit (https://github.com/teambit).

Identification of Duplications and Grouping of Paralogs

We used *ParaHunter* (Miller et al. 2022) to identify gene duplicates. All genes with $>50\%$ amino acid identity and $>50\%$ sequence length overlap were grouped together in clusters; most clusters were composed of two genes. Multiple sequence alignments were generated using Muscle (Edgar 2004), and, from these, codon alignments were made using PAL2NAL (Suyama et al. 2006). We then used CODEML from the PAML software package (Yang 2007) to estimate d_S and d_N values. Codon alignments yielding d_S estimates greater than 5 were excluded from further analyses.

Estimation and Comparison of Expression Levels

To account for RNAseq reads that potentially map well to more than one gene copy, we used a combination of Bowtie2 and BLASTN to discriminate between uniquely mapping reads and those that map equally well to more than one gene copy. Specifically, we used Bowtie2 to identify the total amount of reads mapping to each cluster of paralogous genes. Next, we used BLASTN to identify reads that match with 100% sequence identity to more than one gene in each cluster of paralogs; these ambiguously mapping reads were excluded from analysis. Expression values for each specific gene/paralog in a cluster (for paralog-vs.-paralog comparisons) were not calculated in paralog clusters where more than 10% of the reads were removed due to ambiguity. Ambiguously mapping reads were included in the estimation of bulk-cluster gene expression levels (i.e. the total amount of transcripts generated from all paralogs of a specific gene) for estimating transcript dosage of paralogs.

We quantified gene expression as Transcripts Per Million (TPM) to normalize for gene length and the sequencing depth of each RNA-sequenced library. We next used ANOVA to identify differential expression between paralog pairs and to detect interactions between paralog pairs across the three different experimental conditions. Paralogs with a significant interaction showed differences in expression across conditions (e.g. functional partitioning). To minimize data noise in paralog-versus-paralog comparisons, we required a minimum TPM of 10 in all three conditions. To minimize false positives during ANOVA testing, we required at least one read mapping to each of the paralogs in at least 2 of the 5 experimental replicates.

A. marina recA experiments and phylogenetics

For cloning of *recA* genes, we added a multiple cloning site to Addgene plasmid 40779 (resulting in plasmid pRHA) in order to have *recA* genes under the control of a rhamnose promoter. The multiple cloning site was introduced using gBlocks from New England Biolabs. We PCR-amplified four *recA* copies from *A. marina* strain MBIC11017 (AM1_3550, AM1_5031, AM1_5483, AM1_B0414) as well as the *recA* genes from *E. coli* MG1655 and *Cyanothece* sp. strain PCC 7425. Strains and primers used are listed in supplementary table S3, Supplementary Material online. All genes were cloned into the *Spe*I and *Not*I sites of pRHA that had been introduced with the primers. All clones were sequence verified. We introduce the empty vector (as a control) to the *E. coli* strains that either carry the deletion (NoRecA) or have an intact chromosomal copy of *recA* (ChrRecA). Each vector carrying a cloned *recA* copy was introduced into the *rec*-deletion strain (denoted as 3550, 5031, 5483, B0414, respectively, along with RecAe for *E. coli recA*).

To verify *recA* expression, we performed RT-PCR for each cloned copy of *recA* after rhamnose induction. Cells grown overnight in LB broth with ampicillin and 0.2% glucose were inoculated into fresh LB with ampicillin and 0.2% rhamnose. After 3 h of growth at 37°C with agitation, 1 mL of cells was pelleted and stored at −80°C until further processing. A Qiagen RNeasy kit was used to extract RNA. We generated cDNA using Maxima First Strand cDNA synthesis kit (Thermo Scientific). We used primers designed specific to each *recA* copy for detection of each transcript.

Growth of *E. coli* strains was measured in 96-well plates using a Synergy HT plate reader (BioTek). Each strain was grown overnight in LB with ampicillin (100 μg mL^{-1}) and 0.2% (w/v) glucose. These cultures were used at 5% (v/v) to inoculate wells (to a total volume of 200 μL) with LB with ampicillin and 0.2% rhamnose. These cells were grown at 37˚C for 4.5 h, and optical densities of wells were monitored at 600 nm. Doubling times were estimated for four biological replicates and three independent experiments for the exponential growth phase by: $(T_f\text{-}T_0)$*log2/log(OD_f/OD_0), where T_f and T_0 correspond to the last point at which the cells were growing exponentially (determined by plotting the growth curve on a semi log plot) and the first point at which the cells entered exponential phase, respectively, and OD_f and OD_0 correspond to the OD_{600} reading at the T_f and T_0, respectively. To compare growth rates among strains, we performed *t*-tests with a Benjamini–Hochberg False Discovery Rate-adjusted *P* value of 5% estimated with JMP software version 16.2.0 (SAS Institute Inc., Cary, NC).

Acaryochloris marina MBIC11017 cultures were grown in FeMBG-11 medium (IOBG-11 supplemented with iron (III) monosodium salt) at 30 °C with continuous illumination from cool fluorescent lights at ~20 μmol photons m^{-2} s^{-1} and mild agitation. All cultures were grown to an OD_{750} of ~0.15 to 0.20 in 300 mL and split in half to generate the control and experimental cultures. For the H_2O_2 treatment, we exposed cells to 3 mM H_2O_2 for one hour before harvesting cells. For the UV treatment, the cultures were exposed to 300 J m^{-2} using a BioRad GS Gene Linker, followed by 1 h recovery in the dark (to avoid photoreactivation) before harvesting the cells. For harvesting cells, we filtered cells onto 0.6 μm pore Isopore membrane filters (Millipore), followed by flash freezing. Total RNA was isolated using a Direct-zol RNA mini-prep kit (Zymo Research). We added a bead beating step with the kit's TRI-reagent before proceeding according to the manufacturer's protocol. Each prep was checked for genomic DNA contamination using PCR before proceeding to cDNA synthesis; any prep found to be contaminated was treated with additional DNase, cleanup, and another round of PCR. We generated cDNA using a Maxima First Strand cDNA Synthesis Kit with dsDNase (Thermo Scientific) and 50 ng of RNA. We measured relative expression by qPCR using a Stratagene Mx3000p (Agilent) and DyNAmo Flash SYBR Green qPCR kit (Thermo Scientific).

MxPro QPCR software was used to calculate Ct values. We used the comparative Ct method to estimate relative expression levels (Livak and Schmittgen 2001; Schmittgen and Livak 2008). Each sample was normalized to the average expression of reference genes *petB* and *ilvD*. Normalized expression values for control and treatment samples were then used to estimate relative expression. All statistical analyses were performed using the R statistical environment on the raw ΔΔCt values before log transforming for fold-change calculation.

A RecA amino acid phylogeny was reconstructed with IQtree (Nguyen et al. 2015) by maximum likelihood using a LG + R3 model of sequence evolution selected by BIC with ModelFinder (Kalyaanamoorthy et al. 2017) with 1,000 ultrafast bootstrap replicates. The tree was outgroup-rooted with sequences for *Cyanothece* PCC 7425, *Geitlerinema* PCC 7407, and *Microcoleus* FACHB-672 (NCBI accession numbers B8HR52.1, WP_015170153.1, and WP_190665208.1).

Supplementary Material

Supplementary Material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank John R. Roth for providing us with *E. coli* wildtype and *recA* mutant strains, Nikea Ulrich for reconstructing the *ycf27* gene network and three anonymous reviewers for their comments on an earlier version of the manuscript.

Author Contributions

S.R.M. and A.I.G. designed the study; A.L.G. and E.B.S. collected the data; all authors carried out the analyses and wrote the manuscript.

Funding

This work was supported by award NNA15BB04A from the National Aeronautics and Space Administration to S.R.M.

Data Availability

Data files used in our analyses, in addition to the custom python scripts that we used for analysis, are available in the following GitHub repository: https://github.com/Arkadiy-Garber/Supplement-for-Acaryochloris-Duplication-Paper.

Literature Cited

Anderson RP, Roth JR. Tandem genetic duplications in phage and bacteria. Annu Rev Microbiol. 1977:31(1):473–505. https://doi.org/10.1146/annurev.mi.31.100177.002353.

Andersson DI, Hughes D. Gene amplification and adaptive evolution in bacteria. Annu Rev Genet. 2009:43(1):167–195. https://doi.org/10.1146/annurev-genet-102108-134805.

Ashby MK, Mullineaux CW. Cyanobacterial *ycf27* gene products regulate energy transfer from phycobilisomes to photosystems I and II. FEMS Microbiol Lett. 1999:181(2):253–260. https://doi.org/10.1111/j.1574-6968.1999.tb08852.x.

Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006:38(6):636–643. https://doi.org/10.1038/ng1807.

Bar-Zvi S, Lahav A, Harris D, Niedzwiedzki DM, Blankenship RE, Adir N. Structural heterogeneity leads to functional homogeneity in *A. marina* phycocyanin. Biochim Biophys Acta (BBA). 2018:1859(7):544–553. https://doi.org/10.1016/j.bbabio.2018.04.007.

Birchler JA, Yang H. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell. 2022:34(7):2466–2474. https://doi.org/10.1093/plcell/koac076.

Capaldo FN, Ramsey G, Barbour SD. Analysis of the growth of recombination-deficient strains of *Escherichia coli* K-12. J Bacteriol. 1974:118(1):242–249. https://doi.org/10.1128/jb.118.1.242-249.1974.

Casaregola S, D'Ari R, Huisman O. Quantitative evaluation of *recA* gene expression in *Escherichia coli*. Mol Gen Genet. 1982:185(3):430–439. https://doi.org/10.1007/BF00334135.

Chen M, Floetenmeyer M, Bibby TS. Supramolecular organization of phycobiliproteins in the chlorophyll *d*-containing cyanobacterium *Acaryochloris marina*. FEBS Lett. 2009:583(15):2535–2539. https://doi.org/10.1016/j.febslet.2009.07.012.

Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ, Marians KJ. The importance of repairing stalled replication forks. Nature. 2000:404(6773):37–41. https://doi.org/10.1038/35003501.

David L, Marx A, Adir N. High-resolution crystal structures of trimeric and rod phycocyanin. J Mol Biol. 2011:405(1):201–213. https://doi.org/10.1016/j.jmb.2010.10.036.

Domain F, Houot L, Chauvat F, Cassier-Chauvat C. Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. Mol Microbiol. 2004:53(1):65–80. https://doi.org/10.1111/j.1365-2958.2004.04100.x.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:32(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999:151(4):1531–1545. https://doi.org/10.1093/genetics/151.4.1531.

Gallagher AL, Miller SR. Expression of novel gene content drives adaptation to low iron in the cyanobacterium *Acaryochloris*. Genome Biol Evol. 2018:10(6):1484–1492. https://doi.org/10.1093/gbe/evy099.

Gill WP, Harik NS, Whiddon MR, Liao RP, Mittler JE, Sherman DR. A replication clock for *Mycobacterium tuberculosis*. Nat Med. 2009:15(2):211–214. https://doi.org/10.1038/nm.1915.

Gorna AE, Bowater RP, Dziadek J. DNA repair systems and the pathogenesis of *Mycobacterium tuberculosis*: varying activities at different stages of infection. Clin Sci (Lond). 2010:119(5):187–202. https://doi.org/10.1042/CS20100041.

Haack KR, Roth JR. Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. Genetics. 1995:141(4):1245–1252. https://doi.org/10.1093/genetics/141.4.1245.

Hernández-Prieto MA, Li Y, Postier BL, Blankenship RE, Chen M. Far-red light promotes biofilm formation in the cyanobacterium *Acaryochloris marina*. Environ Microbiol. 2018:20(2):535–545. https://doi.org/10.1111/1462-2920.13961.

Hernández-Prieto MA, Lin Y, Chen M. The complex transcriptional response of *Acaryochloris marina* to different oxygen levels. G3 (Bethesda). 2016:7(2):517–532. https://doi.org/10.1534/g3.116.036855.

Hooper SD, Berg OG. Duplication is more common among laterally transferred genes than among indigenous genes. Genome Biol. 2003:4(8):R48. https://doi.org/10.1186/gb-2003-4-8-r48.

Hu Q, Marquardt J, Iwasaki I, Miyashita H, Kurano N, Mörschel E, Miyachi S. Molecular structure, localization and function of biliproteins in the chlorophyll *a/d* containing oxygenic photosynthetic prokaryote *Acaryochloris marina*. Biochim Biophys Acta (BBA). 1999:1412(3):250–261. https://doi.org/10.1016/S0005-2728(99)00067-5.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Keller TE, Yi SV. DNA methylation and evolution of duplicate genes. Proc Natl Acad Sci U S A. 2014:111(16):5932–5937. https://doi.org/10.1073/pnas.1321420111.

Kolowrat C, Partensky F, Mella-Flores D, Le Corguillé G, Boutte C, Blot N, Ratin M, Ferréol M, Lecomte X, Gourvil P, et al. Ultraviolet stress delays chromosome replication in light/dark synchronized cells of the marine cyanobacterium *Prochlorococcus marinus* PCC9511. BMC Microbiol. 2010:10(1):204. https://doi.org/10.1186/1471-2180-10-204.

Kondo K, Geng XX, Katayama M, Ikeuchi M. Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. Photosynth. Res. 2005:84(1-3):269–273. https://doi.org/10.1007/s11120-004-7762-9.

Kondo K, Ochiai Y, Katayama M, Ikeuchi M. The membrane-associated CpcG2-phycobilisome in *Synechocystis*: a new photosystem I antenna. Plant Physiol. 2007:144(2):1200–1210. https://doi.org/10.1104/pp.107.099267.

Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. R. Soc. B: Biol. Sci. 2012:279(1749):5048–5057. https://doi.org/10.1098/rspb.2012.1108.

Lan X, Pritchard JK. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. Science. 2016:352(6288):1009–1013. https://doi.org/10.1126/science.aad8411.

Liu H, Weisz DA, Zhang MM, Cheng M, Zhang B, Zhang H, Gerstenecker GS, Pakrasi HB, Gross ML, Blankenship RE. Phycobilisomes harbor FNR_L in cyanobacteria. mBio. 2019:10(2):e00669–e00619. https://doi.org/10.1128/mBio.00669-19.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2_T^{-\Delta\Delta C}$ method. Methods. 2001:25(4):402–408. https://doi.org/10.1006/meth.2001.1262.

Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000:290(5494):1151–1155. https://doi.org/10.1126/science.290.5494.1151.

Lynch M, Conery JS. The origins of genome complexity. Science. 2003:302(5649):1401–1404. https://doi.org/10.1126/science.1089370.

MacColl R. Cyanobacterial phycobilisomes. J Struct Biol. 1998:124(2-3):311–334. https://doi.org/10.1006/jsbi.1998.4062.

Miller RV, Kokjohn TA. General microbiology of *recA*: environmental and evolutionary significance. Annu Rev Microbiol. 1990:44(1):365–394. https://doi.org/10.1146/annurev.mi.44.100190.002053.

Miller SR, Abresch HE, Baroch JJ, Fishman Miller CK, Garber AI, Oman AR, Ulrich NJ. Genomic and functional variation of the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. Microorganisms. 2022:10(3):569. https://doi.org/10.3390/microorganisms10030569.

Miller SR, Abresch HE, Ulrich NJ, Sano EB, Demaree AH, Oman AR, Garber AI. Bacterial adaptation by a transposition burst of an invading IS element. Genome Biol Evol. 2021:13:evab245. https://doi.org/10.1093/gbe/evab245.

Miller SR, Augustine S, Olson TL, Blankenship RE, Selker J, Wood AM. Discovery of a free-living chlorophyll *d*-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. Proc Natl Acad Sci U S A. 2005:102(3):850–855. https://doi.org/10.1073/pnas.0405667102.

Miller SR, Wood AM, Blankenship RE, Kim M, Ferriera S. Dynamics of gene duplication in the genomes of chlorophyll *d*-producing cyanobacteria: implications for the ecological niche. Genome Biol Evol. 2011:3:601–613. https://doi.org/10.1093/gbe/evr060.

Miyashita H, Ikemoto H, Kurano N, Adachi K, Chihara M, Miyachi S. Chlorophyll *d* as a major pigment. Nature. 1996:383(6599):402. https://doi.org/10.1038/383402a0.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.

Ohno S. Evolution by gene duplication. New York: Springer-Verlag; 1970.

Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003:424(6945):194–197. https://doi.org/10.1038/nature01771.

Prasad D, Arora D, Nandicoori VK, Muniyappa K. Elucidating the functional role of *Mycobacterium smegmatis recX* in stress response. Sci Rep. 2019:9(1):10912. https://doi.org/10.1038/s41598-019-47312-3.

Qian W, Liao B-Y, Chang AY-F, Zhang J. Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet. 2010:26(10):425–430. https://doi.org/10.1016/j.tig.2010.07.002.

Reams AB, Kofoid E, Savageau M, Roth JR. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. Genetics. 2010:184(4):1077–1094. https://doi.org/10.1534/genetics.109.111963.

Reams AB, Neidle EL. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Mol Microbiol. 2003:47(5):1291–1304. https://doi.org/10.1046/j.1365-2958.2003.03342.x.

Rodin SN, Riggs AD. Epigenetic silencing may aid evolution by gene duplication. J Mol Evol. 2003:56(6):718–729. https://doi.org/10.1007/s00239-002-2446-6.

Romero D, Palacios R. Gene amplification and genomic plasticity in prokaryotes. Annu Rev Genet. 1997:31(1):91–111. https://doi.org/10.1146/annurev.genet.31.1.91.

Sánchez-Baracaldo P. Origin of marine planktonic cyanobacteria. *Sci. Rep.* 2015:5:17418.

Sandegren L, Andersson DI. Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol. 2009:7(8):578–588. https://doi.org/10.1038/nrmicro2174.

Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C_T method. Nat Protocols. 2008:3(6):1101–1108. https://doi.org/10.1038/nprot.2008.73.

Scott JR. Regulation of plasmid replication. Microbiol Rev. 1984:48(1):1–23. https://doi.org/10.1128/mr.48.1.1-23.1984.

Skarstad K, Boye E. Degradation of individual chromosomes in *recA* mutants of *Escherichia coli*. J Bacteriol. 1993:175(17):5505–5509. https://doi.org/10.1128/jb.175.17.5505-5509.1993.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006:34(Web Server):W609–W612. https://doi.org/10.1093/nar/gkl315.

Swingley WD, Chen M, Cheung PC, Conrad AL, Dejesa LC, Hao J, Honchak BM, Karbach LE, Kurdoglu A, Lahiri S, et al. Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. Proc. Natl. Acad. Sci U S A. 2008:105(6):2005–2010. https://doi.org/10.1073/pnas.0709772105.

Ulrich NJ, Uchida H, Kanesaki Y, Hirose E, Murakami A, Miller SR. Reacquisition of light-harvesting genes in a marine cyanobacterium confers a broader solar niche. Curr Biol. 2021:31(7):1539–-1546.e4. https://doi.org/10.1016/j.cub.2021.01.047.

Watanabe M, Semchonok DA, Webber-Birungi MT, Ehira S, Kondo K, Narikawa R, Ohmori M, Boekema EJ, Ikeuchi M. Attachment of phycobilisomes in an antenna–photosystem I supercomplex of cyanobacteria. Proc Natl Acad Sci. 2014:111(7):2512–2517. https://doi.org/10.1073/pnas.1320599111.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007:39(4):457–466. https://doi.org/10.1038/ng1990.

Wood AM, Miller SR, Li WKW, Castenholz RW. Preliminary studies of cyanobacteria, picoplankton, and virioplankton in the Salton Sea with special attention to phylogenetic diversity among eight strains of filamentous cyanobacteria. Hydrobiologia. 2002:473(1/3):77–92. https://doi.org/10.1023/A:1016573400010.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Associate editor: Brian Golding