2017

CONTRIBUTIONS OF GENE COPY NUMBER VARIATION TO GENOME EVOLUTION AND LOCAL ADAPTATION OF THE CYANOBACTERIUM ACARYOCHLORIS

Amy L. Gallagher
University of Montana, Missoula

CONTRIBUTIONS OF GENE COPY NUMBER VARIATION TO GENOME EVOLUTION AND LOCAL ADAPTATION OF THE CYANOBACTERIUM ACARYOCHLORIS

By

AMY LYNN GALLAGHER

B.S. Biology, University of New Mexico, Albuquerque, New Mexico, 2013
B.S. Psychology, University of New Mexico, Albuquerque, New Mexico, 2013

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
Cellular, Microbial, and Molecular Biology

The University of Montana
Missoula, MT

May 2017

Approved by:

Scott Whittenburg, Dean of The Graduate School
Graduate School

Dr. Scott Miller, Committee Chair
Division of Biological Sciences

Dr. Frank Rosenzweig, Committee Member
Georgia Tech; Biological Sciences

Dr. John McCutcheon, Committee Member
Division of Biological Sciences

Dr. James Elser, Committee Member
Division of Biological Sciences

Gallagher, Amy, M.S., May 2017 Cellular, Microbial, and Molecular Biology

Abstract Title – CONTRIBUTIONS OF GENE COPY NUMBER VARIATION TO GENOME EVOLUTION AND LOCAL ADAPTATION OF THE CYANOBACTERIUM ACARYOCHLORIS

Chairperson: Dr. Scott Miller

Acaryochloris is a recently discovered genus of cyanobacteria, unique in its use of an uncommon chlorophyll as its major photosynthetic pigment, and in its peculiar genome dynamics. Members of this genus exhibit increased genic copy number variation (CNV), which is thought to be primarily derived from gene duplications and horizontal gene transfer (HGT). *Acaryochloris* provides an ideal system to explore mechanisms behind maintenance of gene duplicates and the influence of CNV in local adaptation. Here, I propose a mechanism for retention of gene duplicates of the bacterial recombinase, RecA, in *Acaryochloris* genomes and provide preliminary evidence that these paralogs are becoming functionally divergent. I then focus on idiosyncratic CNV between two strains of *Acaryochloris* which were isolated from very different environments. I provide evidence of local adaptation to iron limitation in one strain, associate it with physiological differences between strains, and show that unique CNV drives changes in gene dosage and is associated with variable fitness and physiology.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Dr. Scott Miller and the rest of the Miller Lab: Dr. Emiko Sano, Reid Longley, Bridget Creel, Kayli Anderson

Family: Mom, Dad, Rachel, Shannon

Friends: Patrick Kelly, Harmen Steele, Charlotte Rice, Alexis Magana-Jaggli, Niko Maffucci, Lot van der Graaf, Charlie Katerba, Charlie Villanueva, Jenelle Dowling, Cedar Mathers-Winn

Dr. Frank Rosenzweig

DBS graduate students: Kory Kolis, Alexis Billings, Brett Addis, Gerard Sapes, Sean Pinnell, Zak Clare-Salzler, Cedar Mathers-Winn

Committee members Dr. John McCutcheon and Dr. Jim Elser

All DBS staff members, especially Jill Burke, Zac Raasch, and Jay Bruns

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Abstract

Bacterial recombinase RecA is a multifunctional protein involved in homologous recombination, DNA damage repair, activation and activities of error-prone DNA polymerases, and the regulation of protein activity and gene expression through its coprotease activity. It is apparent that tight regulation of this gene is necessary for normal cell function, and most bacteria have a single copy of the gene. Members of the recently discovered cyanobacterial clade, *Acaryochloris*, have been found to harbor duplicates of the *recA* gene which are constitutively expressed. Overall, there has been selection against protein change during *Acaryochloris recA* diversification. However, certain codons at functionally important sites are predicted to have experienced bursts of positive selection ($d_N/d_S > 1$), which is suggestive of functional divergence. I propose a model of retention for *Acaryochloris recA* paralogs and test the hypothesis that they are becoming functionally divergent. To do this, an *in vivo E. coli* model was developed by cloning individual *Acaryochloris recAs* into a rhamnose-inducible plasmid and inserting the resulting recombinant DNA into *recA* deficient *E. coli.* To determine phenotypic consequences of *Acaryochloris* RecAs in this background, I (1) performed growth experiments as an indirect assay of *in vivo* recombinase activity and (2) challenged the ability of these proteins to complement wild-type *E. coli* RecA in regulating the SOS response. I observed that *Acaryochloris* RecA paralogs differ in their ability to complement the reduction in growth rate observed in *recA* deficient *E. coli*. RecA paralogs do not appear to have the ability to perform the suite of functions needed for a successful SOS response in *E. coli*. Results from the growth experiment support the hypothesis that *Acaryochloris recA* paralogs are becoming functionally divergent in respect to their recombinase activity.

Introduction

Homologous recombination (HR) by members of the Recombinase A family (including RecA in bacteria, RadA in archaea and Rad51 in eukaryotes; Haldenby *et al*., 2009) makes important creative, stabilizing, and destabilizing contributions to genome structure and organismal fitness. Recombinase-mediated HR can assort standing genetic variation into novel combinations (e.g., in meiosis) as well as lead to the innovation of new gene functions or positive dosage effects by gene duplication (Andersson & Hughes, 2009; Bergthorsson *et al*., 2007; Kondrashov, 2012; Ohno, 1970). Recombinase activity is also essential for maintaining genome integrity by repairing stalled or broken DNA

replication forks (Cox *et al*., 2000). However, this activity may also be a source of genetic instability through genomic rearrangements if not properly regulated (Holthausen *et al*., 2010; Krejci *et al*., 2012). For example, increased recombinase activity due to overexpression of *recA* in *E. coli* resulted in a ten-fold increase in duplication rate (Dimpfl and Echols, 1989). In eukaryotes, Rad51 overexpression can result in the disruption of the cell cycle, and in mammals has been found to be overexpressed in some tumor cell lines (Klein, 2008).

The need for cells to tightly control recombinase activity may explain why the vast majority of bacterial genomes have a single copy of the *recA* gene (de Groot *et al*., 2009; Nahrstedt *et al*., 2005; Norioka *et al*., 1995). Exceptions include the genomes of many insect endosymbionts, which lack a copy of *recA* (Moran *et al*., 2008). There are also a few examples of bacteria with more than one copy of the gene. *Bacillus megaterium* and *Myxococcus xanthus* both have two functional copies of *recA* (Nahrstedt *et al.,* 2005; Norioka *et al.,* 1995). Three copies of the gene are present in radiation-tolerant *Deinococcus deserti* (de Groot *et al.,* 2009), and a strain of *Vibrio cholerae* has been found to have a second copy of *recA* inserted on a genomic island (Rapa *et al.,* 2015). With the exception of two paralogous, plasmid-borne copies in *D. deserti*, multiple copies of *recA* in bacterial genomes are the product of horizontal gene transfer (HGT) rather than gene duplication (Nahrstedt *et al.*, 2005; Norioka *et al.*, 1995; de Groot *et al.*, 2009; Rapa *et al.*, 2015).

Remarkable among these exceptions is *recA* copy number variation in *Acaryochloris*, a recently discovered clade of unicellular cyanobacteria that is unique in their use of Chlorophyll *d* as the principal pigment in photosynthesis (Kühl *et al*., 2005;

60 Miller *et al*., 2005; Miyashita *et al*., 1996). To date, two genomes have been sequenced
61 and are publicly available, *Acaryochloris* strain MBIC11017 and *Acaryochloris* strain
62 CCMEE 5410. These strains have seven and four copies of *recA*, respectively (Swingley
63 *et al*., 2008; Miller *et al*., 2011), that have arisen by gene duplication (Miller *et al*., 2011).
64 All *Acaryochloris recA* copies are constitutively expressed but are differentially regulated
65 in response to various environmental stresses (Sano and Miller, unpublished). Although
66 there has generally been selection against amino acid changes during the diversification
67 of *Acaryochloris recA* paralogs, certain codons at functionally important sites are
68 predicted to have experienced diversifying selection ($d_N/d_S > 1$; Miller *et al*., 2011).
69 Together, this suggests that some paralogs may have diverged in function. To explore
70 this possibility, below I discuss the multiple functions of RecA, with a focus on *E. coli*
71 RecA, which has been widely studied (M M Cox & Lehman, 1981; M M Cox & Lehman,
72 1982; Holthausen *et al*., 2010; Lusetti & Cox, 2002).

73 RecA is a jack of all trades. During HR, a RecA-ssDNA nucleoprotein filament
74 generated as a result of a DNA lesion or double stranded break recognizes homologous
75 dsDNA and promotes both complementary base pairing and strand exchange, resulting in
76 a heteroduplex complex with the undamaged dsDNA strand used as a template for repair
77 (M M Cox & Lehman, 1981; Harmon *et al*., 1996). In addition to these homology search
78 and strand exchange functions, *E. coli* RecA affects changes in gene expression and
79 protein activity through its co-protease activity. For example, the SOS response to DNA
80 damaging agents is made up of over forty proteins regulated by the transcriptional
81 repressor LexA (Indiani *et al*., 2013). DNA damage results in the formation of an active
82 RecA nucleoprotein filament which regulates the response by stimulating the

autocleavage of LexA (Little, 1991). Cleavage of this repressor results in increased transcription of genes involved in SOS response (Indiani *et al.*, 2013; Little, 1991). Among the genetic elements regulated by LexA are *dinB,* and the *umuDC* operon, which encode components of error-prone TLS polymerases Pol IV and Pol V, respectively (Indiani *et al.*, 2013). Active Pol V requires proteolysis of UmuD, which is also mediated by the coprotease activity of the RecA nucleoprotein filament (Bianco & Kowalczykowski, 1998; Jiang *et al*., 2009; Nohmi *et al*., 1988). Finally, RecA monomers themselves are active components of both Pol IV and Pol V (Gruber *et al*., 2015; Jiang *et al*., 2009; Patel *et al*., 2010).

RecA is also a master of none, with improvement of one function often entailing a cost for one or more other functions (McGrew and Knight, 2003; Harmon *et al.*, 1996). RecA mediated strand exchange during HR and its coprotease activity are competing process that cannot occur simultaneously (Harmon *et al.*, 1996). Adaptation of individual sub-functions may therefore be constrained by its pleiotropic consequences, a phenomenon referred to as "adaptive conflict" (Hughes, 1994; Lynch and Katju, 2004; Des Marais and Rausher, 2008). *Acaryochloris recA* duplication may set the stage for the escape from adaptive conflict (EAC) via potential specialization of paralogs on different enzyme sub-functions.

I propose a model of retention for *Acaryochloris recA* paralogs involving subfunctionalization, specifically duplication-degeneration-complementation (DDC) (Force *et al.*, 1999), specialization, or a combination of the two. These models are both predicated upon an ancestral single copy gene with more than one function; the main difference is that DDC does not require adaptive evolution, while specialization does

106 (Hahn, 2009). According to the DDC model, duplicate genes can be retained when they
107 undergo loss-of-function or reduction of expression mutations (degeneration) affecting
108 different sub-functions (Force *et al.*, 1999). This results in paralogs with complementary
109 functions that together perform the same task or suite of tasks as the progenitor gene.
110 These paralogs are "locked into" the genome, because together they are required to
111 perform the full function of the ancestral copy (Hahn, 2009; Lynch and Force, 2000).
112 The specialization model of duplicate retention proposes that once a multifunctional
113 progenitor gene is duplicated, duplicates are able to improve (specialize) on individual
114 ancestral sub-functions. This results in functionally specialized paralogs that often have
115 reduced ability to perform all functions of the progenitor as a consequence (Des Marais &
116 Rausher, 2008; Hahn, 2009; Nasvall *et al*., 2012). Again, this would lock multiple
117 paralogs into the genome in order to provide full ancestral functionality. Specialization
118 can be especially important if the multifunctional progenitor gene is under adaptive
119 conflict before duplication, as duplication followed by specialization is a mechanism for
120 escape from adaptive conflict (EAC) (Des Marais and Rausher, 2008). While this model
121 may free *Acaryochloris recAs* from adaptive conflict, locking multiple *recAs* into a
122 genome may have deleterious effects, as over-expression of the gene is known to be
123 harmful.
124 As a first step in appraising the hypothesis of functional divergence of
125 *Acaryochloris recA* paralogs, I tested predictions of the retention model that paralogs
126 retain yet vary in their *in vivo* recombinase activity. To do this, I developed an *E. coli*
127 model by cloning individual paralogs into *recA* deficient *E. coli*. This mutant exhibits a
128 growth defect, since RecA-mediated recombination is used to bypass stalled replication

129 forks (Capaldo *et al*., 1974). Comparing growth rates of the various constructs was used
130 to indirectly measure recombinase activity. I predicted that all paralogs would have some
131 recombinase activity, but would vary to the degree they could complement the growth
132 defect exhibited by the Δ*recA* construct. In addition, I tested survival following SOS
133 induction in the *E. coli* model by exposing constructs to two DNA damaging agents, UV
134 radiation and mitomycin C (Janion, 2008). I predicted that *Acaryochloris* paralogs would
135 partially complement the sensitivity of the *recA* mutant. I did not expect any to fully
136 complement the defect based on previous experiments with cyanobacterial RecAs in *E.*
137 *coli* backgrounds (Domain *et al.*, 2004; Owttrim and Coleman, 1987; Murphy *et al.*,
138 1987). By testing various RecA functions of *Acaryochloris recA* paralogs, I have taken a
139 first step toward addressing the nature of the individual activities of paralogs as well as
140 drawing preliminary conclusions regarding mechanisms maintaining paralogs in the
141 genome.

142 **Methods**

143 *Preparing constructs*

144 Four copies of *recA* from *Acaryochloris marina* strain MBIC 11017 (AM1_3550,
145 AM1_5031, AM1_5483, AM1_B0414) as well as *E. coli* strain TR6968 *recA* were
146 amplified via PCR, size selected on an agarose gel, and purified using the Omega bio-tek
147 E.Z.N.A Gel Extraction Kit. Purified DNA was then inserted into pUC19 cloning vector
148 (NEB) using restriction enzyme SmaI (NEB). Ligated vectors containing a *recA* gene
149 were then cleaned using the Zymoclean Gel DNA Recovery Kit. Gel purified vectors
150 with inserted *recA* were transformed into DH5α competent *E. coli* cells plated on
151 lysogeny broth (LB) agar plates with ampicillin (amp) (100 µg/ml), IPTG (0.1 mM), and

X-Gal (20 µg/ml). Individual colonies containing pUC19 with an inserted *recA* were inoculated into 2 mL LB + 100 µg/ml amp and grown overnight at 37 °C with agitation. Next, boiling lysis miniprep was used to isolate the plasmids. *recAs* were excised from the purified plasmids, subcloned into pBLU vector, and again transformed into DH5α competent *E. coli* cells. The vectors with *recA* inserts were purified using the Promega mini-prep kit and digested using restriction enzymes SpeI and NotI to isolate *recA* inserts. Agarose gel electrophoresis was performed on the digest, and bands corresponding to the *recA* genes were excised and purified using the Zymoclean Gel DNA Recovery Kit.

Multiple cloning sites were added to Addgene plasmid 40779, resulting in plasmid pRHA. This plasmid was used because it carries the rhamnose inducible promoter of the *rhaB E. coli* gene. Each *recA* gene was inserted into plasmid pRHA and transformed into *E. coli* strain TR6968 Δ*recA* and an empty plasmid was introduced to wild type TR6968. This resulted in four experimental constructs containing *Acaryochloris recA*s, a negative control completely lacking any *recA* gene, and two positive controls containing a copy of native *E. coli recA* on either pRHA or on the chromosome. Constructs are referred to using the following notation: chromosome *recA*/plasmid *recA*. Therefore, our constructs are denoted as -/3550, -/5031, -/5483, -/B0414, -/+, -/-, and WT/- for AM1_3550, AM1_5031, AM1_5483, AM1_B0414, *E. coli recA*, negative control, and WT strain with an empty pRHA respectively (Table 1.1).

Growth experiment

Growth was measured in four different conditions to either repress or induce transcription of the *recA* containing plasmids, pRHA. Rhamnose was supplemented to induce expression, whereas glucose was supplemented to repress expression. Expression

of the plasmid was manipulated both during the overnight pre-growth of cultures, and during the time of measured growth. The four different growth conditions are as follows: A) repressed during both pre-growth and measured growth; B) repressed during pre-growth, induced during measured growth; C) induced during pre-growth, repressed during measured growth; D) induced during both pre-growth and measured growth. A summary of the conditions can be found in Table 1.2.

Overnight pre-growth was carried out at 37ºC with aeration in 2 ml LB + 100 µg/ml ampicillin (amp) + 0.2% (w/v) glucose for repression and in 2 ml LB + 100 µg/ml amp + 0.15% (w/v) glucose + 0.2% (w/v) rhamnose for induction. The growth experiment was carried out by inoculating 20 µl of pre-growth culture into 180 µl of LB + 100 µg/ml amp + 0.2% (w/v) glucose for repression during measured growth conditions or LB + 100 µg /ml amp + 0.2% (w/v) rhamnose for the induction during measured growth conditions, resulting in a total volume of 200 µl per well in a 96 well clear-bottom ThermoScientific assay plate. Absorbance at 600 nm was measured every 15 minutes with a Synergy HT plate reader (BioTek) for 4.5 hours with continual agitation and at 37°C. This experiment was done in biological triplicate for each construct.

Doubling time was estimated using the R package Growthcurver (Sprouffske and Wagner, 2016) and averaged over replicates. For each growth condition, a one-way ANOVA was carried out to determine if construct was a significant predictor of doubling time, followed by post hoc multiple comparisons using Tukey's HSD if the ANOVA analysis was statistically significant at the $\alpha = 0.05$ level.

UV Resistance Assay

Constructs were grown overnight in LB + 100 µg/ml amp at 37°C with aeration, then subcultured at a 1:100 dilution in LB + 100 µg/ml amp + 0.2% (w/v) rhamnose and grown for 3 hours to induce pRHA expression. Samples were then serially diluted into M9 salts and 100 µl of the dilution plated onto LB + 100 µg/ml amp agar plates. All control plates and UV exposed plates of WT/- and -/+ were plated at final dilutions of 10^{-6} and 10^{-7}. UV exposed plates of -/3550, -/5031, -/5483, -/B0414, and -/- were plated at final dilutions of 10^{-4} and 10^{-5}. Experimental plates were exposed to UVC (254nm) from an 8W bulb for three seconds at a distance of 17 cm from the light source (Lamag product number 022.9120) and then immediately placed in the dark. Experimental and control plates were incubated overnight at 37 °C in the dark, colonies were counted the next day and survival rates calculated. A one-way ANOVA was carried out to determine if percent survival after UV exposure differed between constructs. Post hoc comparisons were carried out using Tukey HSD as the ANOVA analysis was statistically significant at the $\alpha = 0.05$ level.

Mitomycin C Resistance Assay

Sensitivity was tested at 0.5 and 1.0 µg/ml mitomycin C. 1.5% agar plates containing LB + 100 µg/ml amp + 0.2% (w/v) rhamnose and either 0.0, 0.5, or 1.0 µg/ml mitomycin C were prepared. Strains were grown overnight in LB + 100 µg/ml amp at 37°C with aeration. Overnight cultures were serially diluted into M9 salts and spot plated (10 µl, 10^{-2} through 10^{-6}) onto the control and MMC agar plates. Plates were incubated at 37°C overnight, and results were recorded the following day.

Results and discussion

To test the hypothesis of functional divergence between *Acaryochloris recA* paralogs, I focused on four *recAs* from *Acaryochloris marina* strain MBIC 11017 (AM1_3550, AM1_5031, AM1_5483, AM1_B0414) for which there are orthologs in the *Acaryochloris* strain CCMEE 5410 genome (Miller *et al.*, 2011). The four *A. marina* and native *E. coli recA* genes were cloned into plasmid pRHA carrying the rhamnose inducible promoter of the *rhaB E. coli* gene (see Methods). These plasmids and an empty plasmid control were introduced to Δ*recA E. coli*. Additionally, I made a construct consisting of an empty plasmid in a WT *E. coli* background, which was used to control for effects the plasmid, independent of *recA*. This resulted in four experimental constructs containing *Acaryochloris recA*s, a negative control lacking a copy of *recA*, and two positive controls containing a copy of native *E. coli recA* on either pRHA or on the chromosome. Constructs are referred to using the following notation: chromosome *recA*/plasmid *recA* (Table 1.1).

Growth experiment: Assaying in vivo recombinase activity

RecA mediated recombination is integral to rescuing stalled replication forks, which occur commonly during DNA replication even under optimal growth conditions (Cox *et al.*, 2000). Δ*recA* mutants tend to exhibit a slower growth phenotype compared with wild type, likely due to the inability to repair stalled replication forks (Cox *et al.*, 2008; Capaldo *et al.*, 1974). To investigate functional divergence in *in vivo* recombinase activity of *Acaryochloris* RecA paralogs, I performed a growth experiment to determine if *Acaryochloris recA*s have differential ability to complement the *recA* null construct's growth defect. If a construct exhibited significantly faster growth than the *recA*-null it

244 was determined to have some ability to complement. Degree of complementation should
245 be positively correlated with recombinase activity. Observing significant differences
246 between paralogs would provide the most compelling evidence for differences in
247 recombinase activity and therefore functional divergence.

248 Growth was measured in four different conditions meant to either repress or
249 induce transcription of plasmid copies of *recA*. Induction or repression was
250 accomplished by taking advantage of the rhamnose inducible promoter inserted into the
251 pRHA plasmid. Rhamnose was supplemented to induce expression, while glucose was
252 supplemented to repress expression. Expression of the plasmid was manipulated both
253 during the overnight pre-growth of cultures and during the time of measured growth. The
254 four different growth conditions were as follows: A) repressed during pre-growth and
255 measured growth (repressed/ repressed); B) repressed during pre-growth, induced during
256 measured growth (repressed/ induced); C) induced during pre-growth, repressed during
257 measured growth (induced/ repressed); D) induced during both pre-growth and measured
258 growth (induced/ induced). A summary of the conditions can be found in Table 1.2.
259 Each growth condition was analyzed separately due to the expected differences in effects
260 of glucose and rhamnose supplementation on growth rate, independent of *recA*
261 expression. Glucose is known to be the preferred carbon source of *E. coli*, resulting in
262 faster growth (Monod, 1949).

263 In condition A, for which expression of the *recA* containing plasmid was
264 repressed during both pre-growth and measured growth periods, construct was a
265 significant predictor of generation time [$F_{(6,14)} = 18.00$, $P < 0.0001$]. As expected, the
266 *recA*-deficient construct exhibited a slow growth phenotype. All constructs containing an

Acaryochloris recA paralog performed better than -/-, but were variable in their ability to complement the observed growth defect. Construct -/3550 did not significantly complement -/-, whereas the three other constructs containing *Acaryochloris* paralogs did complement the null, providing evidence for functional divergence (Fig. 1.1A; Table 1.3). This condition also resulted in generally faster generation time for all constructs, as a consequence of glucose supplementation throughout the experiment (Fig. 1.1).

In condition B, for which expression of the *recA* containing plasmid was repressed during pre-growth and induced during the period of measured growth, construct was a significant predictor of generation time [$F_{(6,14)} = 17.73$, $P = 0.021$]. -/3550 performed significantly worse than -/B0414 and WT/-, but no other comparisons were significant (Fig. 1.1B, Table 1.4). As with condition A, -/3550 exhibited the worst performance of all the constructs containing an *Acaryochloris* paralog. Additionally, all constructs exhibited slower growth in this condition compared with condition A (Fig. 1.1).

In condition C, for which expression of the *recA* containing plasmid was induced during pre-growth and repressed during the period of measured growth, construct was a significant predictor of generation time [$F_{(6,14)} = 241.07$, $P < 0.0001$]. Constructs -/3550 and -/+ exhibited slower growth than the other constructs at highly significant levels, while -/5031 and -/B0414 had significantly slower growth rates than fastest growing WT/- (Fig. 1.1C, Table 1.5). There exists striking difference in performance between the two constructs containing native *E. coli recA* in this condition. Very poor performance of -/+ points to the deleterious effects of *recA* overexpression, as its copy of the gene is on the induced plasmid, while WT/- does not appear to be exhibiting any growth defect.

290 Poor performance of -/+ is matched by -/3550, whose growth is again significantly slower
291 than the other paralog-containing constructs.
292 In condition D, for which expression of the *recA* containing plasmid was induced
293 during both pre-growth and measured growth periods, construct was a significant
294 predictor of generation time [$F_{(6,14)} = 5.15$, $P = 0.0054$]. -/3550 and -/+ exhibited
295 significantly slower growth than WT/-; none of the other comparisons were significant
296 (Fig. 1.1D, Table 1.6). Construct generation time tends to increase under pRHA
297 induction (rhamnose supplementation) conditions, with condition D resulting in slowest
298 generation times for all constructs. As with condition C, we see a striking difference in
299 generation time between the two constructs containing *E. coli recA*. Among the paralog-
300 containing constructs, -/3550 again exhibits the slowest growth.
301 As predicted, *Acaryochloris recA* paralogs varied in their ability to complement
302 the growth defect observed in the *recA*-deficient construct. Variable complementation
303 among *Acaryochloris* paralogs was clearly observed in condition A, pRHA repression
304 during pre and measured growth. In this condition, all paralogs with the exception of -
305 /3550 complemented the null to some degree, indicating that AM1_5031, AM1_5483,
306 and AM1_B0414 have some *in vivo* recombinase activity. Across the remaining
307 experimental conditions, -/3550 was consistent in exhibiting the slowest growth of all the
308 paralogs, often to a statistically significant degree. Unpublished *in vitro* assays suggest
309 AM1_3550 may bind DNA very tightly and interfere with other aspects of DNA
310 metabolism, which may help to explain the poor performance of construct -/3550 in
311 conditions of *recA* induction (Sano, unpublished). Growth defect in the construct
312 containing *E. coli recA* on the inducible plasmid under conditions of induction is likely

313 due to excessive recombination. With the exception of -/3550, constructs containing
314 *Acaryochloris recAs* exhibited a less severe growth defect than that with a plasmid copy
315 of *E. coli recA*, which may indicate that they have reduced recombinase activity. These
316 experiments provide preliminary evidence of functional divergence in *Acaryochloris*
317 *recA* paralogs, specifically in AM1_3550.

318 *SOS response assays*

319 In WT *E. coli*, DNA damage is commonly repaired by the SOS response and HR,
320 both of which are RecA mediated (Indiani *et al*., 2013; Cox *et al*., 2000). The *E. coli*
321 SOS response requires RecA to successfully perform all of its known functions, including
322 formation of a nucleoprotein filament, co-protease activities, and pol involvement
323 (Indiani *et al*., 2013; Little, 1991; Nohmi *et al*., 1988; Jiang *et al*., 2009). To address
324 whether *Acaryochloris* RecAs have the ability to perform these functions and repair DNA
325 damage from mutagens known to induce the SOS response (Schlacher and Goodman,
326 2007; Janion, 2008), I exposed constructs to UV radiation and a potent DNA crosslinker,
327 mitomycin C (MMC).

328 To determine ability to repair DNA damage caused by UV radiation, pRHA
329 expression was induced, constructs were exposed to UVC radiation, and percent survival
330 was estimated. Construct was determined to be a significant predictor of survival rate
331 after UV exposure, as determined by a one-way ANOVA [$F_{(6,14)} = 541.30$, $P < 0.0001$].
332 *Acaryochloris recA*s did not significantly differ in survival rate from the *recA* null. Only
333 constructs containing native *E. coli recA*, -/+ and WT/-, were able to complement the UV
334 sensitivity phenotype exhibited by the *recA* null (Fig. 1.2).

A MMC resistance assay was performed by testing the ability of constructs to grow on agar supplemented with 0.5 and 1.0 µg/ml MMC. All *Acaryochloris recA* paralogs and the *recA* null construct exhibited 100% mortality when exposed to both tested concentrations of MMC. Similar to the UV sensitivity assay, both constructs containing *E. coli recA* were robust to the DNA damage caused by mitomycin C (Fig. 1.3).

Complete failure of *Acaryochloris recA*s to complement these SOS-related defects may be the product of general evolutionary divergence rather than subfunctionalization of *Acaryochloris* paralogs. Cyanobacteria and *E. coli* last shared a common ancestor 2.5 billion years ago (Miller *et al.*, 2005; Summons *et al.*, 1999), and it is known that cyanobacteria and *E. coli* regulate the SOS response through different mechanisms (Domain *et al.*, 2004). Similar assays performed on single-copy cyanobacterial *recAs* from *Anabaena variabilis* and *Synechococcus* PCC 7002 in *ΔrecA* *E. coli* backgrounds resulted in minimal UV resistance (Owttrim and Coleman, 1987; Murphy *et al.*, 1987). While some cyanobacterial *recAs* may have minimal ability to elicit an SOS response, *Acaryochloris* RecAs may be too far diverged from *E. coli* to effectively complement at all. This is supported by the failure of the single copy *recA* of *Cyanothece* PCC 7425, a sister taxon to *Acaryochloris,* to complement the growth defect caused by UV exposure in an experiment using the same methods (Sano & Miller,unpublished).

Concluding remarks

Multiple copies of the multifunctional *recA* gene is an exceedingly rare phenomenon in bacteria, making *Acaryochloris* anomalous. Here, I proposed a

mechanism of retention for these paralogs and hypothesized that paralogs are becoming functionally divergent. I have found preliminary evidence that supports my hypothesis of functional divergence in *Acaryochloris recA* paralogs. This is most apparent in significantly different growth rates among constructs containing paralogs, especially the slow growth phenotype in -/3550. Future experiments should be focused on directly measuring homologous recombination in *Acaryochloris recA* paralogs. Poor survival of constructs containing *Acaryochloris* paralogs after exposure to DNA damaging agents further support evidence that cyanobacterial SOS regulation is too far diverged from *E. coli* SOS regulation for complementation (Domain *et al.*, 2004).

Tables

Table 1.1 Summary of constructs

E. coli strain	*recA* on pRHA	Denoted
ΔrecA TR6968	AM1_3350	-/3550
ΔrecA TR6968	AM1_5031	-/5031
ΔrecA TR6968	AM1_5483	-/5483
ΔrecA TR6968	AM1_B0414	-/B0414
ΔrecA TR6968	*E.coli*	-/+
ΔrecA TR6968	-	-/-
TR6968	-	WT/-

Table 1.2 Summary of growth conditions indicating repression or induction of *recA* containing plasmid. Overnight pre-growth manipulation was accomplished by supplementing 0.2% (w/v) glucose to cultures for repression conditions, 0.15% (w/v) glucose + 0.2% (w/v) rhamnose to cultures for induction. Measured growth manipulation was accomplished supplementing 0.2% (w/v) glucose to cultures for repression and 0.2% (w/v) rhamnose to cultures for induction conditions.

	Treatment			
	A	B	C	D
Overnight pre-growth	Repressed	Repressed	Induced	Induced
Measured growth	Repressed	Induced	Repressed	Induced

369

Table 1.3 Condition A (repressed/ repressed) growth experiment results. (A) Mean generation time in minutes, std. error, and homogenous subset. (B) Post hoc comparisons of constructs, absolute value of mean differences in minutes shown. P: < .001 '***', .01 '**', .05 '*'

A

Construct	Mean	Std. Error	Subset
-/B0414	26.27002	0.264953	d
WT/-	26.35411	0.605625	d
-/5483	26.73211	0.332897	cd
-/5031	26.83748	0.293371	cd
-/+	28.68212	0.546738	bc
-/3550	30.37614	0.422189	ab
-/-	30.96604	0.645076	a

B

	-/B0414	WT/-	-/5483	-/5031	-/+	-/3550	-/-
-/B0414	0.00	0.08	0.46	0.57	2.41*	4.11***	4.70***
WT/-		0.00	0.38	0.48	2.33*	4.02***	4.61***
-/5483			0.00	0.11	1.95	3.64**	4.23***
-/5031				0.00	1.84	3.54**	4.13***
-/+					0.00	1.69	2.28*
-/3550						0.00	0.59
-/-							0.00

370

Table 1.4 Condition B (repressed/ induced) growth experiment results. (A) Mean generation time in minutes, std. error, and homogenous subset. (B) Post hoc comparisons of constructs, absolute value of mean differences in minutes shown. P: < .001 '***', .01 '**', .05 '*'

A

Construct	Mean	Std. Error	Subset
WT/-	33.3644	1.1043968	b
-/B0414	33.77391	0.5418365	b
-/5483	35.49634	0.6018905	ab
-/-	36.65059	2.0845105	ab
-/5031	36.83823	1.1481173	ab
-/+	38.75225	1.8206455	ab
-/3550	39.9693	0.6639814	a

B

	WT/-	-/B0414	-/5483	-/-	-/5031	-/+	-/3550
WT/-	0.00	0.41	2.13	3.29	3.47	5.39	6.60*
-/B0414		0.00	1.72	2.88	3.06	4.98	6.20*
-/5483			0.00	1.15	1.34	3.26	4.47
-/-				0.00	0.19	2.10	3.32
-/5031					0.00	1.91	3.13
-/+						0.00	1.22
-/3550							0.00

371

Table 1.5 Condition C (induced/ repressed) growth experiment results. (A) Mean generation time in minutes, std. error, and homogenous subset. (B) Post hoc comparisons of constructs, absolute value of mean differences in minutes shown. P: < .001 '***', .01 '**', .05 '*'

A

Construct	Mean	Std. Error	Subset
WT/-	29.01193	0.742744	c
-/5483	34.80734	0.978227	bc
-/-	35.12383	1.505831	bc
-/B0414	37.02259	1.715167	b
-/5031	39.39755	0.718867	b
-/+	49.21383	2.986592	a
-/3550	54.7175	1.371775	a

B

	WT/-	-/5483	-/-	-/B0414	-/5031	-/+	-/3550
WT/-	0.00	5.80	6.11	8.01*	10.39**	20.20***	25.71***
-/5483		0.00	0.32	2.22	4.59	14.41***	19.91***
-/-			0.00	1.90	4.27	14.09***	19.59***
-/B0414				0.00	2.37	12.19**	17.69***
-/5031					0.00	9.82**	15.32***
-/+						0.00	5.50
-/3550							0.00

Table 1.6 Condition D (induced/ induced) growth experiment results. (A) Mean generation time in minutes, std. error, and homogenous subset. (B) Post hoc comparisons of constructs, absolute value of mean differences in minutes shown. P: < .001 '***', .01 '**', .05 '*'

A

Construct	Mean	Std. Error	Subset
WT/-	37.78166	1.199528	b
-/5483	51.71957	3.029789	ab
-/B0414	53.74087	5.946096	ab
-/-	58.3682	10.95534	ab
-/5031	59.50297	2.331568	ab
-/3550	87.24069	2.019082	a
-/+	95.01191	19.72382	a

B

	WT/-	-/5483	-/B0414	-/-	-/5031	-/3550	-/+
WT/-	0.00	13.94	15.96	20.59	21.72	49.46*	57.23**
-/5483		0.00	2.02	6.65	7.78	35.52	43.29
-/B0414			0.00	4.63	5.76	33.50	41.27
-/-				0.00	1.13	28.87	36.64
-/5031					0.00	27.74	35.51
-/3550						0.00	7.77
-/+							0.00

374

Figures

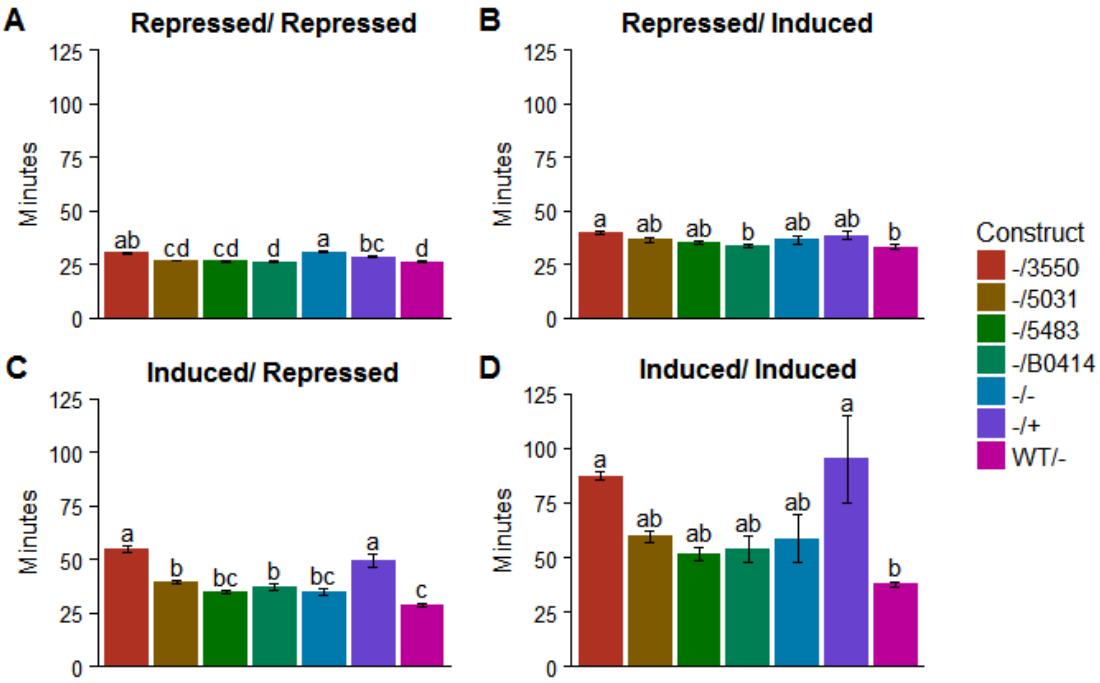
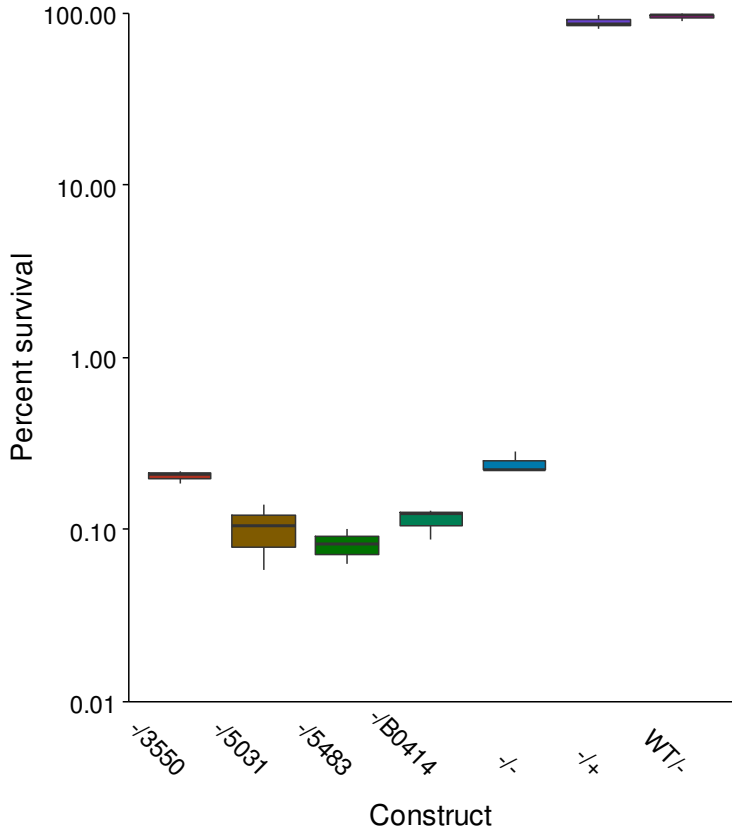


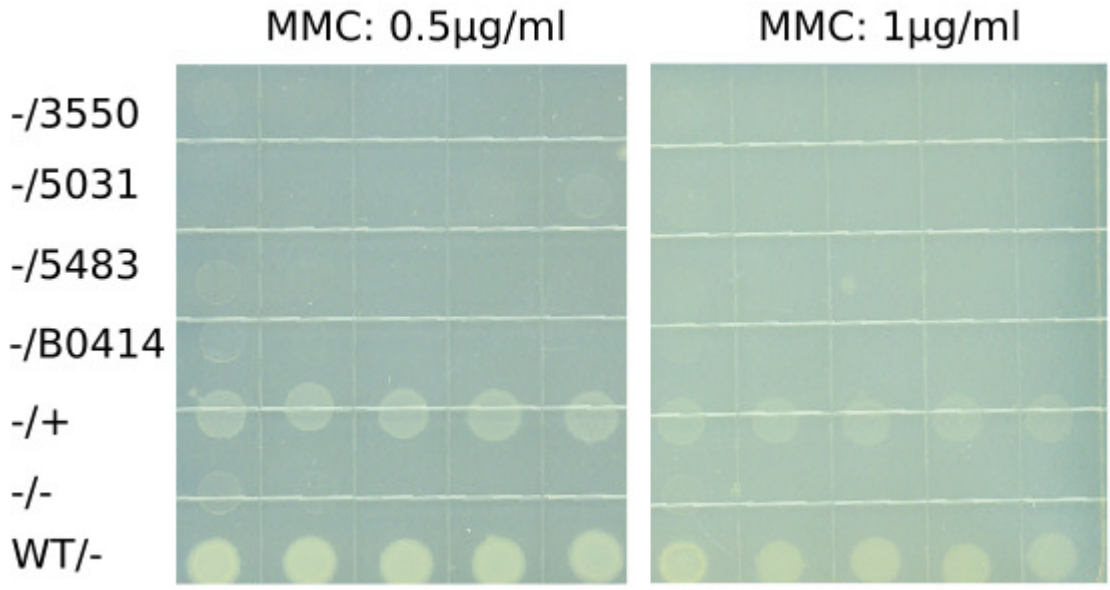
Fig. 1.1 Generation time in minutes of constructs grown in conditions of plasmid expression repression during overnight pre-growth and measured growth (A), repression during overnight pre-growth and induction during measured growth (B), induction during overnight pre-growth and repression during measured growth (C), and induction during overnight pre-growth and measured growth (D). Detailed information regarding induction and repression conditions can be found in Table 1.2. Letters above bars indicate statistically significant homogeneous subsets.



386

387 **Fig. 1.2** Percent survival of constructs after exposure to UVC radiation.

388



389

390 **Fig. 1.3** Constructs grown overnight, subcultured at 1:10 and allowed to grow for three
391 hours before being serially diluted 1:10 and spot plated on LB + 100 µg/ml amp + 0.2%
392 (w/v) rhamnose and 0.5 or 1.0 µg/ml MMC. Media used for overnight and subcultured
393 growth was supplemented with rhamnose to induce expression of the *recA* containing
394 plasmid.
395

Chapter 2

Abstract

Local adaptation of natural populations is thought to be a driver of microbial genetic diversity. Within the cyanobacterial genus *Acaryochloris* are strains that have been isolated from strikingly different environments and have unique gene content which potentially contributes to adaptation in their local, native environments. One strain, MBIC, was isolated from an iron-limited environment and contains increased copy number variation (CNV) of genes involved in iron assimilation. Here, I establish better adaptation to iron limitation in MBIC, as compared to *Acaryochloris* strain CCMEE, which was isolated from a heavy-metal enriched environment. Adaptation is then associated with physiological differences in strains' ability to assimilate iron and finally differences in iron assimilation gene dosage between strains.

Introduction

Local environments impose different selective pressures on natural populations. Over time, this may result in the evolution of locally adaptive traits. A locally adapted population exhibits higher relative fitness in its local environment than populations from other habitats (Kawecki and Ebert, 2004). Microorganisms found in nature tend to be incredibly genetically diverse, which is thought to be a result of local adaptation (Kraemer and Boynton, 2017). An important mechanism by which local adaptation occurs in microorganisms is through changes in gene dosage. Gene dosage, or the amount of a particular gene product transcribed, is commonly affected by changes in transcriptional regulation mechanisms, gene duplication events, and horizontal gene transfer (HGT) (Andersson and Hughes, 2009). Gene duplication and HGT can result in genic copy number variation (CNV), which is commonly associated with increased dosage (Kondrashov, 2012).

Experimental evolution studies of microbial populations have shown that increased gene dosage by duplication events can be an important mechanism for adaptation to nutrient limitation (Brown *et al.*, 1998; Cairns and Foster, 1991; Reams and

425 Neidle, 2003; Riehle *et al.*, 2001). However, there is limited evidence of duplication
426 driving adaptation in natural populations of microorganisms. By contrast, HGT has been
427 shown to be important in niche expansion and local adaptation; it a likely driver of
428 observed phenotypic variability within closely related taxa (Lawrence and Ochman,
429 1998; Ochman *et al.*, 2000; Schönknecht *et al.*, 2013).

430 Members of the recently discovered cyanobacterial genus *Acaryochloris* which
431 are unique in their use of the far-red (>700 nm) light absorbing Chlorophyll *d* (Chl *d*) as
432 the major pigment in photosynthesis (Miyashita *et al.*, 1996; Swingley *et al.*, 2008; Kühl
433 *et al.*, 2005), provide an excellent system to investigate the potential role of CNV in
434 adaptation. *Acaryochloris* has unusual genome evolutionary dynamics, with gene
435 duplication rates considerably higher than other bacteria (Miller *et al.*, 2011). Although
436 this appears to be a generally non-adaptive process, with most duplicates purged from the
437 genome relatively quickly, retained gene duplicates are potentially beneficial in their
438 local, native environments. For example, the genome of *Acaryochloris* strain CCMEE
439 5410 (hereafter, CCMEE), which was isolated from the heavy metal enriched Salton Sea
440 (Miller *et al.*, 2005), retains duplicate genes involved in copper tolerance (Miller *et al.*,
441 2011). Similarly, the genome of *Acaryochloris* strain MBIC 11017 (hereafter, MBIC)
442 has multiple duplicates for genes involved in iron assimilation. This strain was isolated
443 from a biofilm on the underside of *Lissoclinum patella*, a colonial ascidian found in the
444 western Pacific Ocean (Miyashita *et al.*, 1997; Kühl *et al.*, 2005). This is thought to be a
445 severely iron-limited environment due to the low solubility of Fe (III) in circumneutral
446 water (Kranzler *et al.*, 2013) and the ability of *L. patella* to accumulate high levels of iron
447 (Endean, 1954).

In addition to their idiosyncratic retention of duplicated genes, *Acaryochloris* genomes also harbor extensive, plasmid-encoded, novel gene content (Miller *et al.*, 2011) that may in some cases be selectively favored. For example, the MBIC genome exhibits an increased copy number of novel, potentially horizontally-acquired genes involved in iron assimilation compared with the CCMEE genome. These include a cluster of genes responsible for the synthesis of secreted iron-binding molecules called siderophores that appear to be most closely related to a known siderophore-producing cluster in the distantly-related, multicellular cyanobacterium *Anabaena* sp. PCC 7120 (Jeanjean *et al.*, 2008; Fig. 2.1). *Acaryochloris* provides a system to investigate local adaptation by associating fitness in, and physiological responses to particular environmental conditions with idiosyncratic CNV in closely related genomes.

Here, I test whether MBIC has adapted to a low iron environment, and if so, whether this is associated with both enhanced iron assimilation and increased dosage of iron assimilation genes. If MBIC has adapted to low iron, I predict it will exhibit greater fitness relative to CCMEE under iron limitation by growing at faster rate and to a greater final yield. I also predict that MBIC will exhibit an enhanced ability to assimilate iron during recovery from iron starvation. Finally, I expect that physiological performance will be associated with increased dosage of iron assimilation genes in MBIC. Together; this study will enable me to assess the potential contribution of CNV and gene dosage to adaptation of MBIC to low iron conditions.

Methods

Culture conditions

All cultures were grown at 30 °C with constant shaking at 100 rpm and constant illumination of 13-18 µmol photons/m^2s cool white fluorescent light. Cultures were either grown as 100 ml media in 250 ml Erlenmeyer flasks or 600 ml media in 1 L Erlenmeyer flasks. Two types of media were used, one for the high iron condition and one for the low iron condition (Swingley media and Swingley$_0$, respectively). Swingley media was prepared as previously described, though referred to as FeMBG-11 (Swingley *et al.*, 2005). Swingley$_0$ was prepared in the same fashion as Swingley media, while leaving out ferric ammonium citrate and EDTA iron(III) sodium salt. Ferric iron content in Swingley media is 51 µM, and estimated 7.7 nM in Swingley$_0$. In order to minimize iron contamination, all media was prepared using MilliQ filtered water in polycarbonate culture flasks that had been soaked overnight in 1N HCl.

Cell count and optical density regression

Culture turbidity was used to measure growth for both strains, this was done by taking optical density readings at 750 nm (OD_{750}) using a Beckman Coulter DU 530 spectrophotometer. A regression of optical density and cell count for both *Acaryochloris* strains was produced in order to normalize results to cell count and ultimately cell volume. MBIC cells are smaller than CCMEE; their approximate diameters are 1.75 µm and 2.75 µm, respectively. I prepared various dilutions or concentrations of cell cultures while in mid-exponential phase, performed cell counts using a hemocytometer, and took OD_{750} readings. To ensure accuracy, between twelve and fourteen counts at various dilutions were used for each strain. Cell count was then plotted as a function of optical

density at 750 nm and a best fit line was applied in MS Excel. I used the following equations to estimate cell count from optical density in future experiments. The regressions had R^2 of 0.91 for MBIC and 0.97 for CCMEE.

$$MBIC\,\frac{cells}{ml} = 4 \times 10^8\ (OD_{750}) - \ 3 \times 10^6$$

$$CCMEE\,\frac{cells}{ml} = 2 \times 10^8\ (OD_{750}) + \ 3 \times 10^6$$

Growth Experiments

For each strain, triplicate independent cultures derived from the same inoculum were grown in Swingley and Swingley$_0$ media. Media were prepared as described above using 250 ml polycarbonate flasks with a final volume of 100 ml, approximately 1x10^6 stationary phase cells from stocks maintained in Swingley$_0$ medium were used to inoculate each flask. Growth was measured by taking OD_{750} readings every 24-48 hours. Doubling, or generation, times (*G)* were estimated from the exponential growth phase of the culture, as determined by plotting the growth data on a semi-log plot, finding time intervals where cultures were exponentially growing and applying the following formula:

$$G = \frac{T_f - T_0}{3.3 \log\left(\frac{OD_f}{OD_0}\right)}$$

T_0 and T_f are the first and last time points during which the cells are growing exponentially. OD_0 and OD_f are the OD_{750} readings corresponding to T_0 and T_f.

Iron step-up

Five 1 L flasks containing 600 ml Swingley$_0$ media were prepared for each strain for an iron step-up experiment. Approximately 6x10^6 cells were inoculated into each flask from stocks growing in Swingley$_0$, previously described growth conditions were

512 applied. OD_{750} was recorded at regular intervals to measure growth. Once the cultures
513 had been in stationary phase for 7 days I harvested cells for RNA extraction, intracellular
514 iron analysis, chlorophyll *d* content, and took an OD_{750} reading; this is referred to as time
515 0 (t0). After collecting data for t0, I supplemented culture flasks with iron in the
516 following forms and concentrations: EDTA iron(III) sodium salt (28 µM) and ferric
517 ammonium citrate (23 µM). The types and concentrations of iron added were chosen to
518 equal the iron content in Swingley media (Swingley *et al.*, 2005). At 12, 24, and 36
519 hours after iron supplement (t12, t24, t36), an OD_{750} reading was taken and samples were
520 collected for both intracellular iron and chlorophyll *d* content analyses. At 36 hours post
521 iron addition, cells were again collected for RNA isolation.
522 *Chlorophyll d extraction and concentration estimation*
523 Chlorophyll *d* levels were monitored by harvesting 1.2 ml of culture by
524 centrifugation at 16,000 X *g* for 5 minutes. Supernatant was then aspirated and cell
525 pellets were resuspended in 1.2 ml ice cold 100% methanol by vortexing. Samples were
526 stored on ice in the dark for 12-15 minutes to extract pigments, after which they were
527 again centrifuged at 16,000 X *g* to pellet debris (Schliep et al., 2010). Absorbance was
528 measured at 696 nm on a Beckman Coulter DU 530 spectrophotometer and concentration
529 of chlorophyll *d* in µg/ml was determined using the published mass extinction coefficient
530 of chl *d* (77.62 L g^{-1} cm^{-1}; Li *et al*., 2012).
531 *Intracellular iron collection, digestion, and analysis*
532 To determine intracellular iron content, 10 ml of culture from each sample was
533 filtered onto 0.6 µm pore size polycarbonate membrane filters (MILLIPORE product
534 DTTP02500). Filters were inserted into 2 ml screw-top microcentrifuge tubes and 1 ml

of 5 mM EDTA pH 7.8 was added. Tubes with filters and EDTA were vortexed until cells were resusupended in the solution and filters were removed with a toothpick. Samples were then centrifuged for 10 minutes at 16,000 X *g* to pellet cells and supernatant was aspirated. Next, 1 ml sterile Swingley0 media was added to the tubes and pellets were resuspended by vortexing, samples were again centrifuged 10 minutes at 16,000 X *g* and supernatant was aspirated. Cell pellets were stored at -20 ºC until chemically digested for iron content analysis using optical emission spectroscopy.

A modified protocol of EPA method 3050B was used to digest cell pellets for iron content analysis (Environmental Protection Agency). Millipore water and 70% TraceMetal Grade nitric acid were added in 1:1 ratio to microcentrifuge tubes containing cell pellets to a final volume of 1 ml, tubes were then vortexed to resuspend the pellet and incubated at 85 ºC for 4 hours. Samples were removed from heat and allowed to cool. Once cool, 150 µl of a 30% solution of hydrogen peroxide was added to all samples and they were incubated for 30 minutes at 60 - 70 ºC. Digestions were then added to 19 ml of a 2% nitric acid solution for a final acid concentration of approximately 4.4% and final volume of approximately 20 ml. Once digested, optical emission spectroscopy was performed on each sample by The University of Montana's Environmental Biogeochemistry Laboratory to determine the amount of iron per milliliter *Acaryochloris* culture. Iron concentration could then be normalized to approximate *Acaryochloris* biomass.

To account for any iron precipitation, I used blank controls. At t0, when cells were collected before iron was added, there was no iron detected in the blank sample. Iron concentrations in blanked samples after iron was supplemented were not negligible

and varied approximately two fold. This variation was likely not meaningful, as it is likely the result of accidentally aspirating some of the precipitated iron. Because of this, I averaged the iron concentration in the blank samples and subtracted this number from all iron concentrations taken from samples collected after iron addition (t12, t24, t36).

Cell collection for RNA-seq

RNA-seq was performed on both strains of *Acaryochloris* under three environmental conditions: stationary phase while iron starved, during recovery from iron starvation, and while growing exponentially in iron limited media. For all conditions, there were 5 replicates for each strain. Stationary phase under iron starvation and during recovery correspond to t0 and t36 of the iron step-up experiment, which is discussed above. For the condition of exponential growth in iron limited media both strains were grown to mid-exponential phase in $Swingley_0$ media. Experimental procedures were identical to the beginning of the iron step-up experiment, but cells were collected at an earlier phase in their growth and no iron was added. 200ml of cell culture from each sample was collected for RNA isolation. Cell collection was carried out via vacuum filtration onto 1.2 µm pore size polycarbonate membrane filters (MILLIPORE product RTTP04700). Using sterilized forceps, filters with cells were carefully inserted into 15 ml Falcon tubes. Tubes containing cells on filters were immediately flash frozen in liquid nitrogen and stored at -80 ºC until RNA extraction.

RNA extraction

A guanidium-thiocyanate-phenol-chloroform extraction with PGTX extraction buffer was used to isolate RNA. PGTX buffer was prepared as described by Pinto *et al*., 2009; the extraction protocol used is a combination of the "PGTX 95" protocol outlined

581 in the publication, the Meeks Lab RNA isolation from *Nostoc punctiforme* (Meeks), and
582 the Qiagen RNeasy Mini Handbook with modifications. Falcon™ tubes containing cells
583 on polycarbonate filters were removed from the -80 ºC freezer and 2 ml warmed PGTX
584 reagent was added. Samples were vortexed to resuspend cells and incubated for 5
585 minutes at 95 ºC with occasional vortexing. Immediately, samples were incubated on ice
586 for 5 minutes. During this incubation period, filters were removed with sterile pipette
587 tips. Next, 400 µl chloroform was added and samples were incubated for 10 minutes at
588 room temperature with occasional vortexing. Phase separation was then facilitated by
589 centrifugation for 15 minutes at 4 ºC and 12,000 X *g*. The aqueous layer was transferred
590 to a new tube and an equal volume of chloroform added. Again, extractions were
591 incubated at room temperature for 10 minutes with occasional vortexing and centrifuged
592 for 15 minutes at 4 ºC and 12,000 X *g*. To precipitate RNA, the aqueous layer was
593 transferred to a new tube, 1/10 volume 3M sodium acetate at pH 5.2 and 2.5 volume
594 100% ice cold ethanol were added. Tubes were mixed by inversion and precipitated
595 overnight at -20 ºC.

596 The following day, samples were briefly chilled to -80 ºC and centrifuged 20
597 minutes at 4 ºC and 12,000 X *g* to pellet RNA. Supernatant was aspirated and pellets
598 were washed by resuspension in 1 ml 75% ethanol and pelleted again by centrifugation
599 for 10 minutes at 4 ºC and 12,000 X *g*. Supernatant was again aspirated and a 75%
600 ethanol wash was repeated. At this point I performed the Qiagen RNeasy Mini RNA
601 Cleanup protocol, including a DNase step. This protocol began with resuspending RNA
602 pellets in 100 µl RNase-free water and was followed to the end. Supplementing 10 µl β-
603 mercaptoethanol to 1 mL buffer RLT was taken to further inhibit RNases. Finally,

samples were eluted twice with 35 µl fresh RNase-free water, for a total elution volume of 70 µl for each sample and stored at -80 ºC. PCR was used to check for genomic DNA contamination, which was present, so a second DNase treatment was applied. Again, I used Qiagen DNase and an RNeasy Mini column. A modified version of a protocol found on the lab website of Christopher Mason, PhD was used (Zumbo, 2011). I eluted samples twice with 25 µl fresh RNase-free water, for a total elution volume of 50 µl for each sample. A small aliquot from each sample was taken for quality control and quantification and was stored at -80 ºC.

RNA QA/QC, quantification, sequencing, and data analysis

To check for genomic DNA contamination in RNA samples, I performed 25 rounds of PCR using *isiA* primers; there was no amplification in any of these RNA samples. These primers were chosen because they were available and were known to successfully amplify (Miller lab, unpublished data).

Fragment analysis of RNA was done on an Agilent Technologies TapeStation using an RNA ScreenTape, RIN values for the samples ranged from 6.4 – 8.2. RNA was quantified using a Qubit Fluorometer with the Broad Range RNA Assay Kit. High quality RNA was sent to the Washington State University, Spokane Genomics Core for library prep with TruSeq Stranded Total with Ribo-zero (Illumina) and 50-bp single read sequencing on a HiSeq-2500.

Illumina reads were received with adapters trimmed, analysis was performed using a Galaxy server maintained by the University of Montana (Afgan *et al.*, 2016). FASTQC (Andrews, 2010) was used to verify sequence quality. Both *Acaryochloris* species, CCMEE 5410 (NBIC taxon ID 310037; assembly GCA_000238775.2) and

MBIC 11017 (NBIC taxon ID 329726; assembly GCA_000018105.1) were mapped to their respective genome assemblies using Bowtie2 (Langmead and Salzberg, 2012). Resulting sorted BAM files were analyzed using the Cufflinks suite (Trapnell *et al.*, 2010). Mapped reads were assembled into transcripts with Cufflinks, using reference GFF annotations found on NCBI. Cuffmerge was used to merge all Cufflinks output of a single strain to create an annotation file used for differential expression analysis using Cuffdiff. Cuffdiff output was further analyzed using CummeRbund in R programming language (Trapnell *et al.*, 2012).

Results and discussion

Acaryochloris MBIC has higher fitness under low iron condition

Acaryochloris MBIC was isolated from an environment thought to be severely iron limited (Miyashita *et al.*, 1997; Kühl *et al.*, 2005; Boyd *et al.*, 2007; Endean, 1954). Conversely, there is no evidence suggesting iron limitation in the environment from which *Acaryochloris* CCMEE was isolated (Miller *et al.*, 2005). To test the hypothesis that MBIC is better adapted to low iron environments, I assayed generation times and final yield for both strains in media containing either low (7.7 nM) or high (51 μM) concentrations of iron. I expected the *Acaryochloris* strains to exhibit differential fitness with MBIC exhibiting a faster generation time and a greater final yield in both iron conditions.

A two-way ANOVA was carried out to determine the main effects of strain and iron condition on growth rate. Both strain [$F_{(1,8)} = 203.01$, $P < 0.0001$] and iron condition [$F_{(1,8)} = 347.52$, $P < 0.0001$] were significant predictors of growth rate; additionally, there was a significant interaction of the two variables [$F_{(1,8)} = 5.912$, $P = 0.04$]. MBIC grew

significantly faster than CCMEE under both iron replete (0.63 ± 0.003 vs. 0.45 ± 0.012 days per generation, $P < 0.0001$ by Tukey's HSD) and iron-limited conditions (0.41 ± 0.013 vs. 0.28 ± 0.012 days per generation, $P = 0.0001$). My observation that the main effect of iron condition was a significant predictor of generation time is not surprising, as the low iron condition was intended to induce iron starvation and therefore retard culture growth. The significant effect of strain is likely due to the difference in cell size between the strains. MBIC is smaller than CCMEE and therefore has a larger surface area to volume ratio (SA:V); SA:V is positively correlated with bacterial growth rate (Foy, 1980; Banse, 1976). However, the significant strain x iron interaction indicates that differences in growth rate cannot be solely explained by differences in cell size. MBIC responded more positively to higher iron availability, with growth rate increasing on average 0.23 generations/day between low and high iron conditions, compared with 0.17 generations/day for CCMEE (Fig. 2.2). Analysis of growth rate showed that MBIC has an even higher relative fitness advantage in high iron conditions compared with CCMEE, suggesting it may be better able to assimilate available iron.

CCMEE grew to a much lower final yield under iron limitation relative to MBIC, which suggests that CCMEE is significantly worse at scavenging iron (Fig 2.3). A two-way ANOVA was carried out to determine the main effects of strain and iron condition on final yield. Final yield was inverse transformed to conform to the assumptions of the model. Strain [$F_{(1,8)} = 24.08$, $P = 0.001$], iron condition [$F_{(1,8)} = 24.52$, $P = 0.001$], and their interaction [$F_{(1,8)} = 19.68$, $P = 0.002$] were all significant predictors of final yield. Post hoc analysis using Tukey HSD revealed that CCMEE in the iron-limited condition grew to a significantly lower yield than CCMEE in high iron condition and MBIC in both

673 iron conditions. Significant differences in final yield for CCMEE grown in the two iron
674 conditions, along with the absence of these differences in final yield of MBIC support the
675 hypothesis that MBIC is better adapted to low iron conditions. As with the differences in
676 growth rate, a significant interaction effect between strain and iron condition indicate
677 strain-dependent physiological responses.

678 *Acaryochloris strains differ in physiology of iron assimilation*

679 In order to assess differences between *Acaryochloris* strains in recovery from iron
680 starvation, an iron step-up experiment was performed. Cultures of both *Acaryochloris*
681 strains were grown to stationary phase in low iron media and then maintained for seven
682 additional days to ensure iron starvation. A final concentration of 51 µM Fe (III) was
683 added to culture flasks and I monitored the early stages of recovery from iron starvation
684 by measuring cell density (approximated by OD_{750}), rate of iron assimilation, and chl *d*
685 concentration.

686 Measurements were taken immediately prior to iron supplementation (t0) as well
687 as 12, 24, and 36 hours after iron addition (t12, t24, t36). I expected MBIC to exhibit
688 physiological changes indicative of recovery from iron starvation after the nutrient was
689 supplemented. I predicted that MBIC will assimilate iron and synthesize chl *d* more
690 quickly than CCMEE as it begins recovery and prepares to resume growth.

691 There was a significant increase in intracellular iron content and chl *d* (µg/volume
692 one cell) for both strains over the time series. MBIC assimilated iron significantly more
693 rapidly than CCMEE during the recovery period (Fig. 2.4A). According to two-way
694 ANOVA, while time after iron addition and strain were both significant predictors of
695 intracellular iron levels ([$F_{(1,36)}$ =61.40, $P < 0.0001$] and [$F_{(1,36)}$ =38.17, $P < 0.0001$],

696 respectively), there was also a significant strain x time after iron addition interaction
697 [$F_{(1,36)}$ =13.89, P= 0.0007]. This points to physiological differences between
698 *Acaryochloris* strains in their ability to assimilate iron during recovery from starvation.
699 Chlorophyll *d* content was predicted by main effects of strain [$F_{(1,36)}$ = 161.17, P <
700 0.0001], and time after iron addition [$F_{(1,36)}$ = 11.6, P = 0.002], however there was no
701 significant interaction effect of the two predictor variables [$F_{(1,36)}$ = 0.11, P = 0.74].
702 MBIC produced more chl *d* than CCMEE at all time points, but both strains increased
703 production of the pigment at the same rate during recovery (Fig. 2.4B).
704 *Acaryochloris strains differ in iron assimilation gene dosage*
705 The MBIC genome has a greater number of genes involved in iron assimilation
706 compared with the CCMEE genome as a consequence of both gene duplication and the
707 acquisition of novel gene content by horizontal transfer (Miller *et al.*, 2011). RNA-seq
708 was performed under a variety of conditions to determine if differences in fitness and iron
709 recovery observed between strains are correlated with increased dosage of iron
710 assimilation gene transcripts in MBIC. Iron assimilation genes broadly fall into three
711 categories based on known mechanisms of bacterial iron acquisition: siderophore
712 synthesizing genes, siderophore transporters, and fur transcription regulators. A brief
713 summary of bacterial iron assimilation is discussed below.
714 Bacteria assimilate iron from the extra-cellular environment through the use of
715 low molecular weight compounds called siderophores that have a high affinity for iron.
716 These compounds are synthesized by non-ribosomal peptide synthetases (NRPSs) and
717 polyketide synthetases (PKSs) (Kranzler *et al.*, 2013). They are then transported out of
718 the cell, a process mediated by ATP-binding cassette (ABC) superfamily proteins, where

719 they chelate iron and are finally shuttled back inside the cell, which is mediated by TonB-
720 dependent transporters (Kranzler *et al.*, 2013). Transcription of many siderophore
721 synthesis and transport genes are regulated by Fur transcriptional regulators, which
722 exhibit metal-dependent repression (Escolar *et al.*, 1999). When intracellular iron is high,
723 transcription of genes under Fur regulation is repressed and when intracellular iron is low
724 their transcription is derepressed (Andrews *et al.*, 2003).

725 Conditions tested in the RNA-seq experiment were iron starved cells in stationary
726 phase (t0 from the iron step-up experiment described above), cells beginning recovery
727 from iron starvation (t36 from the iron step-up experiment), and cells during active
728 growth under low iron (sw0). I predicted that the observed differences between strains in
729 fitness under iron limitation and in physiological response to recovery from iron
730 starvation will be associated with more transcripts of iron assimilation genes in MBIC
731 compared to CCMEE. To assess potential contributors to increased dosage, I divided
732 iron assimilation genes in MBIC into three groups. These groups were (1) single copy
733 genes present in both genomes, (2) genes with paralogs in MBIC and an ortholog in
734 CCMEE, and (3) genes that are novel to the MBIC genome. To compare gene expression
735 between strains I used normalized gene counts, generated by Cufflinks (Trapnell *et al.*,
736 2010).

737 First, I considered genes with a single copy in each *Acaryochloris* genome. If
738 these genes contribute to increased dosage of iron assimilation genes in MBIC, I would
739 expect them to be much more highly transcribed than their orthologs in CCMEE.
740 Seventeen genes fell into this category, ten of which exhibited greater estimated
741 expression in CCMEE in all three conditions. To focus on genes that were highly

differentially expressed between strains, I calculated log2-fold difference in expression of genes between strains for each growth condition. Six genes in this group had a log2-fold differences in expression between strains with an absolute value of at least 2 in one or more condition. Four highly differentially expressed genes were transporters, with three exhibiting greater expression in CCMEE in at least one condition and one exhibiting greater expression in MBIC in at least one condition. One siderophore synthesizing gene was highly differentially expressed, with more transcripts in CCMEE. Finally, one Fur transcriptional regulator was highly differentially expressed, with MBIC exhibiting a greater number of transcripts. Genes present as single copies in both strains do not appear to contribute to increased dosage in MBIC (Fig. 2.5).

The second group of genes I considered were genes with at least one duplicate, or paralogous copy, in the MBIC genome and an ortholog in the CCMEE genome. When genes are duplicated, dosage often increases because of the additional gene copy, which can be adaptive (Kondrashov, 2012). However, increased gene dosage as a result of gene duplication can have deleterious effects, a result of maladaptive stoichiometry (Hooper and Berg, 2003). In some cases paralog expression decreases, and ultimately there is no change in gene dosage pre and post duplication (Qian *et al.*, 2010). If genes in this group are contributing to increased dosage in MBIC, I would expect the total number of transcripts for paralogs to be greater than the number of transcripts for the single copy ortholog in CCMEE. Nine genes fell into this category, six of which had higher expression in CCMEE in at least one condition. However, when I focused on highly differentially expressed genes, as defined in the previous paragraph, there was only one with greater expression in CCMEE under at least one condition; this gene coded for

765 proteins involved in siderophore transport. Two Fur transcriptional regulators in this
766 group exhibited significantly greater expression in MBIC under at least one condition
767 (Fig. 2.6).

768 Finally, I considered novel gene content in MBIC. Twenty-five genes fell into
769 this category, three of which have paralogs in the genome. Included in this group was a
770 plasmid-encoded cluster of nine genes that is homologous to a known siderophore
771 producing gene cluster in the filamentous cyanobacterium *Anabaena* (Jeanjean *et al.*,
772 2008). These are likely the result of HGT (Fig. 2.1). While the other genes in this group
773 have not undergone phylogenetic analysis to determine their origin, up to eleven
774 additional genes in this group may be the result of HGT. This is due to their presence on
775 plasmids without a paralogous copy on the chromosome, novel gene content in
776 *Acaryochloris* genomes tends to cluster on plasmid DNA (Fig. 2.7; Miller *et al.*, 2011).
777 Certain genes in this group have much greater expression than their homologs in the other
778 two groups discussed. For example, a gene involved in siderophore transport found as a
779 single copy on a plasmid in MBIC is one of the most highly expressed genes overall in
780 the strain. Furthermore this novel transporter, exhibits nearly 3-fold greater expression
781 than the next most highly expressed siderophore transport gene in MBIC, which also
782 happens to be unique to that genome. Additionally, the cluster that shares homology with
783 the siderophore producing cluster in *Anabaena* accounts for the majority of transcripts
784 mapping to siderophore producing genes (Fig 2.8).

785 The final two groups, which include genes in MBIC with at least one paralog,
786 allow us to analyze differential expression levels for genes related by duplication. In
787 studies on the role of gene regulation in mammalian speciation, it has been shown that

gene regulatory elements can result in allele-specific expression in hybrids, specifically when an allele is in a novel regulatory background (Mack *et al.*, 2016). Paralogs in MBIC are effectively alleles of the same gene, and appear to exhibit allele-specific differences in expression. Paralogs most commonly had one gene located on the chromosome and its paralog(s) on plasmid(s). Of those genes with duplicates, four stood out as having differential expression levels between copies. Three of these genes, two involved in siderophore transport and a Fur transcriptional regulator, exhibited greater expression of the paralog located on the chromosome. One Fur transcriptional regulator exhibited greater expression of paralogs located on plasmids compared with the chromosomal copy (Fig. 2.6).

The iron step-up experiment afforded an opportunity to link differences in gene content among *Acaryochloris* strains with expression levels and ultimately fitness. Transcriptomic data were tightly associated with observed physiological differences between MBIC and CCMEE as they recovered from iron starvation. The observed faster rate of iron assimilation in MBIC after supplementation likely reflects enrichment for siderophore-producing and transport genes in this strain as compared to CCMEE (Fig. 2.8). Ultimately, this experiment shows that novel gene content in MBIC may account for the majority of observed increased gene dosage in iron assimilation genes between *Acaryochloris* strains. MBIC has both a large number of novel iron-assimilation genes, and greater dosage of iron-assimilation gene transcripts.

Concluding remarks

In this chapter of my thesis, I showed that *Acaryochloris* MBIC is adapted to low iron conditions. I associated this with increased ability to assimilate iron and ultimately

with positive gene dosage for iron assimilation genes in MBIC (Fig. 2.8). MBIC, which was isolated from an iron-poor environment, has increased copy number of genes involved in iron assimilation compared to CCMEE (Miller *et al.*, 2011). Results of experiments performed in this chapter show that MBIC exhibits greater fitness under low-iron conditions and is better able to recover after iron starvation. Furthermore, RNA-seq showed that MBIC is enriched in transcripts of novel iron assimilation genes compared to CCMEE, supporting the hypothesis that gene copy number variation can give rise to increased gene dosage and ultimately have a positive effect on fitness under certain conditions.

Figures

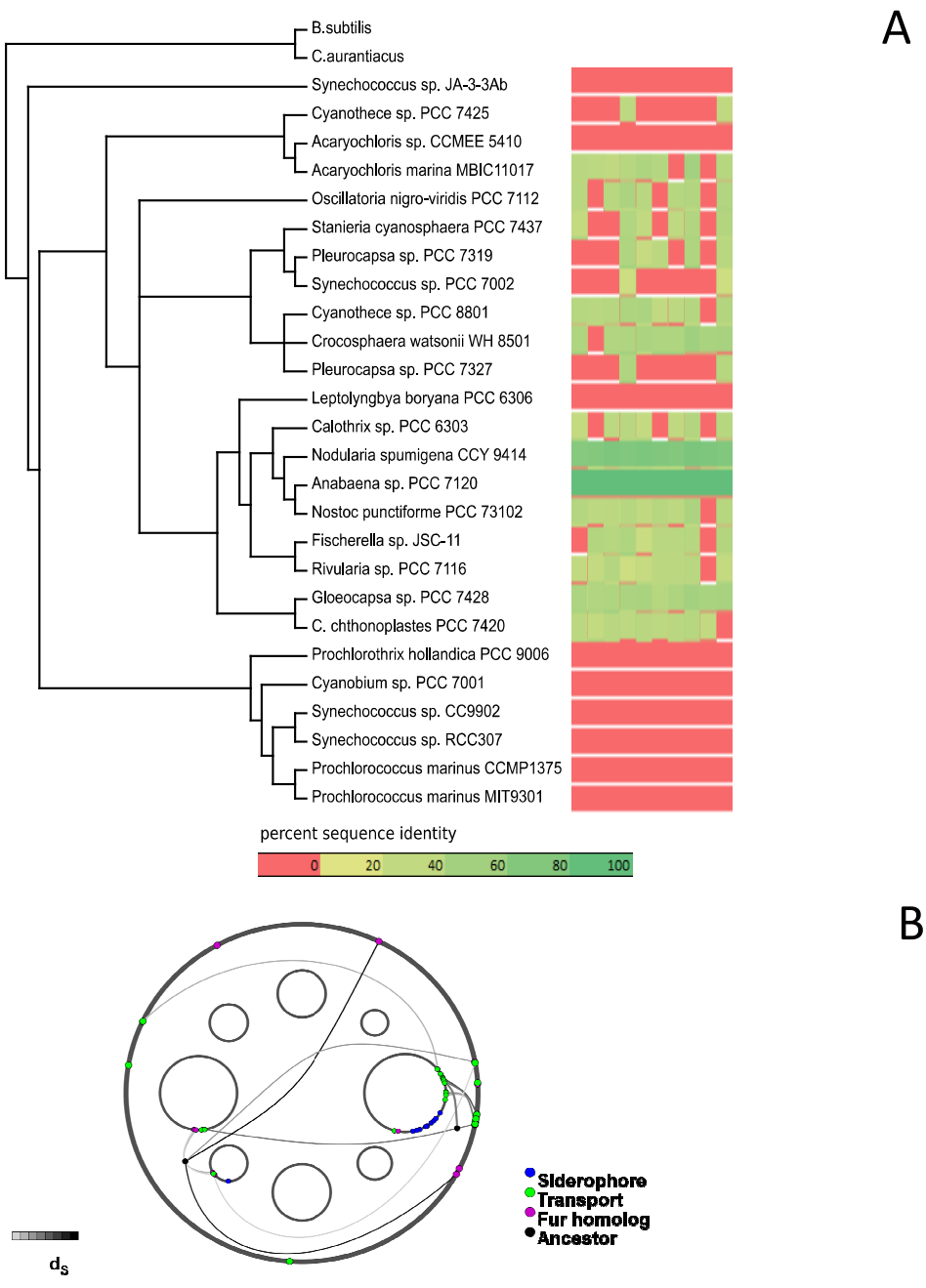


Fig. 2.1 (A) Bayesian phylogeny of cyanobacteria, built using 16s sequences showing presence or absence of genes with homology to 10 known siderophore producing genes in *Anabaena* sp. PCC 7120. (B) Representation of *Acaryochloris* MBIC genome, the large outer circle is the chromosome and smaller inner circles are plasmids. This shows the location of some iron assimilation genes with CNV, represented by the small colored circles. The large plasmid-born cluster of siderophore producing genes shares homology with the *Anabaena* sp. PCC 7120 siderophore cluster. Genes connected by lines indicate genes related by duplication

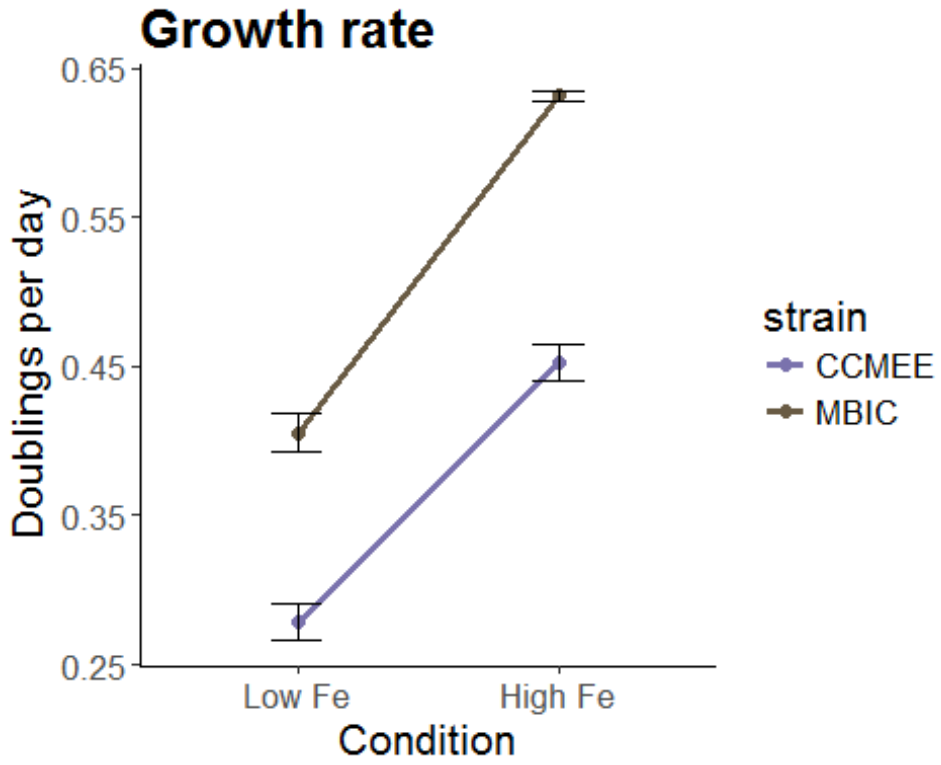


Fig. 2.2 Generations per day in *Acaryochloris* strains under low and high iron conditions, error bars indicating standard error. MBIC increased doubling time by 0.23 days between low and high iron conditions, CCMEE increased doubling time by 0.17 days between the two conditions.

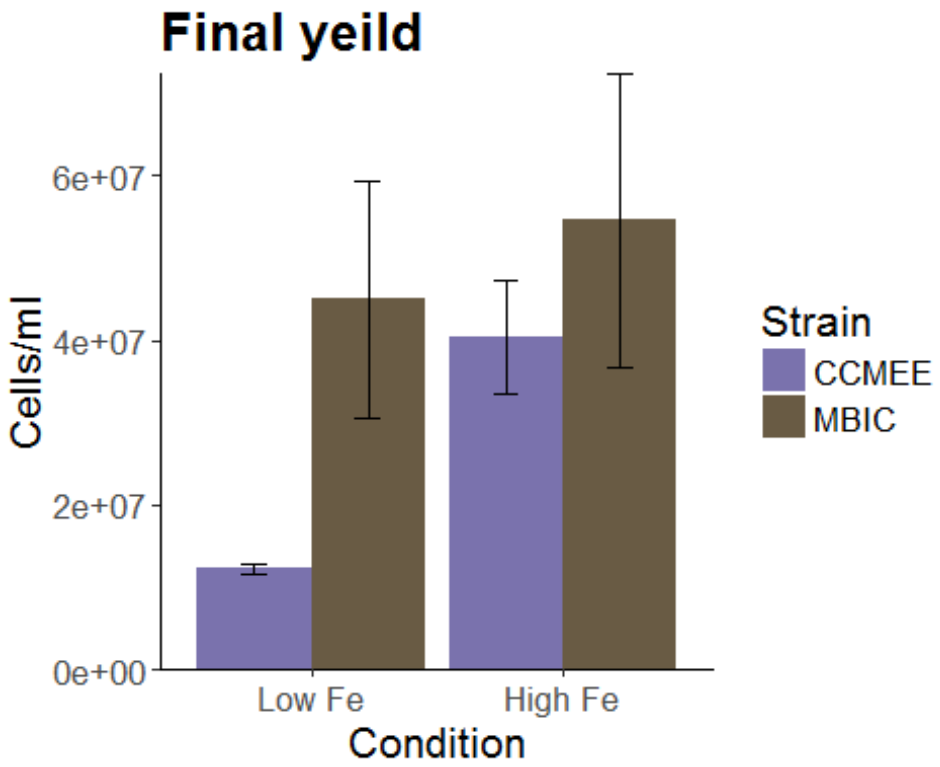


Fig. 2.3 Final yield in cells/ml for *Acaryochloris* strains grown in low and high iron conditions, error bars indicating standard error. Final yield of CCMEE in low iron is significantly lower than the final yield of MBIC in both conditions and CCMEE in high iron.

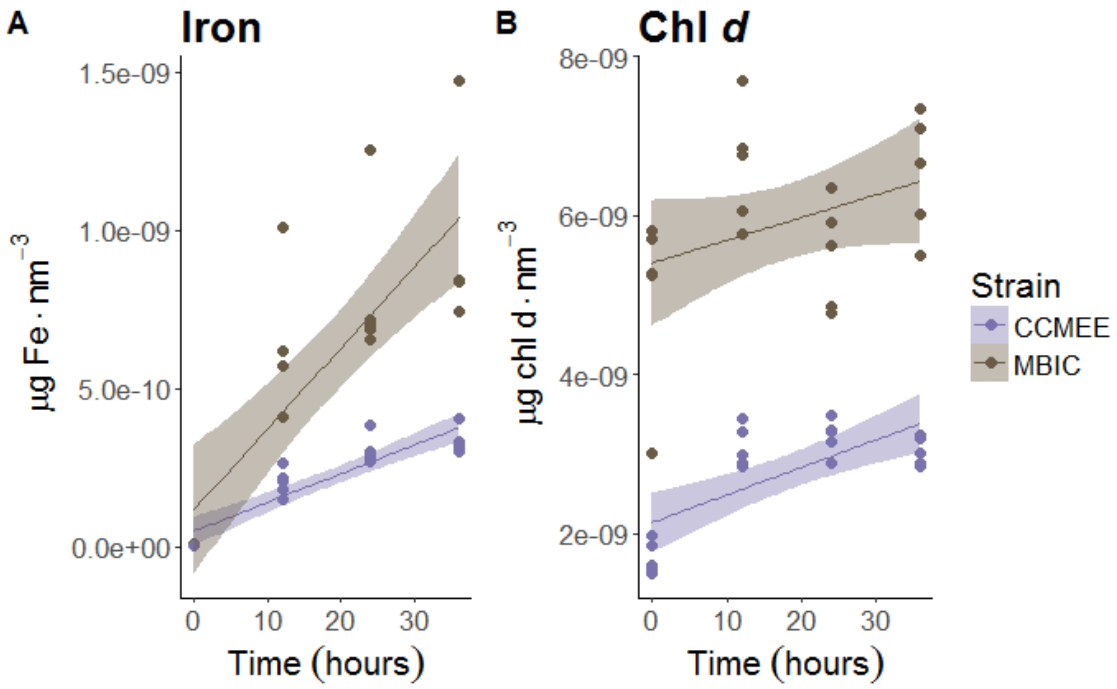


Fig. 2.4 (A) Intracellular iron content and (B) chlorophyll *d* content normalized to single cell volume after iron addition. Linear model with 95% CI shown. Rate at which iron is assimilated differs between MBIC and CCMEE.

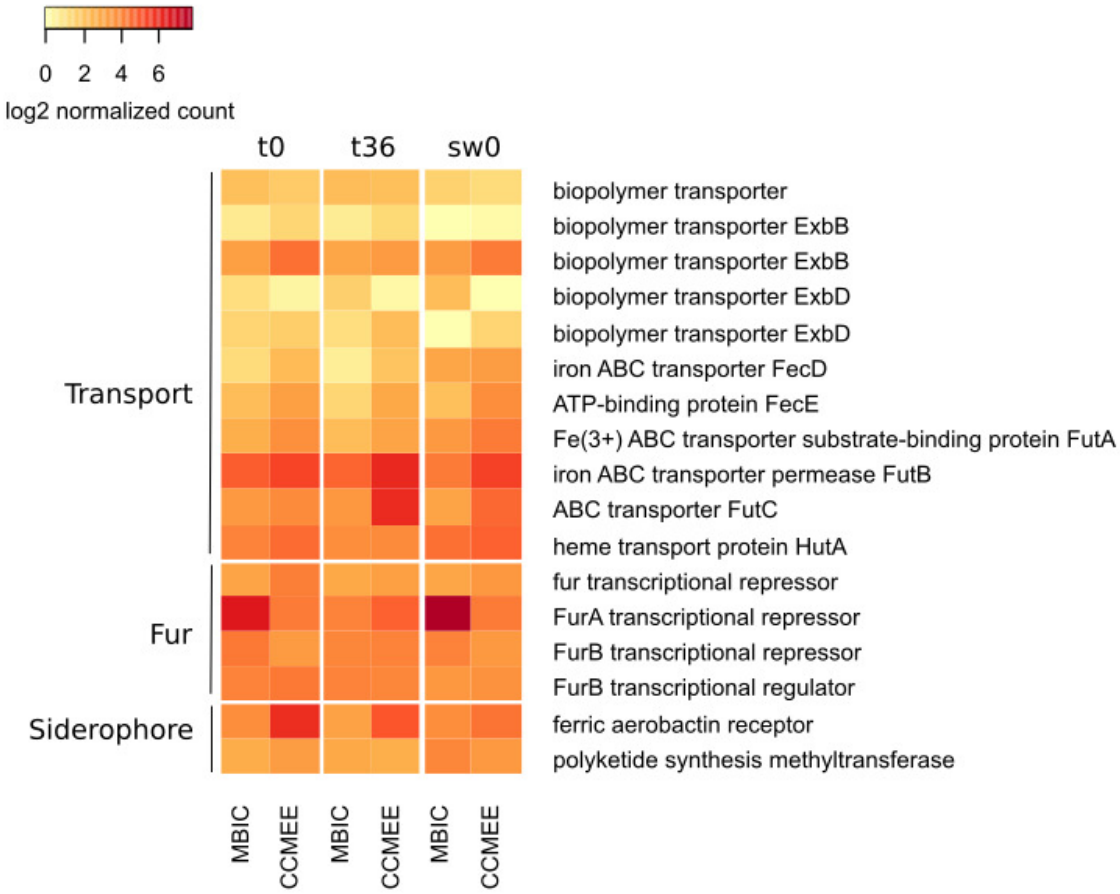


Fig. 2.5 Heatmap of log2 transformed normalized gene counts for genes present as a single copy in both MBIC and CCMEE genomes.

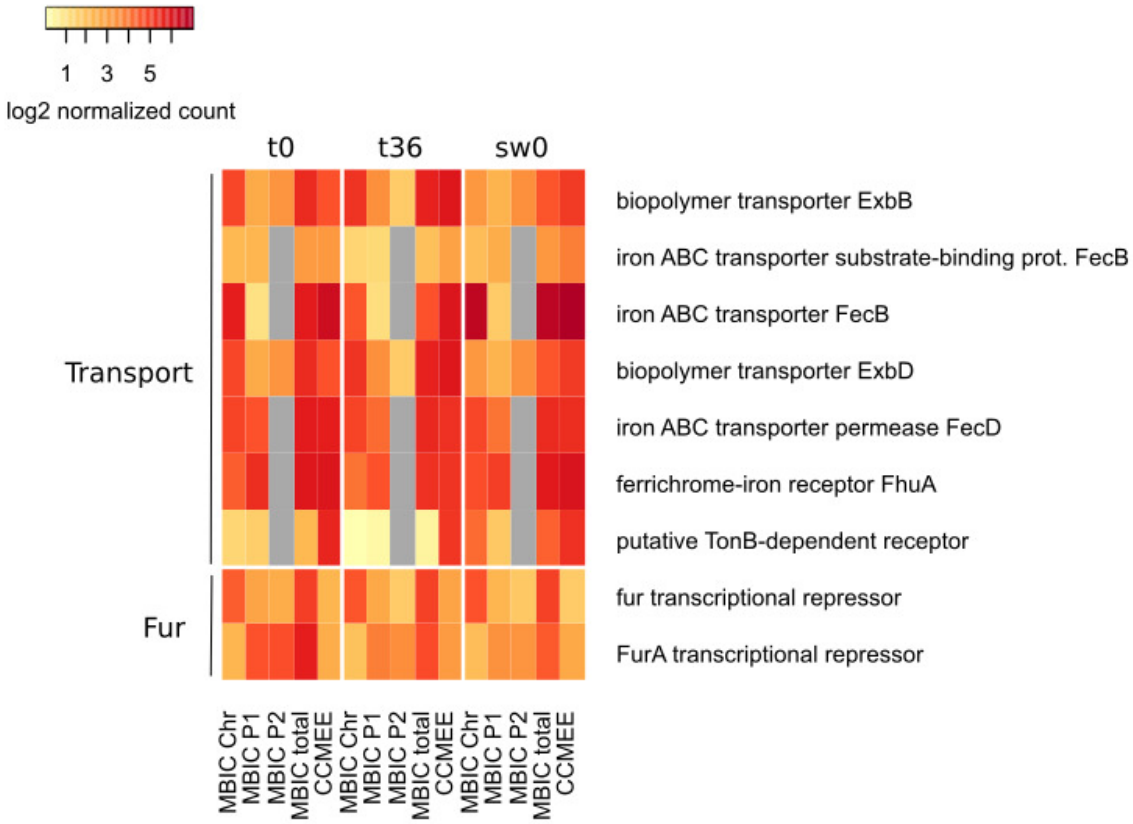


Fig. 2.6 Heatmap of log2 transformed normalized gene counts for genes with at least one paralog in MBIC and an ortholog in CCMEE. Within each group of columns representing a condition, MBIC Chr corresponds to genes located on the MBIC chromosome; MBIC P1 and MBIC P2 correspond to paralogs located on plasmids. Summed expression for MBIC paralogs corresponds to MBIC total, and expression of single copy ortholog in CCMEE is given in columns labeled CCMEE.

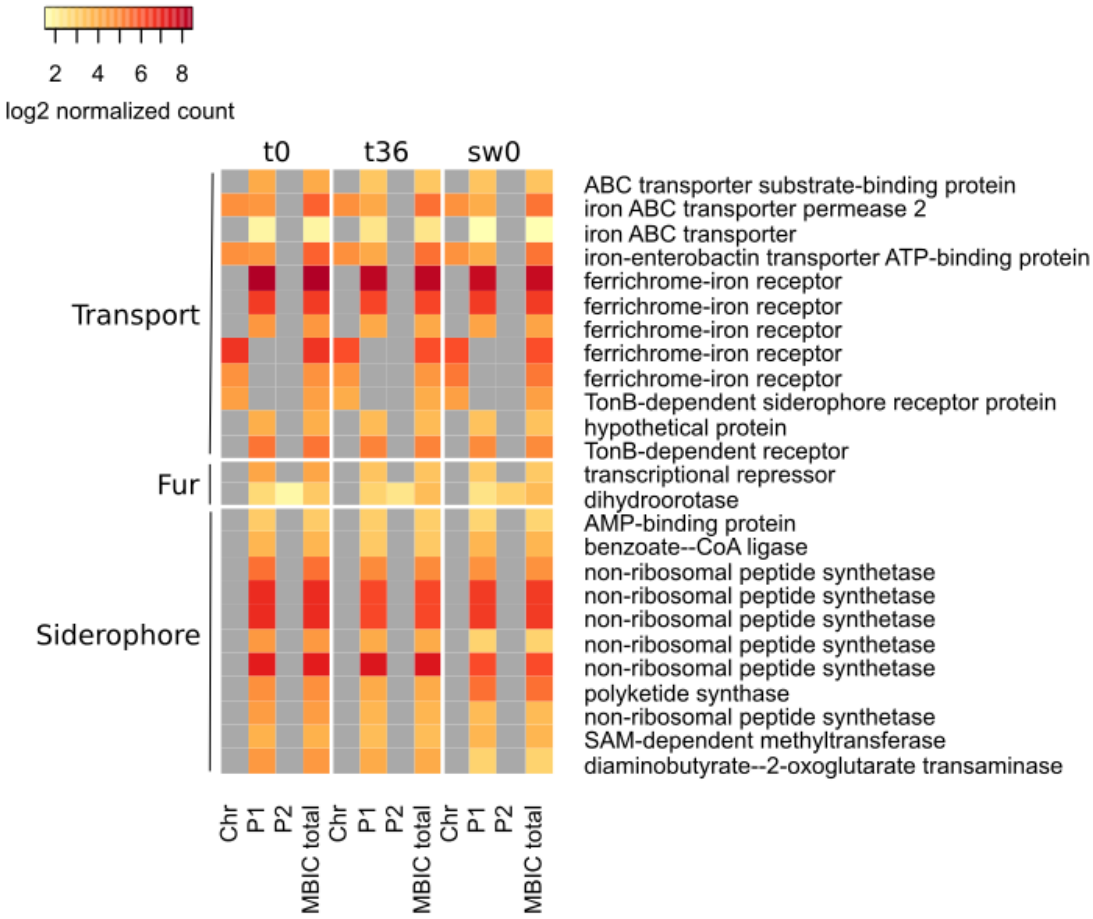
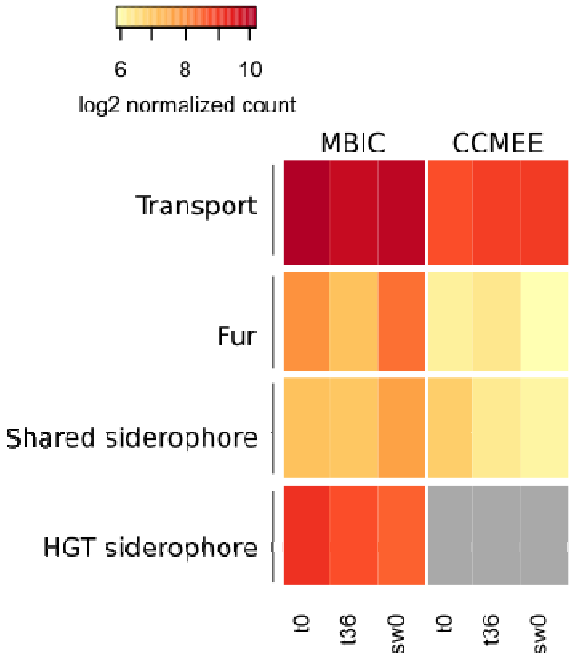


Fig. 2.7 Heatmap of log2 transformed normalized gene counts for genes novel to the MBIC genome. Within each group of columns representing a condition, Chr corresponds to genes located on the MBIC chromosome; P1 and P2 correspond to paralogs located on plasmids. Summed expression for MBIC paralogs corresponds to MBIC total.



865
866 **Fig. 2.8** For each strain and condition, normalized gene counts for all genes involved in a
867 given aspect of iron assimilation were added and log2 transformed. HGT siderophore
868 corresponds to genes in MBIC with evidence of horizontal origin (Fig. 2.1).
869

Literature Cited

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, *et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3–W10.

Andersson DI, Hughes D. (2009). Gene amplification and adaptive evolution in bacteria. *Annu Rev Genet* **43**: 167–195.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Andrews SC, Robinson AK, Rodríguez-Quiñones F. (2003). Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**: 215–237.

Banse K. (1976). RATES OF GROWTH, RESPIRATION AND PHOTOSYNTHESIS OF UNICELLULAR ALGAE AS RELATED TO CELL SIZE?A REVIEW [2]. *J Phycol* **12**: 135–140.

Bergthorsson U, Andersson DI, Roth JR. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci U S A* **104**: 17004–17009.

Bianco PR, Kowalczykowski SC. (1998). RecA protein. In: Vol. 32. *Encyclopedia of Life Sciences*. pp 69–76.

Boyd PW, Jickells T, Law CS, Blain S, Boyle E a, Buesseler KO, *et al.* (2007). Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science* **315**: 612–617.

Brown CJ, Todd KM, Rosenzweig RF. (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol Biol Evol* **15**: 931–942.

Cairns J, Foster PL. (1991). Adaptive Reversion of a Frameshift Mutation in Escherichia coli. *Genetics* **128**: 695–701.

Capaldo FN, Ramsey G, Barbour SD. (1974). Analysis of the growth of recombination-deficient strains of Escherichia coli K-12. *J Bacteriol* **118**: 242–9.

Cox JM, Li H, Wood EA, Chitteni-Pattu S, Inman RB, Cox MM. (2008). Defective dissociation of a "slow" RecA mutant protein imparts an Escherichia coli growth defect. *J Biol Chem* **283**: 24909–21.

Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ, Marians KJ. (2000). The importance of repairing stalled replication forks. *Nature* **404**: 37–41.

Cox MM, Lehman IR. (1982). recA Protein-promoted DNA Strand Exchange. **257**: 8523–8532.

Cox MM, Lehman IR. (1981). recA protein of Escherichia coli promotes branch migration, a kinetically distinct phase of DNA strand exchange. *Proc Natl Acad Sci U S A* **78**: 3433–3437.

Dimpfl J, Echols H. (1989). Duplication Mutation as an SOS Response in Escherichia coli: Enhanced Duplication Formation by a Constitutively Activated RecA. *Genetics* 255–260.

Domain F, Houot L, Chauvat F, Cassier-Chauvat C. (2004). 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* **53**: 65–80.

Endean R. (1954). STUDIES OF THE BLOOD AND TESTS O F SOME AUSTRALIAN ASCIDIANS. *Mar Freshw Res* **6**: 157–164.

915 Environmental Protection Agency U. EPA Method 3050B (SW-846): Acid Digestion of
916 Sediments, Sludges, and Soils.
917 Escolar L, Pérez-Martín J, De Lorenzo V. (1999). Opening the iron box: Transcriptional
918 metalloregulation by the fur protein. *J Bacteriol* **181**: 6223–6229.
919 Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J. (1999). Preservation
920 of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**: 1531–
921 1545.
922 Foy RH. (1980). THE INFLUENCE OF SURFACE TO VOLUME RATIO ON THE
923 GROWTH RATES OF PLANKTONIC BLUE-GREEN ALGAE. *Br phycol J* **15**: 279–
924 289.
925 de Groot A, Dulermo R, Ortet P, Blanchard L, Guérin P, Fernandez B, *et al.* (2009).
926 Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium
927 Deinococcus deserti. *PLoS Genet* **5**: e1000434.
928 Gruber AJ, Erdem AL, Sabat G, Karata K, Jaszczur MM, Vo DD, *et al.* (2015). A RecA
929 Protein Surface Required for Activation of DNA Polymerase V. *PLOS Genet* **11**:
930 e1005066.
931 Hahn MW. (2009). Distinguishing among evolutionary models for the maintenance of
932 gene duplicates. *J Hered* **100**: 605–617.
933 Haldenby S, White MF, Allers T. (2009). RecA family proteins in archaea: RadA and its
934 cousins. *Biochem Soc Trans* **37**: 102–107.
935 Harmon FG, Rehrauer WM, Kowalczykowski SC. (1996). Interaction of Escherichia coli
936 RecA Protein with LexA Repressor. *J Biol Chem* **271**: 23865–23873.
937 Holthausen JT, Wyman C, Kanaar R. (2010). Regulation of DNA strand exchange in
938 homologous recombination. *DNA Repair (Amst)* **9**: 1264–1272.
939 Hooper SD, Berg OG. (2003). On the nature of gene innovation: Duplication patterns in
940 microbial genomes. *Mol Biol Evol* **20**: 945–954.
941 Hughes AL. (1994). The evolution of functionally novel proteins after gene duplication.
942 *R Soc* **256**: 119–124.
943 Indiani C, Patel M, Goodman MF, O'Donnell ME. (2013). RecA acts as a switch to
944 regulate polymerase occupancy in a moving replication fork. *Proc Natl Acad Sci U S A*
945 **110**: 5410–5415.
946 Janion C. (2008). Inducible SOS Response System of DNA Repair and Mutagenesis in
947 Escherichia coli. *Int J Biol Sci Int J Biol Sci*.
948 Jeanjean R, Talla E, Latifi A, Havaux M, Janicki A, Zhang CC. (2008). A large gene
949 cluster encoding peptide synthetases and polyketide synthases is involved in production
950 of siderophores and oxidative stress response in the cyanobacterium Anabaena sp. strain
951 PCC 7120. *Environ Microbiol* **10**: 2574–2585.
952 Jiang Q, Karata K, Woodgate R, Cox MM, Goodman MF. (2009). The active form of
953 DNA polymerase V is UmuD'(2)C-RecA-ATP. *Nature* **460**: 359–363.
954 Kawecki TJ, Ebert D. (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–
955 1241.
956 Klein HL. (2008). The consequences of Rad51 overexpression for normal and tumor
957 cells. *DNA Repair (Amst)* **7**: 686–693.
958 Kondrashov FA. (2012). Gene duplication as a mechanism of genomic adaptation to a
959 changing environment. *Proc R Soc B Biol Sci* 5048–5057.
960 Kraemer SA, Boynton PJ. (2017). Evidence for microbial local adaptation in nature. *Mol*

961 *Ecol* **26**: 1860–1876.
962 Kranzler C, Rudolf M, Keren N, Schleiff E. (2013). Iron in Cyanobacteria.
963 Krejci L, Altmannova V, Spirek M, Zhao X. (2012). Homologous recombination and its
964 regulation. *Nucleic Acids Res* **40**: 5795–5818.
965 Kühl M, Chen M, Ralph PJ, Schreiber U, Larkum AWD. (2005). Ecology: A niche for
966 cyanobacteria containing chlorophyll d. *Nature* **433**: 820.
967 Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat*
968 *Methods* **9**: 357–359.
969 Lawrence JG, Ochman H. (1998). Molecular Archaeology of the Escherichia coli
970 Genome. *Source Proc Natl Acad Sci United States Am* **95**: 9413–9417.
971 Li Y, Scales N, Blankenship RE, Willows RD, Chen M. (2012). Extinction coefficient for
972 red-shifted chlorophylls: Chlorophyll d and chlorophyll f. *Biochim Biophys Acta -*
973 *Bioenerg* **1817**: 1292–1298.
974 Little JW. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of
975 RecA coprotease. *Biochimie* **73**: 411–422.
976 Lusetti SL, Cox MM. (2002). The Bacterial RecA Protein and the Recombinational DNA
977 Repair of Stalled Replication Forks. *Annu Rev Biochem* **71**: 71–100.
978 Lynch M, Force A. (2000). The probability of duplicate gene preservation by
979 subfunctionalization. *Genetics* **154**: 459–473.
980 Lynch M, Katju V. (2004). The altered evolutionary trajectories of gene duplicates.
981 *Trends Genet* **20**: 544–549.
982 Mack KL, Campbell P, Nachman MW. (2016). Gene regulation and speciation in house
983 mice. *Genome Res* **26**: 451–61.
984 Des Marais DL, Rausher MD. (2008). Escape from adaptive conflict after duplication in
985 an anthocyanin pathway gene. *Nature* **454**: 762–765.
986 McGrew D a, Knight KL. (2003). Molecular design and functional organization of the
987 RecA protein. *Crit Rev Biochem Mol Biol* **38**: 385–432.
988 Meeks JC. RNA isolation from Nostoc punctiforme.
989 http://microbiology.ucdavis.edu/meeks/xpro7a.htm.
990 Miller SR, Augustine S, Olson T Le, Blankenship RE, Selker J, Wood AM. (2005).
991 Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid
992 proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc Natl Acad Sci U S A* **102**:
993 850–5.
994 Miller SR, Wood a M, Blankenship RE, Kim M, Ferriera S. (2011). Dynamics of gene
995 duplication in the genomes of chlorophyll d-producing cyanobacteria: implications for
996 the ecological niche. *Genome Biol Evol* **3**: 601–13.
997 Miyashita H, Adachi K, Kurano N, Ikemoto H, Chihara M, Miyachi S. (1997). Pigment
998 composition of a novel oxygenic photosynthetic prokaryote containing chlorophyll d as
999 the major chlorophyll. *Plant Cell Physiol* **38**: 274–281.
1000 Miyashita H, Ikemoto H, Kurano N. (1996). Chlorophyll d as a major pigment. *Nature*
1001 **383**: 402.
1002 Monod J. (1949). The growth of bacterial cultures. *Annu Rev Microbiol* **3**: 371–394.
1003 Moran NA, McCutcheon JP, Nakabachi A. (2008). Genomics and evolution of heritable
1004 bacterial symbionts. *Ann Rev Genet* **42**: 165–190.
1005 Murphy RC, Bryant DA, Porter RD, de Marsac NT. (1987). Molecular cloning and
1006 characterization of the recA gene from the cyanobacterium Synechococcus sp. strain PCC

7002. *J Bacteriol* **169**: 2739–47.
Nahrstedt H, Schröder C, Meinhardt F, Schro C, Meinhardt F. (2005). Evidence for two recA genes mediating DNA repair in Bacillus megaterium. *Microbiology* **151**: 775–87.
Nasvall J, Sun L, Roth JR, Andersson DI. (2012). Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science (80-)* **338**: 384–387.
Nohmi T, Battista JR, Dodson L a, Walker GC. (1988). RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proc Natl Acad Sci U S A* **85**: 1816–1820.
Norioka N, Hsu MY, Inouye S, Inouye M. (1995). Two recA genes in Myxococcus xanthus. *J Bacteriol* **177**: 4179–4182.
Ochman H, Lawrence JG, Groisman E a. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
Ohno S. (1970). Evolution by Gene Duplication. Springer Berlin Heidelberg: Berlin, Heidelberg.
Owttrim GW, Coleman JR. (1987). Molecular cloning of a recA-like gene from the cyanobacterium Anabaena variabilis. *J Bacteriol* **169**: 1824–9.
Patel M, Jiang Q, Woodgate R, Cox MM, Goodman MF. (2010). A New Model for SOS-induced Mutagenesis: How RecA Protein Activates DNA Polymerase V. *Crit Rev Biochem Mol Biol* **45**: 171–184.
Pinto FL, Thapper A, Sontheim W, Lindblad P. (2009). Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Mol Biol* **10**: 79.
Qian W, Liao B-Y, Chang AY-F, Zhang J. (2010). Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet* **26**: 425–30.
Rapa RA, Islam A, Monahan LG, Mutreja A, Thomson N, Charles IG, *et al.* (2015). A genomic island integrated into recA of Vibrio cholerae contains a divergent recA and provides multi-pathway protection from DNA damage. *Environ Microbiol* **17**: 1090–1102.
Reams AB, Neidle EL. (2003). Genome plasticity in Acinetobacter: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol* **47**: 1291–1304.
Riehle MM, Bennett AF, Long AD. (2001). Genetic Architecture of Thermal Adaptation in Escherichia coli. *Source Proc Natl Acad Sci United States Am* **98**: 525–530.
Sano E, Miller S. No Title.
Schlacher K, Goodman MF. (2007). Lessons from 50 years of SOS DNA-damage-induced mutagenesis. *Nat Rev Mol Cell Biol* **8**: 587–594.
Schliep M, Crossett B, Willows RD, Chen M. (2010). 18O Labeling of Chlorophyll d in Acaryochloris marina Reveals That Chlorophyll a and Molecular Oxygen Are Precursors. *J Biol Chem* **285**: 28450–28456.
Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, *et al.* (2013). Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote. *Science (80-)* **339**: 1207–1210.
Sprouffske K, Wagner A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**. e-pub ahead of print, doi: 10.1186/s12859-016-1016-7.
Summons RE, Jahnke LL, Hope JM, Logan GA. (1999). 2-Methylhopanoids as

biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**: 554–557.
Swingley WD, Chen M, Cheung PC, Conrad AL, Dejesa LC, Hao J, *et al.* (2008). Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium Acaryochloris marina. *Proc Natl Acad Sci U S A* **105**: 2005–10.
Swingley WD, Hohmann-Marriott MF, Le Olson T, Blankenship RE. (2005). Effect of iron on growth and ultrastructure of Acaryochloris marina. *Appl Environ Microbiol* **71**: 8606–8610.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, *et al.* (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–578.
Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, *et al.* (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**. e-pub ahead of print, doi: 10.1038/nbt.1621.
Zumbo P. (2011). Purify Total RNA (≤ 45 µg) with DNase Treatment. http://physiology.med.cornell.edu/faculty/mason/lab/zumbo/files/ZUMBO_rna_cleanup.pdf (Accessed November 21, 2016).