Integrative Genomics Approach to Identify Genes Important for H_2 Production by *Rhodopseudomonas*

Somsak Phattarasukol

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Caroline S. Harwood, Chair

Roger E. Bumgarner

John E. Mittler

Program Authorized to Offer Degree:

Microbiology

University of Washington

Abstract

Integrative Genomics Approach to Identify Genes Important for H_2 Production by *Rhodopseudomonas*

Somsak Phattarasukol

Chair of the Supervisory Committee:

Professor Caroline S. Harwood

Microbiology

Hydrogen gas (H_2) is a clean-burning fuel and energy source that can be produced biologically by bacteria. Photosynthetic bacteria are especially promising as biocatalysts for H_2 production because energy from light can be used to drive this thermodynamically difficult process. The photosynthetic bacterium *Rhodopseudomonas* draws on the functioning of three major metabolic modules to produce H_2. These are photophosphorylation to generate ATP from light, carbon compound catabolism to generate electrons, and nitrogenase, an enzyme that combines electrons from carbon compounds with protons from water to generate H_2 by an ATP-intensive process. Each of metabolic modules is complicated and when considered together it is clear that H_2 production requires the integration of dozens of metabolic reactions. For example, acquisition of metals needed for synthesis of a functional nitrogenase is not easily identified as being associated with H_2 production. To identify a full set of genes from *Rhodopseudomonas* involved in H_2 production, I compared the genomes and transcriptomes of 16 closely related

strains of *Rhodopseudomonas*. In addition, I constructed co-expression networks of a set of 7 other less closely related *Rhodopseudomonas* strains and correlated gene expression modules with nitrogenase activities and H_2 production yields of the strains. I identified a set of 54 genes that were highly expressed in all 16 closely related strains grown under the H_2-producing condition in high light. These are candidates for genes that contribute to H_2 production. The co-expression analysis suggested that expression of nitrogenase genes is an essential but not a limiting factor for H_2 production. In contrast, the expression of light-harvesting genes appeared to be important for H_2 production under high and low light. In the process of generating data for this thesis, I developed an easy-to-use integrated tool for processing RNA-seq data, called Xpression. This tool is well suited to analyze gene expression data generated from bacteria and Achaea and should be useful to research laboratories that do not routinely carry out gene expression studies.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor Caroline Harwood for her support, patience and generosity. In the past seven years, I have learned so much from her scientifically, professionally and personally. Her encouragement and guidance have been instrumental in making me a smarter scientist and a better person. I would also like to acknowledge my committee members: John Leigh, Roger Bumgarner and John Mittler for their invaluable discussions and guidance in my research work. I would next like to thank Eric Schadt, John Castle, Bin Zhang, and Jun Zhu for sharing many tips and techniques for analyzing complex data.

I would like to especially thank Yasuhiro Oda and Colin Lappala for their instrumental helps in generating data for my research, and Amy Schaefer and Kathryn Fixen for numerous life-saving advices and for being great friends to me. I would also like to thank Kieran Pechter, Claudine Baraquet, Sandy Thao, Sudha Chugani, James McKinlay, Jennifer O'Connor, Jason Hickman, Jean Huang and Tuzun Güverner for encouragement, and my fellow graduate students, Erin Heiniger and Varisa Huangyutitham, for friendship and support.

I would like to personally thank my wife Yada Chaiyabutr and my parents Weera and Wannalee Chausiriphattana for unconditional loves and supports, and for believing in me. It had been a long and winding journey and I would not have made it without all of you.

CHAPTER 1

Introduction

INTRODUCTION

Hydrogen Gas - a Clean Fuel for the Future

Hydrogen gas (H_2) is a promising alternative fuel for the future. It is a clean-burning energy that yields only water after combustion, and can be converted to electricity in hydrogen fuel cells for powering electric vehicles. Currently, H_2 is mostly produced via industrialized methods such as steam reformation of natural gas, petroleum refining, and coal gasification (Rupprecht *et al.*, 2006), which are all energy intensive and emit considerable amounts of greenhouse gas to the environment. In contrast, biological H_2 production is much more environmentally friendly because it is produced at much lower temperatures and pressures. In addition, the process can be coupled with treatment of waste materials from the agricultural and food industries (Kapdan and Kargi, 2006), and thus provides a ideal solution to environmental problems of both meeting increasing energy needs and waste management.

Biological Production of H_2

The fundamental process of biological H_2 production is carried out with H_2-producing enzymes, which catalyze the reduction of protons to H_2. At present, two enzymes carrying out this reaction are known: hydrogenase and nitrogenase (Manish and Banerjee, 2008). While nitrogenase produces H_2 as an obligatory aspect of the catalytic cycle of the enzyme, hydrogenase produces H_2 directly as a result of proton reduction (McKinlay and Harwood, 2010a). Both nitrogenases and hydrogenases produce H_2 only in the context of living cells. Thus researchers have been investigating three major types of microorganisms for biological H_2 production: 1) anaerobic fermentative bacteria, which generate electrons from organic compounds and produce H_2 as a fermentation end product; 2) cyanobacteria and green algae, which generate electrons and produce H_2 from water via photophosphoryation; and 3) anoxygenic photosynthetic bacteria, which generate electrons from organic compounds and generate energy from light to drive H_2 production.

Rhodopseudomonas as a Platform for H_2 Production

Rhodopseudomonas is a photosynthetic bacterium that is widely distributed in natural environments (Larimer *et al.*, 2004). It is metabolically versatile, capable of growing chemotrophically as well as phototrophically, and produces large amounts of H_2 using nitrogenases when converting nitrogen gas to ammonia. It is considered an ideal platform for H_2 production for a number of reasons. First, *Rhodopseudomonas* is capable of deriving electrons from cheap and abundant compounds such as lignin monomers (the second most abundant polymer on earth) (Larimer *et al.*, 2004) and thiosulfate (a byproduct of several industrial processes) for H_2 production (Huang *et al.*, 2010). Second, unlike fermentative bacteria, *Rhodopseudomonas* generates energy from light to drive energy-intensive reactions, and thus can potentially use all of the electrons derived from an electron-donating substrate to H_2 production (Harwood, 2008). Third, unlike cyanobacteria and algae, *Rhodopseudomonas* catalyzes H_2 production under anaerobic conditions and does not need to protect nitrogenase from inactivation by oxygen. Lastly, *Rhodopseudomonas* can be genetically engineered to produce H_2 exclusively (Rey *et al.*, 2007) and continuously for four months (Gosse *et al.*, 2010).

Molecular Mechanism of H_2 Production

When grown anaerobically in light and deprived of fixed nitrogen from the environment, *Rhodopseudomonas* uses nitrogenase to reduce N_2 to ammonia and produces H_2 as an obligate byproduct (Dixon and Kahn, 2004). Nitrogenase is a complex metalloenzyme containing two components, which are named to reflect the metal composition they contain (Seefeldt *et al.*, 2009). The smaller component is known as the Fe protein (dinitrogenase reductase, encoded by *nifH*) while the larger component is known as the MoFe protein (dinitrogenase, encoded by *nifDK*). The Fe protein is a homodimer responsible for transferring electrons (one at a time) to the MoFe protein via a reaction that is dependent on ATP hydrolysis (two ATPs per the transfer of one electron). In contrast, the MoFe protein is a heterotetrameric protein containing two

metal centers called the FeMo cofactor and the P cluster, which is the substrate reduction site and the component involved in internal electron transfer, respectively. The process of N_2 reduction occurs as follows. First, an electron transfer protein such as ferredoxin and flavodoxin reduces the Fe protein. Second, a single electron is transferred from the Fe protein to the MoFe protein. Lastly, the electron is transferred internally in the MoFe protein by the P cluster to the FeMo cofactor, two protons are reduced to H_2, and subsequently N_2 is reduced to ammonia (Dixon and Kahn, 2004). The chemical equation of N_2 reduction (H_2 production) is as follows.

$$N_2 + 8e^- + 8H^+ + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i$$

Note that even in the absence of N_2, nitrogenase is capable of reducing protons to hydrogen as shown below (McKinlay and Harwood, 2010a).

$$8e^- + 8H^+ + 16ATP \rightarrow 4H_2 + 16ADP + 16P_i$$

Nitrogenase Synthesis and Regulation

Not only is nitrogenase an expensive enzyme to operate, the synthesis and assembly of a mature nitrogenase is complicated (Rubio and Ludden, 2005; Seefeldt *et al.*, 2009). As a result, the synthesis and activity of nitrogenase are tightly repressed if fixed nitrogen compounds, such as ammonia, are available (Dixon and Kahn, 2004). *Rhodopseudomonas* uses a number of transcriptional mechanisms to regulate nitrogenase synthesis as explained below.

- *PII Proteins*. *Rhodopseudomonas* senses ammonium availability using PII signal transduction proteins (encoded by *glnB*, *glnK1* and *glnK2*) (Connelly *et al.*, 2006). When the intracellular level of α-ketoglutarate is high (a signal of nitrogen deficiency) UTase/UR (encoded by *glnD*) uridylylates the PII proteins, which alters conformation of the proteins and changes their ability to interact with target proteins such as the NtrB protein (see below). Conversely, when the intracellular level of glutamine is high (a sign of nitrogen sufficiency), glutamine binds to the UTase/UR and causes the enzyme to

switch activity. As a result, the PII proteins are de-uridylylated and that allows them to properly interact with their targets (Leigh and Dodsworth, 2007; Merrick and Edwards, 1995).

- *NtrBC System*. One of the binding targets for PII proteins is NtrBC. It is a the two-component regulator that is involved in the activation of nitrogen starvation genes in many proteobacteria (Merrick and Edwards, 1995). When the intracellular level of glutamine is high (a sign of nitrogen sufficiency), the de-uridylylated form of PII proteins binds to the NtrB protein and stimulates dephosphorylation of NtrC. When the intracellular level of α-ketoglutarate is high (a signal of nitrogen deficiency), the uridylylated form of the PII proteins does not bind to the NtrB protein. That allows the NtrB to phosphorylate and activate the NtrC protein, which in turn facilitates the binding of σ54-RNA polymerase to the promoters of genes such as *nifA* (see below) (Dixon and Kahn, 2004; Merrick and Edwards, 1995).

- *NifA Protein*. The master transcriptional regulator of genes required for nitrogenase synthesis is the NifA protein (Dixon and Kahn, 2004). It is an enhancer DNA binding protein with three major domains: a GAF domain, an AAA ATPase domain, and a helix-turn-helix DNA binding domain. In alpha proteobacterial systems, the GAF domain interacts with the AAA ATPase domain and thus inhibits the protein's activity when fixed nitrogen is available (Arsene *et al.*, 1996). Since the anti-activator *nifL* gene (Larimer *et al.*, 2004) is not found in *Rhodopseudomonas* genome, it is believed that the NifA protein is posttranslationally activated by the uridylylated form of the PII proteins, as happens in its close relative, *Rhodospirillum rubrum* (Zhang *et al.*, 2004).

Other Factors that Influence H_2 Production

H_2 production is a complex process. It requires the integration of dozens of metabolic reactions carried out in the context of a complex web of molecular interactions within the cell. In

addition to nitrogenase synthesis and regulation, there are a number of factors that could influence H_2 production, as explained below.

- *Photophosporylation*. *Rhodopseudomonas* drives energy-intensive reactions like H_2 production by generating energy from light. Embedded in its photosynthetic membranes are two types of light-harvesting complexes referred to as "core" and "peripheral". The core complex, also known as light-harvesting 1 (LH-1), is associated with bacteriochlorophylls and carotenoids, forming a donut-shaped structure that surrounds a reaction center. The peripheral complex, also known as light-harvesting 2 (LH-2), is non-covalently bound to bacteriochlorophylls and carotenoids, which altogether assemble into arrays around the core complex (Cogdell *et al.*, 2006). LH-2 is normally expressed under high light to moderate light conditions, while two variants of the peripheral complex, light harvesting 3 (LH-3) and light-harvesting 4 (LH-4), are normally expressed under low light conditions (Kotecha *et al.*, 2012). *Rhodopseudomonas* strains vary in the number of different types of peripheral light-harvesting complexes that they synthesize (Evans *et al.*, 2005; Hartigan *et al.*, 2002). This might provide a competitive advantage to some strains in harvesting energy from light under certain light conditions (Oda *et al.*, 2008).

- *Carbon Dioxide Fixation*. *Rhodopseudomonas* is capable of fixing carbon dioxide (CO_2) into cell material via the Calvin Benson Bassham pathway, as shown below.

$$3CO_2 + 6NADPH + 5H_2O + 9ATP \rightarrow C_3H_5O_3—PO_3 + 3H^+ + 9ADP + 8P_i$$

  The process, however, has a high demand for reductant and potentially competes for electrons that could be used in H_2 production. To examine the interplay between the two processes, McKinlay and Harwood grew mutants that were capable of both CO_2 fixation and H_2 production. They found that electrons were shifted away from CO_2 fixation

toward H_2 production, and that the Calvin cycle flux significantly decreased (but did not shut off completely) in response to H_2 production (McKinlay and Harwood, 2010b).

- *Uptake Hydrogenase*. *Rhodopseudomonas* has an ability to recapture H_2 that it produces to reuse the electrons in H_2 in metabolic reactions (Larimer *et al.*, 2004). The process is mediated by uptake hydrogenase, a membrane-bound nickel-iron enzyme that catalyzes the oxidation of H_2. Rey and colleagues found that *Rhodopseudomonas* strains that were defective in hydrogen utilization produced significantly more H_2 than a strain that was not defective (Rey *et al.*, 2006).

Challenges

It is known that to support H_2 production *Rhodopseudomonas* draws on the functioning of three major metabolic modules, which are photophosporylation to generate ATP from light, carbon metabolism to generate reduced electrons, and nitrogenase to reduce N_2. In the past, we have used microarray analysis to identify a relatively small number of key genes that are important to H_2 production (Rey *et al.*, 2007). This approach has likely missed genes in peripheral metabolic modules. For example, genes involved in oxygen stress response, nitrogen acquisition, reductant supply and iron acquisition, all of which likely affect the ability of cells to generate H_2.

A complex process like H_2 production can be compared to a complex human disease, which also derives from the interplay of dozens of metabolic reactions. Recent studies took advantage of the naturally occurring variation in human subjects and experimental mouse populations by leveraging them for defining the molecular basis of complex human diseases (Schadt *et al.*, 2005; Zhu *et al.*, 2004; 2008). This concept might also be applied to the study of *Rhodopseudomonas* H_2 production, as the process is also a result of the integration of dozens of metabolic reactions within the cell. To do so, we first assembled a group of closely related strains of *Rhodopseudomonas* that produced different amounts of H_2. Then, we analyzed the naturally

occurring genetic variation among strains. Next, we examined the differences in transcriptional responses when cells were grown under three growth conditions: 1) non-H_2-producing, high light, (PM[1]-high), 2) H_2-producing (nitrogen-fixing), high light (NF-high) and 3) H_2-producing (nitrogen-fixing), low light (NF-low). The rationale is that *Rhodopseudomonas* produces H_2 under the NF-high and NF-low conditions, but not under the PM-high condition (because ammonia provided in the growth media represses the synthesis of nitrogenase), and energy generation limits the rate of H_2 production in the NF-low condition as compared to the NF-high condition. Lastly, we investigated whether differences, either in gene content or gene expression, were associated with changes in H_2 yields.

The advent of high-throughput sequencing technology, which can be applied to RNA as well as DNA samples, provides an opportunity to study DNA and RNA variation in a population of bacteria more economically and comprehensively than traditional technologies could. For example, RNA-seq, which is a technique for global analysis of mRNA transcripts by directly sequencing an RNA library (Wang *et al.*, 2009), has improved sensitivity, increased dynamic range and lower cost, compared to microarrays. Despite many advantages, the adoption of RNA-seq is hampered partially by a lack of easy-to-use, integrated, open-source tools for processing of the nucleotide sequence data that are generated as the output of the technique. In chapter 2, I explain the development of such tools for specifically processing various types of bacterial RNA-seq data. Due to their smaller genome sizes, bacterial RNA-seq libraries can be bundled for sequencing (multiplexing) to reduce per-sample sequencing costs. Moreover, the libraries can be generated using various techniques to preserve the directional information of transcripts (Armour *et al.*, 2009). As a result, the direction of sequencing reads can be in a native or a reverse-complement direction relative to the orientation of the open reading frame. This makes it difficult to process the data using available bioinformatics tools, since none of them were specifically designed and developed to process multiplexed and various types of strand-specific

[1] Non-H_2-producing medium is traditionally referred as photosynthetic mineral medium or PM.

RNA-seq data. Thus, I created Xpression, an integrated, easy-to-use tool for processing prokaryotic RNA-seq data (Phattarasukol *et al.*, 2012), and then used it for calculating the relative abundance of transcripts of *Rhodopseudomonas* grown under three conditions.

To examine genetic and transcriptomic variations among closely related strains of *Rhodopseudomonas*, I selected sixteen strains that had 16S rRNA sequence identities of 99.8% or greater from the Harwood Laboratory collection for my study. The initial challenge I faced was a lack of genomic data, as only two out of the sixteen strains had been sequenced. Consequently, we needed to rely on high-throughput sequencing technology and *de novo* genome assembly for constructing draft genomes of the fourteen strains. Moreover, genes from bacterial strains, even those from the closely related ones, often have different genomic locations. Sequences of orthologous genes, furthermore, are often not identical. As a result, we needed to perform an orthologous gene analysis and create gene-to-gene associations among the sixteen strains, so that gene content and gene expression profiles among strains could be correctly compared. In chapter 3, I describe how we assembled and annotated draft genomes, and identified the presence or absence of orthologous genes in each strain. I also discuss the differences in nitrogenase gene expression among strains and whether the differences are correlated with changes in H_2 yields. Lastly, I describe additional genes that might be important to H_2 production because they were similarly up-regulated or down-regulated in all strains.

Since H_2 production is a complex process that derives from the interplay of dozens of metabolic reactions, it is sensible to study the transcriptional activity of all annotated genes in a genome. Genes participating in a common biochemical pathway or metabolic process often exhibit similar expression patterns. Thus, in large-scale gene expression data analysis, it is common to cluster genes according to the similarity in their expression patterns (Eisen *et al.*, 1998; Stuart *et al.*, 2003). Doing so facilitates the characterization of unknown genes and the discovery of genes that are associated to phenotypes of interest. (De Smet and Marchal, 2010). I

was interested in finding clusters of genes from *Rhodopseudomonas* that work in concert when cells are producing H_2, and thus adopted the Weighted Gene Co-expression Network Analysis (WGCNA) method (Zhang and Horvath, 2005) for examining the expression patterns of *Rhodopseudomonas* genes. WGCNA is a relatively new method that has been successfully used in a number of eukaryotic microarray studies (Fuller *et al.*, 2007; Ghazalpour *et al.*, 2006; Haas *et al.*, 2012; MacLennan *et al.*, 2009; Park *et al.*, 2011; Presson *et al.*, 2008; Saris *et al.*, 2009). In chapter 4, I describe how I adapted WGCNA for constructing co-expression networks from bacterial RNA-seq data and used the results for inferring additional genes that may be important for H_2 production by *Rhodopseudomonas* when light is limited and when light is plentiful.

Lastly, I listed in the appendix the work I have done for papers published with collaborators both inside and outside the Harwood laboratory.

REFERENCES

Armour, C.D., Castle, J.C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J.K., Dey, J., Rohl, C.A., Johnson, J.M., *et al.* (2009). Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nat. Methods *6*, 647–649.

Arsene, F., Kaminski, P.A., and Elmerich, C. (1996). Modulation of NifA activity by PII in *Azospirillum brasilense*: evidence for a regulatory role of the NifA N-terminal domain. J. Bacteriol. *178*, 4830–4838.

Cogdell, R.J., Gall, A., and Köhler, J. (2006). The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to *in vivo* membranes. Q. Rev. Biophys. *39*, 227–324.

Connelly, H.M., Pelletier, D.A., Lu, T.-Y., Lankford, P.K., and Hettich, R.L. (2006). Characterization of pII family (GlnK1, GlnK2, and GlnB) protein uridylylation in response to nitrogen availability for *Rhodopseudomonas palustris*. Anal. Biochem. *357*, 93–104.

De Smet, R., and Marchal, K. (2010). Advantages and limitations of current network inference methods. Nat. Rev. Microbiol. *8*, 717–729.

Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. *2*, 621–631.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. *95*, 14863–14868.

Evans, K., Fordham-Skelton, A.P., and Mistry, H. (2005). A bacteriophytochrome regulates the synthesis of LH4 complexes in *Rhodopseudomonas palustris*. Photosynth Res 85, 169–180.

Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusis, A.J., and Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm. Genome *18*, 463–472.

Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., Brozell, A., Schadt, E.E., Drake, T.A., Lusis, A.J., *et al.* (2006). Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet. *2*, e130.

Gosse, J.L., Engel, B.J., Hui, J.C.-H., Harwood, C.S., and Flickinger, M.C. (2010). Progress toward a biomimetic leaf: 4,000 h of hydrogen production by coating-stabilized nongrowing photosynthetic *Rhodopseudomonas palustris*. Biotechnol. Prog. *26*, 907–918.

Haas, B.E., Horvath, S., Pietiläinen, K.H., Cantor, R.M., Nikkola, E., Weissglas-Volkov, D., Rissanen, A., Civelek, M., Cruz-Bautista, I., Riba, L., *et al.* (2012). Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. BMC Med Genomics *5*, 61.

Hartigan, N., Tharia, H.A., Sweeney, F., Lawless, A.M., and Papiz, M.Z. (2002). The 7.5-A electron density and spectroscopic properties of a novel low-light B800 LH2 from *Rhodopseudomonas palustris*. Biophys. J. *82*, 963–977.

Huang, J.J., Heiniger, E.K., McKinlay, J.B., and Harwood, C.S. (2010). Production of hydrogen

gas from light and the inorganic electron donor thiosulfate by *Rhodopseudomonas palustris*. Appl. Environ. Microbiol. *76*, 7717–7722.

Kapdan, I.K., and Kargi, F. (2006). Bio-hydrogen production from waste materials. Enzyme and Microbial Technology *38*, 569–582.

Kotecha, A., Georgiou, T., and Papiz, M.Z. (2012). Evolution of low-light adapted peripheral light-harvesting complexes in strains of *Rhodopseudomonas palustris*. Photosynth Res *114*, 155–164.

Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., *et al.* (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat. Biotechnol. *22*, 55–61.

Leigh, J.A., and Dodsworth, J.A. (2007). Nitrogen regulation in bacteria and archaea. Annu. Rev. Microbiol. *61*, 349–377.

MacLennan, N.K., Dong, J., Aten, J.E., Horvath, S., Rahib, L., Ornelas, L., Dipple, K.M., and McCabe, E.R.B. (2009). Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. Mol. Genet. Metab. *98*, 203–214.

Manish, S., and Banerjee, R. (2008). Comparison of biohydrogen production processes. Int J Hydrogen Energy 33, 279–286.

McKinlay, J.B., and Harwood, C.S. (2010a). Photobiological production of hydrogen gas as a biofuel. Curr. Opin. Biotechnol. *21*, 244–251.

McKinlay, J.B., and Harwood, C.S. (2010b). Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. Proc. Natl. Acad. Sci. U.S.A. *107*, 11669–11675.

Merrick, M.J., and Edwards, R.A. (1995). Nitrogen control in bacteria. Microbiol. Rev. *59*, 604–622.

Oda, Y., Larimer, F.W., Chain, P.S.G., Malfatti, S., Shin, M.V., Vergez, L.M., Hauser, L., Land, M.L., Braatsch, S., Beatty, J.T., *et al.* (2008). Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. Proc. Natl. Acad. Sci. U.S.A. *105*, 18543–18548.

Park, C.C., Gale, G.D., de Jong, S., Ghazalpour, A., Bennett, B.J., Farber, C.R., Langfelder, P., Lin, A., Khan, A.H., Eskin, E., *et al.* (2011). Gene networks associated with conditional fear in mice identified using a systems genetics approach. BMC Syst Biol *5*, 43.

Phattarasukol, S., Radey, M.C., Lappala, C.R., Oda, Y., Hirakawa, H., Brittnacher, M.J., and Harwood, C.S. (2012). Identification of a *p*-coumarate degradation regulon in *Rhodopseudomonas palustris* by Xpression, an integrated tool for prokaryotic RNA-seq data processing. Appl. Environ. Microbiol. *78*, 6812–6818.

Presson, A.P., Sobel, E.M., Papp, J.C., Suarez, C.J., Whistler, T., Rajeevan, M.S., Vernon, S.D., and Horvath, S. (2008). Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Syst Biol *2*, 95.

Rey, F.E., Heiniger, E.K., and Harwood, C.S. (2007). Redirection of metabolism for biological

hydrogen production. Appl. Environ. Microbiol. *73*, 1665–1671.

Rey, F.E., Oda, Y., and Harwood, C.S. (2006). Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. J. Bacteriol. *188*, 6143–6152.

Rubio, L.M., and Ludden, P.W. (2005). Maturation of nitrogenase: a biochemical puzzle. J. Bacteriol. *187*, 405–414.

Rupprecht, J., Hankamer, B., Mussgnug, J.H., Ananyev, G., Dismukes, C., and Kruse, O. (2006). Perspectives and advances of biological H2 production in microorganisms. Appl. Microbiol. Biotechnol. *72*, 442–449.

Saris, C.G.J., Horvath, S., van Vught, P.W.J., van Es, M.A., Blauw, H.M., Fuller, T.F., Langfelder, P., DeYoung, J., Wokke, J.H.J., Veldink, J.H., *et al.* (2009). Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. BMC Genomics *10*, 405.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C., *et al.* (2005). An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. *37*, 710–717.

Seefeldt, L.C., Hoffman, B.M., and Dean, D.R. (2009). Mechanism of Mo-dependent nitrogenase. Annu. Rev. Biochem. *78*, 701–722.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249–255.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.

Zhang, Y., Pohlmann, E.L., and Roberts, G.P. (2004). Identification of critical residues in GlnB for its activation of NifA activity in the photosynthetic bacterium *Rhodospirillum rubrum*. Proc. Natl. Acad. Sci. U.S.A. *101*, 2782–2787.

Zhu, J., Lum, P.Y., Lamb, J., GuhaThakurta, D., Edwards, S.W., Thieringer, R., Berger, J.P., Wu, M.S., Thompson, J., Sachs, A.B., *et al.* (2004). An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet. Genome Res. *105*, 363–374.

Zhu, J., Zhang, B., Smith, E.N., Drees, B., Brem, R.B., Kruglyak, L., Bumgarner, R.E., and Schadt, E.E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat. Genet. *40*, 854–861.

CHAPTER 2

Xpression – an Integrated Tool for Prokaryotic RNA-seq Data Processing

Published As: Phattarasukol, S., Radey, M.C., Lappala, C.R., Oda, Y., Hirakawa, H., Brittnacher, M.J., and Harwood, C.S. (2012). Identification of a *p*-coumarate degradation regulon in *Rhodopseudomonas palustris* by Xpression, an integrated tool for prokaryotic RNA-seq data processing. Appl. Environ. Microbiol. 78, 6812–6818.

INTRODUCTION

RNA-seq is a recently developed technique for global analysis of mRNA transcripts that involves the use of high-throughput sequencing technology (Wang *et al.*, 2009). It has a number of advantages over traditional microarray-based technologies including improved sensitivity, increased dynamic range and lower cost. As a result, it is becoming the preferred tool for gene expression studies. Despite many advantages, widespread adoption of RNA-seq is impeded by a lack of easy-to-use, integrated, open source tools to process the nucleotide sequence data that are generated as the output of the technique. Millions of raw sequence reads are generated for each RNA-seq experiment, which makes it impossible to process the sequencing data without bioinformatic tools.

A number of tools have been developed to automatically process RNA-seq data. Commercial solutions like Avadis NGS (http://www.avadis-ngs.com) and Illumina CASAVA (http://www.illumina.com/software/genome_analyzer_software.ilmn) offer rich features, but their costs are prohibitive for small laboratories. Non-commercial tools such as ArrayExpressHTS (Goncalves *et al.*, 2011) and rnaSeqMap (Leśniewska and Okoniewski, 2011) have recently been released, but none of the existing tools is specifically designed for processing prokaryotic RNA-seq data. Due to their smaller genome sizes, prokaryotic RNA-seq data can be multiplexed by adding a barcode to each sample to reduce per-sample sequencing costs. In addition, strand-specific library construction methods can be used to preserve the directional information of prokaryotic transcripts (Armour *et al.*, 2009; Hirakawa *et al.*, 2011). These methods yield sequences in a native direction as well as in a reverse-complement direction with respect to the orientation of the open reading frame (Armour *et al.*, 2009; Hirakawa *et al.*, 2011). Programming skills are required to customize existing bioinformatic tools to process these types of RNA-seq data.

Here we describe Xpression, an integrated tool that we developed to process prokaryotic RNA-seq data generated with Illumina sequencing technology. The tool accepts simple commands from users via a graphical interface, is fully automated and finishes all processing tasks starting from sequence extraction to generation of a general visualization format file that can be opened by visualization softwares such as Artemis (Carver *et al.*, 2008) or the Integrative Genomics Viewer (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). It will process data that are not strand specific. But it is also designed to analyze multiplexed and strand-specific data. It extracts and trims specific sequences from files and separately quantifies sense and antisense reads in the final results. Outputs from Xpression can also be conveniently used in downstream analysis. For example, users can apply a statistical software such as DESeq (Anders and Huber, 2010) to gene expression reports to identify differentially expressed genes.

A recent genetic and biochemical study of the purple nonsulfur phototrophic bacterium *Rhodopseudomonas palustris* revealed that *couAB* genes, which encode an enoyl-CoA lyase/hydratase and a coenzyme A ligase, are required for the degradation of the plant lignin monomers *p*-coumarate, ferulate and caffeate (Hirakawa *et al.*, 2012b). In the same study, a MarR family repressor protein named CouR was identified as binding *p*-coumaroyl-CoA to derepress *couAB* gene expression. Results from quantitative reverse transcriptase-PCR experiments showed that the *couR* mutant had levels of *couAB* expression 30-40-fold higher than those of the wild type. Here we used Xpression to process strand-specific RNA-seq data to further investigate the CouR regulon. This resulted in the identification of 11 additional genes that are likely regulated by CouR.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

R. palustris wild type strain CGA009 and a *couR* deletion mutant derived from CGA009 (Hirakawa *et al.*, 2012b) were grown anaerobically in light with succinate (10 mM) as the carbon source, as previously described (Hirakawa *et al.*, 2012b; Kim and Harwood, 1991). Cells in the mid-logarithmic phase of growth, where they express *p*-coumarate degradation genes at high levels (Pan *et al.*, 2008), were chilled in an ice-water bath, harvested by centrifugation and the pellets were frozen in liquid nitrogen and then stored at -80°C.

Electrophoretic Mobility Gel Shift Assays

CouR was purified as described (Hirakawa *et al.*, 2012b) and electrophoretic mobility gel-shift assays were carried out as described previously (Hirakawa *et al.*, 2012b) except that probes specific for the promoter of each gene were generated by PCR amplification with genomic *R. palustris* CGA009 DNA as template. For each probe, the entire intergenic region was amplified.

Strand-specific cDNA Library Construction for RNA-seq

Cells previously stored at -80°C were thawed and disrupted by bead beating, and RNA was then purified from cells as previously described (Hirakawa *et al.*, 2011). A strand-specific cDNA library was prepared from total RNA by a previously described method called Not-so-Random (NSR) RNA-seq (Armour *et al.*, 2009). First and second strand synthesis, NSR RNA-seq library construction and DNA sequencing on an Illumina GA2 were carried out as described (2,8). For this we specified nucleotide read lengths of 36 bases. The DNA raw sequencing reads have been deposited in the NCBI Gene Expression Omnibus under accession number GSE39025.

Xpression Installation

Xpression is freely available for download from the Harwood Laboratory website (https://depts.washington.edu/cshlab/html/rnaseq.html). Due to the nature of the dependent software Biopython (Cock *et al.*, 2009), SAMtools (Li *et al.*, 2009), Pysam (http://code.google.com/p/pysam) and the Burrows-Wheeler Alignment (BWA) tool (Li and

Durbin, 2010), the installation of Xpression requires a properly configured Unix-like operating system. The website provides two alternatives for getting Xpression onto a desktop computer. For those who have a Linux or Unix-like operating system, the best option is to use the provided automatic script that will install all required software from the source. For those with a Windows or Mac OS operating system, we have provided a fully operational, system-independent graphical environment (Xpression VE) that can run Xpression. The only software that Xpression VE needs is a free virtualization software called VirtualBox (https://www.virtualbox.org/). Please consult the Xpression virtual system user guide available on the Harwood laboratory website for point-and-click directions for installation of Xpression VE onto a computer. This program can be easily installed on a desktop, laptop or netbook computer.

RNA-seq Data Processing

First, reference files for the genome sequence being queried, *R. palustris* CGA009 in this case, were uploaded. We obtained these files from NCBI as described in the supporting documentation available on the Harwood web site. The FASTA file was uploaded into Xpression with the "FASTA Reference" button shown in Figure 2.1A and the Genbank file was uploaded into Xpression with the "Genbank Reference" button shown in Figure 2.1 A. The RNA-seq data (the FASTQ files from the Illumina sequencer) of the wild type were uploaded with the "Sequencing FASTQ" button. Next, we entered "Sample Information" (Figure 2.1 A), which included a specified sample ID, "wild_type" in the case shown in Figure 2.1 A. Also the barcode for the samples to be analyzed was entered into the appropriate window. We selected "4 - Generate Visualization" to be the final step in the analysis. In the next step we specified "Sample Options" (Figure 2.1 B), after clicking on the "Options" menu located as shown in Figure 2.1 A. The sequencing data were strand-specific and we entered this in the "Sample Options" panel. We also indicated that we would allow two nucleotide mismatches when aligning reads against a

reference genome, and that the start position of the biological sequence is the 6th nucleotide from the 5' end. This trims the barcode plus one additional nucleotide off the sequence reads, effectively reducing them from 36 to 31 nucleotides in length. Finally, we clicked "Add to Queue" (Figure 2.1 A) to queue up additional files to be analyzed. Once everything was set, we pressed the "Start Run" button and waited for Xpression to finish all the tasks in its internal workflow (Figure 2.2). The processing tasks included sequence extraction, filtering, trimming, alignment, quantification, normalization and visualization as described below in the Results. Xpression generated a mapping statistics table and an expression profile presented as a comma-separated-values (CSV) file. Xpression also displayed mapping positions and numbers of uniquely-mapped reads computed in the previous steps in customized Wiggle plots (http://genome.ucsc.edu/goldenPath/help/wiggle.html/), which we loaded onto the Artemis genome browser (Carver *et al.*, 2008). The presentation of the mapped reads in Artemis was as described previously (Hirakawa *et al.*, 2011).

Identifying Differentially Expressed Genes

We used the statistical software DESeq (Anders and Huber, 2010) to evaluate whether, for a given region, an observed difference in read counts between the *couR* mutant and the wild type was significant. DESeq takes into account technical and biological variability of count data and, consequently yields more balanced and accurate results than simple calculation. Features with *p*-values <0.05 and fold change ratios >3.0 were considered to be differentially expressed.

RESULTS

Previously we identified *couA* and *couB* as encoding enzymes that are regulated by a transcriptional repressor protein that we named CouR (Hirakawa *et al.*, 2012b). CouR binds *p*-coumaroyl-CoA to derepress the expression of these genes. CouA is an enoyl-CoA

hydratase/lyase and CouB, a coenzyme A ligase required for removing the side chain of the phenylpropanoid, *p*-coumarate, and converting it to *p*-hydroxybenzaldehyde. The same enzymes remove the side chains of ferulate and caffeate. It is often the case that genes required for the transport of a particular compound or for the degradation of related compounds are coordinately regulated with degradation genes. To identify other genes that may be regulated by CouR, we compared the transcriptome of succinate-grown wild-type cells with that of succinate-grown cells of the *couR* mutant using RNA-seq and we used Xpression to analyze the sequence data. Succinate is a tricarboxylic acid cycle intermediate and a good growth substrate for *R. palustris*. *p*-Coumarate is degraded to acetyl-CoA, which then enters the tricarboxylic acid cycle (Hirakawa *et al.*, 2012b; Pan *et al.*, 2008).

Sequence Read Extraction, Filtering, and Trimming

Xpression scanned sequences from the input FASTQ files and extracted only those whose sequence started with the specified barcode. To ensure that the extracted reads were of high quality, the tool assessed them using base-calling-accuracy scores. Reads whose accuracy was 99% or greater (called “high-quality reads”) were kept for further processing while reads whose accuracy was lower (called “low-quality reads”) were discarded. As shown in Table 2.1, 79.59 % of the wild-type RNA-seq data were of high quality and 81.92 % of the *couR* mutant data were of high quality. Next, Xpression trimmed the reads as described in Materials and Methods to remove the barcode so that the reads could be accurately aligned to the *R. palustris* genome in the next step.

Sequence Read Alignment and Classification

Xpression used the BWA tool to map high-quality, trimmed reads from the previous step against the FASTA reference genome sequence. Once finished, the tool categorized reads according to the mapping results into four classes. First, reads that aligned to unique locations in the reference genome were called “uniquely-mapped reads”. Second, reads that aligned to

unique locations but with more than two nucleotide mismatches were called “partially-mapped reads”. Third, reads that aligned to more than one location were called “non-uniquely-mapped reads”. Reads that could not be aligned were called “unmapped reads.”

The mapping statistics generated by Xpression are shown in Table 2.1. Xpression aligned 92.48% of the high-quality reads from the wild-type data to the reference genome. Of these, 28.97% of the reads mapped to a single place on the genome (uniquely mapped). A large number of reads, 66.08%, mapped to more than one location. A further analysis revealed that most of these non-uniquely-mapped reads aligned to the two ribosomal RNA operons. Percentages of mapped reads from the *couR* mutant data were similar (Table 2.1). Unlike other classes of reads, uniquely-mapped reads were aligned unambiguously to locations in the reference genome and Xpression quantified this class of reads to infer transcript levels in the next step.

Sequence Read Quantification, Normalization and Visualization

Xpression first collected all annotated features in the reference genome from the Genbank reference annotation file. The tool then counted how many uniquely-mapped reads were located inside each gene and provided raw and normalized numbers for each gene (Table 2.2). Since our RNA-seq data were strand-specific, the tool also took into account the alignment direction of reads with respect to the transcriptional direction of genes. It thus separated gene counts into “sense counts” (sense), if reads were mapped in the same direction of the gene, and “antisense counts” (antis), if reads were mapped in the opposite direction (Figure 2.3).

The tool also counted how many uniquely-mapped reads were located inside each intergenic region, which was defined as the nucleotides between two adjacent annotated features. These reads were collectively called “intergene counts”. Since there is no transcriptional direction for intergenic regions to be referred to, Xpression separated intergene counts based on which DNA strand the reads aligned to. They were either “intergene-top” counts, abbreviated to “igtop”,

which were the number of reads whose sequences matched intergenic sequences on the top (leading) DNA strand, or "intergene-bottom", abbreviated to "igbot", which were those whose sequences matched intergenic sequences on the bottom DNA strand (Figure 2.3). Table 2.2 shows the number of reads that mapped to *couA* and the intergenic region upstream of *couA*. In the sense direction, there were 16 uniquely-mapped reads for the wild type while there were 1,636 uniquely-mapped reads for the *couR* mutant in the *couA* gene. In the intergenic region, 6 and 16 reads matched the top strand for the wild type and the *couR* mutant respectively.

To facilitate comparison of read numbers between annotated regions and between samples, Xpression normalized raw counts to reads per million uniquely-mapped reads (RPM) and reads per kilobase per million uniquely-mapped-reads (RPKM). This normalization method was similar to that of Mortazavi *et al.* (Mortazavi *et al.*, 2008). The only difference was that Xpression normalized raw counts to the total number of uniquely-mapped reads, instead of the total number of reads, to reduce the weight of the ribosomal RNA reads. When normalized, the *couA* gene in the wild-type strain and the *couR* mutant had 8.08 and 976.27 RPKM, respectively (Table 2.2).

Since our RNA-seq data are strand-specific, Xpression created four plots for each sample as illustrated in Figure 2.4. A comparison of Figure 2.4 A with Figure 2.4 B shows that genes *rpa1782- rpa1793* were expressed at higher levels in the *couR* mutant compared to the wild type. An exception is *rpa1790*, predicted to encode a diguanylate cyclase, which was not differentially expressed between the two strains.

The CouR Regulon

Xpression identified 13 genes that were expressed at three-fold or higher levels in the *couR* mutant relative to the wild type (Table 2.3). Consistent with our previous RT-PCR data, the RNA-seq data indicate that *couA* and *couB* are repressed by CouR. In fact, with the exception of *rpa1790*, genes *rpa1782 – rpa1793* near *couAB* were all repressed by CouR. *Rpa1782-1784* are

predicted to encode a TrapT family transporter and *rpa1791-1793* to encode a predicted ABC transporter. In addition, *rpa1789* encodes a periplasmic binding protein that has been expressed and shown by a fluorescence-based thermal shift assay to bind *p*-coumarate, ferulate, and cinnamate (Giuliani *et al.*, 2011). Genes *rpa4198* and *rpa4199* were also strongly repressed by CouR (Table 2.3). These genes, separated by 148 bp, are predicted to encode an amidohydratase and halide hydrolase.

We found a CouR binding motif [GTTATA NNN TATAAC] (9) upstream of *rpa1782*, *rpa1786*, and in the intergenic region between *rpa1793* and *couR* (*rpa1794)*. In addition, a CouR binding motif with a three nucleotide substitution was present upstream of *rpa4198-4199* (Figure 2.5 A). In gel-shift assays, using as probes the regions 170-200 bp upstream of genes with predicted CouR binding motifs, we found that CouR bound to the promoter regions of *rpa1782*, *rpa1793-couR*, and *rpa4198* (Figure 2.5 B). We have previously shown that CouR binds to the *rpa1786* (*couA*) promoter region (Figure 2.1 **Error! Reference source not found.**B)(Hirakawa *et al.*, 2012b).

DISCUSSION

Here we defined a CouR regulon of 13 genes using Xpression, a nucleotide sequence processing tool that we designed for the quantification and visualization of data that are generated by the method of RNA-seq. Our RNA-seq experimental design involved isolating RNA from wild-type cells and *couR* mutant cells grown on succinate and then reverse transcribing the RNA in a strand-specific manner to generate cDNA, which was then sequenced on an Illumina platform. CouR is a transcriptional repressor protein that binds *p*-coumaroyl-CoA to derepress gene expression. We found that CouR not only controls the expression of *couAB* genes for *p*-coumarate degradation as we previously described (Hirakawa *et al.*, 2012b), but also transport systems that are likely involved in the uptake of *p*-coumarate and structurally related

compounds into cells. The functions of the predicted amidohydratase and halidohydrolase genes that are regulated by CouR are unclear, but one could speculate that they are involved in preparing structurally modified forms of phenylpropanoids to enter the *p*-coumarate degradation pathway. *P*-Coumarate metabolism is of interest for several reasons. *P*-Coumarate is produced in large amounts by green plants as a precursor of lignin (Whetten and Sederoff, 1995). It is also a breakdown product of lignin and is used as a carbon source by diverse soil bacteria (Hirakawa *et al.*, 2012b; Trautwein *et al.*, 2012). In addition we discovered that *R. palustris* synthesizes an unusual acyl-homoserine lactone quorum sensing signal, *p*-coumaroyl-HSL, from *p*-coumarate that it obtains from its environment (Schaefer *et al.*, 2008). A quantitative proteome and microarray study (Pan *et al.*, 2008) suggested that at least 40 genes and their encoded proteins are upregulated during growth on *p*-coumarate compared to succinate. Some of these are regulated by *p*-coumaroyl-HSL and the transcription protein RpaR and others are regulated by CouR.

Xpression is well suited to analyze gene expression data generated in eubacterial and archeal RNA-seq projects. In the example presented here we analyzed strand specific data. This allowed us to capture "sense" reads that reflect the expression levels of genes and also "antisense" reads that may reflect the expression levels of antisense RNAs. We found no evidence for antisense RNAs in the example shown here, but in other recent work, Xpression revealed the presence of an antisense RNA that modulates quorum sensing in *R. palustris* (Hirakawa *et al.*, 2011; 2012a). Intergenic data can be useful for visualizing the transcription start site of a gene and this can be seen in the igtop reads mapped to the genome in Figure 2.4 B. How accurate this is depends on the method used to prepare cDNA for sequencing. The preparation method that we used generates cDNAs that map to within about 10 bp of transcription start sites (unpublished data). Intergenic data may also reveal the presence of small RNA transcripts. If a user analyzes data that are not prepared in a strand specific manner then both sense and antisense read counts will

be folded into the “genic” count and both igtop and igbot read counts will be folded into the “inter” count.

Xpression should be useful to research laboratories that do not routinely carry out gene expression studies. Recent dramatic decreases in sequencing costs coupled with the ease of use of Xpression should allow individual investigators to function autonomously to generate and process gene expression data generated by RNA-seq methods.

ACKNOWLEDGEMENTS

This work was supported by the Office of Science (BER), U.S. Department of Energy [grant DE-FG02-08ER64482: S.P., C.L. Y.O. and C.S.H.], and the National Institutes of Health, National Institute of Allergy and Infectious Diseases [grant U54 AI057141: M.R. and M.B.]. Hidetada Hirakawa received funding from the Japan Society for the Promotion of Science (JSPS), the Uehara Memorial Foundation, and the Cell Science Research Foundation.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Armour, C.D., Castle, J.C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J.K., Dey, J., Rohl, C.A., Johnson, J.M., *et al.* (2009). Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nat. Methods *6*, 647–649.

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B.G., Parkhill, J., and Rajandream, M.-A. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics *24*, 2672–2676.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

Giuliani, S.E., Frank, A.M., Corgliano, D.M., Seifert, C., Hauser, L., and Collart, F.R. (2011). Environment sensing and response mediated by ABC transporters. BMC Genomics *12 Suppl 1*, S8.

Goncalves, A., Tikhonov, A., Brazma, A., and Kapushesky, M. (2011). A pipeline for RNA-seq data processing and quality assessment. Bioinformatics *27*, 867–869.

Hirakawa, H., Harwood, C.S., Pechter, K.B., Schaefer, A.L., and Greenberg, E.P. (2012a). Antisense RNA that affects *Rhodopseudomonas palustris* quorum-sensing signal receptor expression. Proc. Natl. Acad. Sci. U.S.A. *109*, 12141–12146.

Hirakawa, H., Oda, Y., Phattarasukol, S., Armour, C.D., Castle, J.C., Raymond, C.K., Lappala, C.R., Schaefer, A.L., Harwood, C.S., and Greenberg, E.P. (2011). Activity of the *Rhodopseudomonas palustris p*-coumaroyl-homoserine lactone-responsive transcription factor RpaR. J. Bacteriol. *193*, 2598–2607.

Hirakawa, H., Schaefer, A.L., Greenberg, E.P., and Harwood, C.S. (2012b). Anaerobic *p*-coumarate degradation by *Rhodopseudomonas palustris* and identification of CouR, a MarR repressor protein that binds *p*-coumaroyl coenzyme A. J. Bacteriol. *194*, 1960–1967.

Kim, M.K., and Harwood, C.S. (1991). Regulation of benzoate-CoA ligase in *Rhodopseudomonas palustris*. FEMS Microbiology Letters.

Leśniewska, A., and Okoniewski, M.J. (2011). rnaSeqMap: a Bioconductor package for RNA sequencing data exploration. BMC Bioinformatics *12*, 200.

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

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

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and

quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Pan, C., Oda, Y., Lankford, P.K., Zhang, B., Samatova, N.F., Pelletier, D.A., Harwood, C.S., and Hettich, R.L. (2008). Characterization of anaerobic catabolism of *p*-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. Mol. Cell Proteomics *7*, 938–948.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Schaefer, A.L., Greenberg, E.P., Oliver, C.M., Oda, Y., Huang, J.J., Bittan-Banin, G., Peres, C.M., Schmidt, S., Juhaszova, K., Sufrin, J.R., *et al.* (2008). A new class of homoserine lactone quorum-sensing signals. Nature *454*, 595–599.

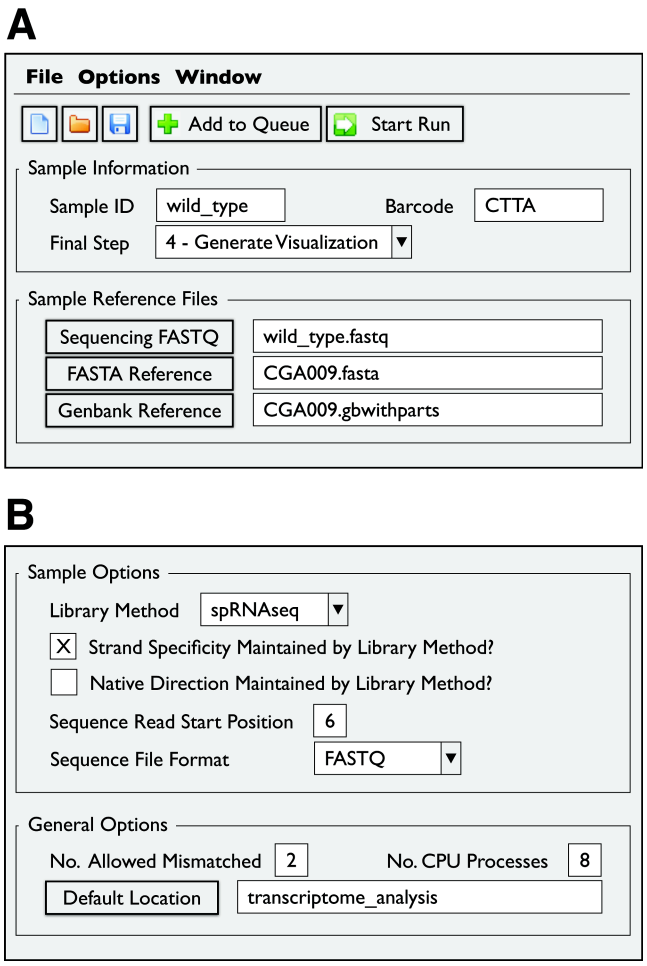
Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinformatics *14*, 178–192.

Trautwein, K., Wilkes, H., and Rabus, R. (2012). Proteogenomic evidence for β-oxidation of plant-derived 3-phenylpropanoids in "*Aromatoleum aromaticum*" EbN1. Proteomics *12*, 1402–1413.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

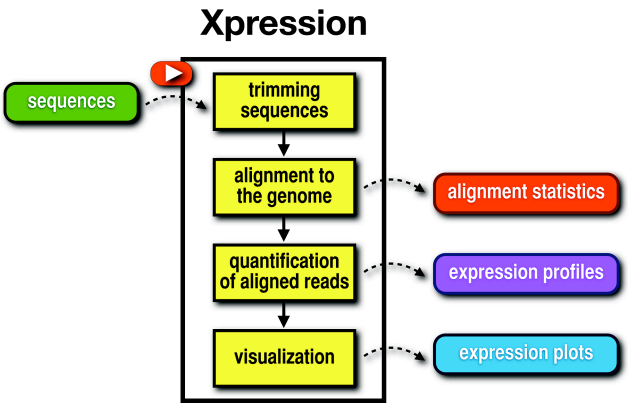
Whetten, R., and Sederoff, R. (1995). Lignin Biosynthesis. Plant Cell *7*, 1001–1013.

Figure 2.1. A depiction of the Xpression graphical interface.



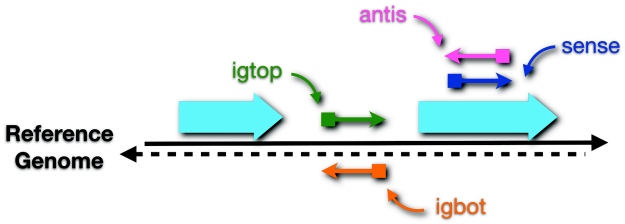
(A) Settings to analyze the wild type RNA-seq data are shown as an example, and (B) Depiction of the sample options window.

Figure 2.2. Tasks carried out the internal workflow of Xpression.



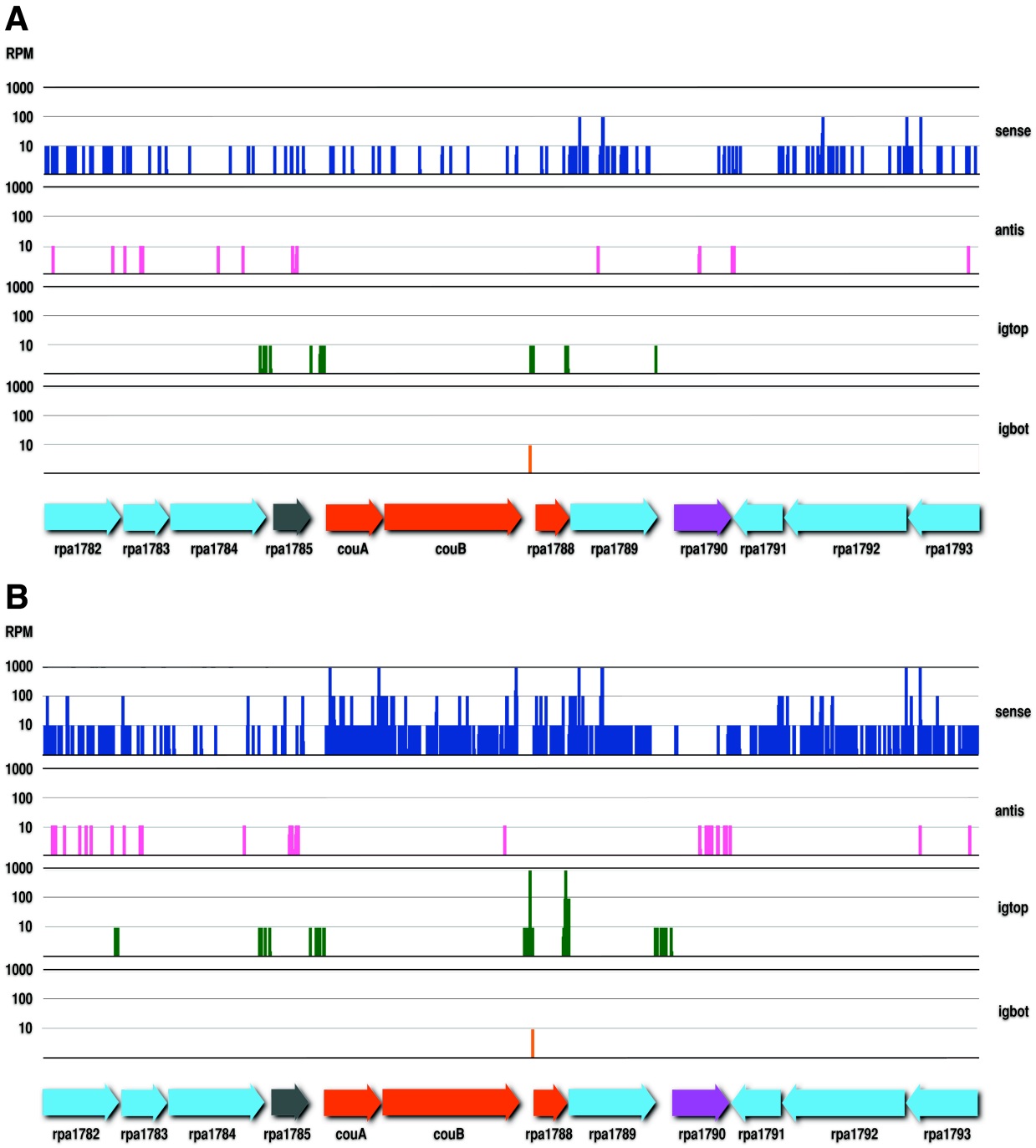
Tasks that are automatically carried out include sequence extraction, filtering, trimming, alignment, quantification, normalization and visualization. Once all the processing tasks are completed, Xpression provides the user with alignment statistics, gene expression profiles and gene expression plots.

Figure 2.3. A depiction of the types of DNA sequence reads that are quantified by Xpression.



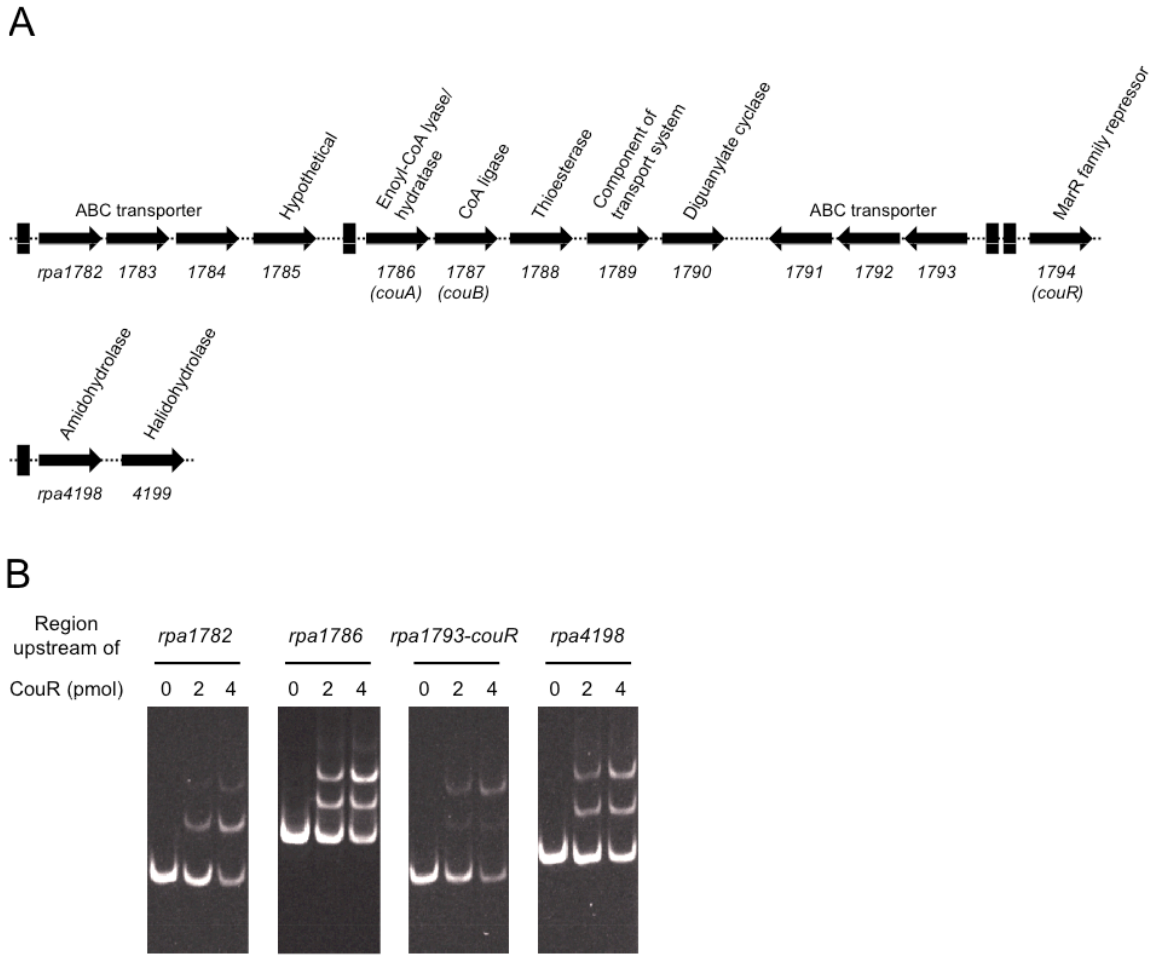
Strand-specific reads are counted separately according to their mapping position and direction. Reads that map inside an annotated feature are counted as either “sense” if they map to the same strand as an open reading frame or “antisense” (abbreviated to “antis”) if they map to the opposite strand of an open reading frame. Reads that map between two annotated features are counted as either “intergene-top ” (abbreviated to “igtop”) if the sequence matches that of the top (leading) DNA strand sequence or “intergene-bottom” (abbreviated to “igbot”) if they the sequence matches that of the bottom DNA strand.

Figure 2.4 Visualization of RNA-seq data using output from Xpression.



Comparison of RNA-seq profiles of wild type (A) and *couR* mutant (B) cells grown on succinate. Numbers of sequence reads (RPM) were binned (1-10, 11-100 or 101-1000) as indicated on the y-axes and the start location of the reads was overlaid on the R. palustris CGA009 genome using Artemis software as described in the Materials and Methods. The type of DNA sequence read is indicted by line color: sense reads, blue; antisense reads, pink; intergenic regions on the top DNA strand, green; and intergenic regions on the bottom DNA strand, orange. Predicted gene functions are color-coded as follows: blue, transport; grey, hypothetical; red, enzyme; purple, diguanylate cyclase. This is a simplified version of the figure generated in Artemis.

Figure 2.5. Map of genes in the CouR regulon and gel-shift assay.



(A) Map of genes in the CouR regulon. CouR binding motifs are indicated by the black boxes. Note that rpa1790 is not regulated by CouR. (B) Gel-shift assay showing binding of CouR to the promoter regions of genes found to be regulated by CouR in RNA-seq experiments. CouR protein (0, 2 or 4 pmol) was added to reaction mixtures containing 0.3 pmol of DNA probe. The relative gene sizes are not accurately depicted. For this information see Figure 2.4.

Table 2.1. Sequence read statistics of RNA-seq data generated from *Rhodopseudomonas palustris* wild type and the *couR* mutant.

	Wild Type		***couR* Mutant**	
Total Reads	**11,507,621**	**100.00 %**	**11,132,011**	**100.00 %**
High-quality Reads	9,159,246	79.59%	9,119,736	81.92%
Low-quality reads	2,348,375	20.41%	2,012,275	18.08%
Total High-quality Reads	**9,159,246**	**100 %**	**9,119,736**	**100 %**
Mapped Reads	8,470,854	92.48%	8,413,620	92.26%
Unmapped Reads	688,393	7.52%	706,117	7.74%
Total Mapped Reads	**8,470,854**	**100 %**	**8,413,620**	**100 %**
Uniquely-mapped Reads	2,453,640	28.97%	2,076,547	24.68%
Partially-mapped Reads	419,433	4.95%	305,605	3.63%
Non-uniquely-mapped Reads	5,597,780	66.08%	6,031,467	71.69%

Table 2.2. *CouA* read counts derived from RNA-seq data from wild type and *couR* mutant cells.

				Wild Type			*CouR* Mutant		
Locus	**Size (bp)**	**Kind of Read[a]**	**Gene**	**Raw**	**RPM[b]**	**RPKM[c]**	**Raw**	**RPM**	**RPKM**
rpa1786	297	igbot		0	0	0	0	0	0
rpa1786	297	igtop		6	2.45	8.23	16	7.71	25.94
rpa1786	807	sense	*couA*	16	6.52	8.08	1,636	787.85	976.27
rpa1786	807	antis		0	0	0	0	0	0

[a] See Figure 2.3 for a description of kinds of reads.
[b] Reads per million uniquely-mapped reads
[c] Reads per kilobase per million uniquely-mapped reads

Table 2.3. Genes regulated by CouR.

Locus	Kind of Read	Gene Name	Annotation[a]	Exp. Ratio couR vs. WT[b]
rpa1782[c,d*]	sense		TrapT family transporter, periplasmic binding protein	4.4
rpa1783	sense		Trap transporter, small permease component	7.3
rpa1784	sense		Trap transporter,large permease component	4.8
rpa1785	sense		Hypothetical protein	6.7
rpa1786*	sense	*couA*	Enoyl-CoA lyase/hydratase	81.6
rpa1787	sense	*couB*	Coenzyme A ligase	96.4
rpa1788	igtop			67.7
rpa1788	sense		Thioesterase	30.8
rpa1789	igtop			16.0
rpa1789	sense		Periplasmic binding protein	10.0
rpa1791	sense		ABC transporter, ATP binding protein	15.3
rpa1792	sense		ABC transporter, ATP binding protein	9.5
rpa1793*	sense		ABC transporter, membrane protein	9.9
rpa4198*	sense		Amidohydrolase 2	30.1
rpa4199	sense		Putative 2-haloacid halidohydrolase	20.4

[a] As defined by the Joint Genome Institute site (www.jgi.doe.gov)
[b] Genes and intergenic regions that were changed more than three-fold with a *p*-value less than 0.05 are listed. Expression ratios were calculated by DESeq, which took into account technical and biological variability of count data, and consequently they were slightly different from those obtained from simple calculation.
[c] Those genes regulated by *p*-coumarate in Affymetrix Genechip experiments (Pan *et al.*, 2008) are indicated by bold lettering.
[d] Elements with promoters that were examined by gel-shift experiments are indicated by an asterisk.

CHAPTER 3

Comparative Genomic and Transcriptomic Analysis of *Rhodopseudomonas* Strains Provides Insights into Determinants of Microbial Hydrogen Gas Production

INTRODUCTION

Rhodopseudomonas is a photosynthetic bacterial genus that belongs to the alpha proteobacteria. It is widely distributed in natural environments and is considered one of the most metabolically versatile bacteria studied to date (Larimer *et al.*, 2004). Importantly, *Rhodopseudomonas* can produce hydrogen gas (H_2) via the process of nitrogen fixation. H_2 has received serious consideration as an alternative fuel to petroleum because it is a clean-burning fuel that can be produced biologically (Das and Veziroglu, 2001). *Rhodopseudomonas* is considered an ideal platform to develop as a biocatalyst for H_2 production because of its metabolic versatility and ability to draw on abundant natural resources such as sunlight and biomass for H_2 production (Rey *et al.*, 2007).

A number of *Rhodopseudomonas* strains have been cultivated and their strain-to-strain genotypic and phenotypic diversity has been studied. First, Oda and colleagues applied BOX-PCR genomic DNA fingerprinting to characterize 75 isolates from sediment samples at three different sites in The Netherlands (Oda *et al.*, 2002). They found not only genotypic differences but also significant phenotypic differences among strains. It had been thought that all *Rhodopseudomonas* strains were capable of degrading benzoate. Oda and colleagues, however, found that not of all *Rhodopseudomonas* isolates they characterized were able to do so. Furthermore, Oda and colleagues sequenced the genomes of four *Rhodopseudomonas* strains and compared them to the genome of *Rhodopseudomonas* strain CGA009 (Oda *et al.*, 2008). *Rhodopseudomonas* strain CGA009 is the type strain studied in the Harwood Laboratory and its genome was sequenced in 2004 (Larimer *et al.*, 2004). They found that the five strains belong to the same taxonomic unit and have many characteristics in common. For example, the genomes are about 5.5-Mbp in size, comprised of a single circular chromosome, and encode about 5,000 genes. However, they also found substantial genetic variation between the five strains. For example, there are about 2,750 genes (approximately 55% of the genes from each

strain) that are present in all five genomes, and there are a significant number of strain-specific genes (ranging from 420 to 859 genes).

I was interested in studying genetic and transcriptomic variations among related strains of *Rhodopseudomonas*, and determining whether the differences in gene content and gene expression affect H_2 production. H_2 production is a complex process that requires the integration of dozens of metabolic reactions carried out in the context of a complex web of molecular interactions within the cell. Figure 3.1 illustrates the metabolic modules (photophosporylation to generate ATP from light, carbon metabolism to generate reduced electrons, and nitrogenase) required for H_2 production. By studying closely related strains, we can take advantage of the naturally occurring variation in the bacterial population and use them to investigate differences in gene content and gene expression among strains and their effect on H_2 production on a finer scale than the previous studies.

I selected sixteen *Rhodopseudomonas* strains that had 16S rRNA sequence identities of 99.8% or greater from the Harwood Laboratory collection for my analysis. The initial challenge I faced was that only two out of sixteen in my strain selection (strain CGA009 and TIE-1) had been sequenced and annotated, and thus we needed to sequence, assemble and annotate the genomes of the fourteen strains. Although the advent of high-throughput sequencing technologies makes genome sequencing more accessible and economical to individual laboratories, the process of *de novo* genome assembly, especially from short nucleotide sequences, is still a challenge. Moreover, genes from bacterial strains, even closely related ones, often have different genomic locations and their gene sequences are not identical. Consequently, I needed to use information from an orthologous gene analysis to create gene-to-gene associations among the sixteen strains so that I could accurately compare their gene content and gene expression profiles.

In this chapter, I describe how we assembled draft genomes from high-throughput sequencing data, annotated the draft genomes and used information from an orthologous analysis to identify the presence or absence of genes in each strain. I also explain how I examined transcriptional profiles from three growth conditions: 1) non-H_2-producing, high light, (PM[2]-high), 2) H_2-producing (nitrogen-fixing), high light (NF-high) and 3) H_2-producing (nitrogen-fixing), low light (NF-low). The rationale is that *Rhodopseudomonas* produces H_2 under the NF-high and NF-low conditions, but not under the PM-high condition (because ammonia provided in the growth media represses the synthesis of nitrogenase, the catalyst of H_2 production), and energy generation limits the rate of H_2 production in the NF-low condition as compared to the NF-high condition. Then, I discuss whether there were variations in nitrogenase synthesis gene content and gene expression and whether the differences affected H_2 yields. Lastly, I explain how I identified orthologous genes that were similarly up-regulated or down-regulated in all strains and whether the differences in expression changes of these genes influenced H_2 production.

MATERIAL AND METHODS

Bacterial Strains, Growth Conditions and Phenotypes

I selected sixteen *Rhodopseudomonas* strains that had 16S rRNA sequence identities of 99.8% or greater from the Harwood Laboratory collection (Table 3.1). These strains were isolated from different geographic locations around the world, including the United States, the Netherlands, and Japan. Later, when comparing expression profiles, strain CGA010 was additionally included in the group. Strain CGA010 is a derivative of strain CGA009, whose *hupV* gene was repaired so that uptake hydrogenase becomes functional (Rey *et al.*, 2006).

[2] Non-H_2-producing medium is traditionally referred as photosynthetic mineral medium or PM.

All laboratory work, explained briefly here, was done by Dr. Yasuhiro Oda and Colin Lappala, a research scientist and research assistant in the Harwood laboratory. *Rhodopseudomonas* strains were grown anaerobically in light at 30°C in 10 ml non-nitrogen-fixing medium (PM) or nitrogen-fixing medium (NF) (Oda *et al.*, 2005) containing 20 mM acetate, 10 μl VCl_3, and Wolfe's vitamin solution (5 ml/l). Cultures were illuminated with 60W incandescent lamps for high-light conditions or 15W incandescent lamps for low-light conditions. The medium was sufficiently depleted of nickel such that uptake hydrogenase was not expected to be present to complicate H_2 measurement (Rey *et al.*, 2006). H_2 yield (μmol H_2/mg protein) was measured by gas chromatography using a thermal conductivity detector and nitrogenase activity (nmol C_2H_4 formed/min/mg protein) was measured by gas chromatography using the acetylene reduction assay as explained in (Oda *et al.*, 2005; Rey *et al.*, 2007).

Preparation of DNA and RNA for Sequencing

Dr. Yasuhiro Oda and Colin Lappala used the QIAGEN Genomic-tip kit to isolate DNA from cells grown in 30 ml PM containing 10 mM acetate and 0.3% yeast extract. Then, library construction was performed according to following the manufactures' instructions. After the DNA was diluted, sheared, ligated, amplified and purified, products of 400 to 500 bp were size-selected, amplified and sequenced using an Illumina HiSeq sequencer. For RNA sequencing, the miRNeasy Mini kit (QIAGEN) was used to purify RNA harvested from cells from mid-log phase following the manufacture's instruction. After the RNA was treated with DNase and purified, a cDNA library was prepared for sequencing as explained in (Chugani *et al.*, 2012; Hirakawa *et al.*, 2011).

***De novo* Genome Assembly**

The program Velvet (Zerbino and Birney, 2008) was used to assemble draft genomes of three *Rhodopseudomonas* strains: 1a1, BIS3 and CEA001. In brief, I called the *velveth* command to create an internal data structure (a hash table) in order to represent sequencing

reads, and then the *velvetg* command to build *de Bruijn* graphs, simplify and correct errors in the graphs, and output sequences of a draft genome. The process was repeated about sixty times with different combinations of parameters (*i.e.,* kmer, expected coverage and coverage cutoff) and a draft genome consisting of the least number of scaffolds and/or having the largest scaffolds was selected. To evaluate the quality of a draft genome, I used Prodigal (Hyatt *et al.*, 2010) to predict protein sequences, used Blast (http://blast.ncbi.nlm.nih.gov) to compare the predicted protein sequences against known protein sequences in the NCBI database, and then counted how many of the predicted protein sequences matched to known protein sequences. The rationale is that low quality draft genomes contain high numbers of artificial nucleotide sequences, which in turn results in artificial protein sequences that do not match or match with a very low percentage identity to real protein sequences.

To determine the possibility of using short sequencing reads for constructing *Rhodopseudomonas* genomes, I selected three strains, 1a1, BIS3 and CEA001, for preliminary sequencing. We used an Illumina GA-II sequencer for sequencing genomic libraries of the three and generated approximately 8 million pairs of 76-base reads (with an estimated coverage of 220×) per library. I then constructed draft genomes from these sequencing data and was able to obtain results with a relatively small number of scaffolds and the size of the largest scaffold more than 500 kilobase (Table 3.2). In addition to quantitative assessment, I also evaluated the assembly results qualitatively. I found that 92% to 98% of the predicted protein sequences matched with percentages of sequence identity 90% or higher to known protein sequences in the NCBI database (Table 3.2). This demonstrated that the draft genomes I constructed likely contained very few assembly errors.

In 2012, an Illumina Hi-Seq sequencer was available to us and we decided to use it for sequencing genomic libraries of the fourteen strains (including re-sequencing the three strains that I previously assembled). Since the Illumina Hi-Seq sequencer is capable of generating

longer sequencing reads than the Illumina GA-II sequencer, we expected assembly results to be even better than the results we had already obtained. The machine generated approximately 1.2 million pairs of 101-base reads (with an estimated coverage of 45×) per library. Matthew Ready, a bioinformaticain at the sequencing facility of the Department of Microbiology, then used an in-house assembly pipeline (based on Velvet) to assemble draft genomes for us. Each draft genome he constructed consisted of a relatively high number of scaffolds, with the largest scaffolds ranging from about 81 to 257 kilobase (Table 3.2). Quantitatively, the draft genomes assembled from the Illumina HiSeq sequencing data looked worse than those assembled from the Illumina GA-II sequencing data because the numbers of scaffolds constituting a draft genome are much higher and the sizes of the largest scaffold in a draft genome are significantly lower. This is likely due to the fewer numbers of reads per library that were generated (1.2 millions reads from Illumina HiSeq compared to 8 millions reads from Illumina GA-II), even though the reads are longer (101-base reads from Illumina Hiseq compared to 76-base reads from Illumina GA-II). This makes it difficult for the assembly software to assemble long and contiguous pieces of DNA. I again evaluated the draft genomes qualitatively and found that that 90% to 96% of the predicted protein sequences matched with percentages of sequence identity 90% or higher to known protein sequences in the database (Table 3.2). These numbers are comparable to those from the draft genomes assembled from the Illumina GA-II sequencing data, and demonstrated that, despite the high numbers of scaffolds, the draft genomes assembled from the Illumina HiSeq sequencing data were still of high quality. To ensure technical consistency among the draft genomes, I decided to discard the three draft genomes assembled from the Illumina GA-II sequencing data and use the fourteen draft genomes assembled from the Illumina HiSeq sequencing data in my analysis.

Genome Annotation

The fourteen draft genomes were submitted to the Integrated Microbial Genomes Expert Review (IMG/ER) system (http://img.jgi.doe.gov/er) for comprehensive genome annotation (Mavromatis *et al.*, 2009). In brief, the system first detected non-coding RNA genes in the draft genomes using tRNAScan (Lowe and Eddy, 1997) and RNAmmer (Lagesen *et al.*, 2007) for tRNA and rRNA identification, respectively. Then, it validated other non-coding sequences by blasting them against a database containing known ncRNA genes. To identify protein-coding genes, it used either GeneMark (Lukashin and Borodovsky, 1998) or Metagene (Noguchi *et al.*, 2006) for prediction, compared results to protein families (e.g. COGs, Pfam) before product names were assigned.

Orthologous Gene and Gene Expression Analysis

Genes from *Rhodopseudomonas* strains often have different genomic locations and the sequences of orthologous genes are often not identical. Thus it was necessary to create gene-to-gene associations by categorizing genes from different strains into orthologous groups. This was accomplished with help from Sagar Utturkar, a graduate student at the University of Tennessee at Knoxville, who used OrthoMCL (Li *et al.*, 2003) to analyze the sixteen *Rhodopseudomonas* genomes. The process is explained briefly here. First, proteins shared among the sixteen strains were searched against the KEGG database, using an e-value cutoff of 1e-05. Next, putative orthologous relationships were identified from reciprocal best hits and the relationship information was then converted into a graph. Lastly, the Markov Cluster algorithm was applied to the graph to create groups of orthologous genes.

In the next step, the collective expression levels of genes in each orthologous group were determined. In brief, I wrote software to define the presence or absence of an orthologous group in each strain, and used Xpression (Phattarasukol *et al.*, 2012), see Chapter 2) to process *Rhodopseudomonas* RNA-seq data and calculate the expression levels of individual genes. Then, I wrote another software to map the expression levels of individual genes in Xpression outputs

to orthologous groups, and compute the sum of expression levels of all genes in an orthologous group in each strain. Next, genes that were similarly up-regulated and down-regulated in all *Rhodopseudomonas* strains were identified. To do this, I used DESeq (Anders and Huber, 2010) to identify orthologous groups whose expression levels were significantly changed. These are 1) genes that have expression levels of 10 or more and have expression ratios of 2.0 fold or more (*i.e.,* up-regulated), or 0.5 fold or less (*i.e.,* down-regulated) under two conditions, or 2) genes that have expression levels of 10 or more and have the statistical significance of expression changes (*i.e., p*-value) less than or equal to 0.05. Lastly, genes that met the criteria from each strain were listed, and only genes that were present in every list were selected for analysis.

RESULTS

General Genome Features of 14 *Rhodopseudomonas* Strains

The fourteen draft genomes that we constructed were submitted to the Integrated Microbial Genomes Expert Review (IMG/ER) system for comprehensive genome annotation, and results showed that the fourteen genomes are about 5.5 megabases in size, with the exception of strain 1a1, which has a slightly smaller genome of 5.3 megabases (Table 3.3). The number of protein-coding genes in each genome varies from 5,352 to 6,082, which is slightly more than what was predicted in the complete genome of strain CGA009 at 4,838. The total numbers of tRNAs and rRNAs are also similar, ranging from 64 to 73, compared to 80 in strain CGA009. All fourteen *Rhodopseudomonas* genomes have similarly high GC contents of approximately 65%.

Genetic and Transcriptomic Variations among *Rhodopseudomonas* Strains

To determine how many genes these *Rhodopseudomonas* strains have in common, we performed an orthologous analysis on the sixteen genomes. We found that the 88,340 total protein-coding genes from the sixteen strains could be categorized into 6,976 orthologous

groups, and that 3,999 groups (about 72% of the genes from each strain) are shared by all the sixteen strains, while only 26 groups (less than 1%) are strain-specific (Table 3.4). The strain-specific genes included mostly unknown genes and genes encoding for ribosomal protein L11 methylase in strain AP1, a chemotaxis signal transduction protein in strain BIS3, a mannosyltransferase in strain DSM8283, site-specific recombinase XerD in strain KD1, a nucleotide sugar dehydrogenase in strain RCH350, and P-loop ATPase in strain RSP24 (Table 3.5). As expected, the percentage of genes shared by the sixteen closely related strains (about 72% of the genes from each strain) is higher than the percentage of genes shared by the five more distantly related strains previously analyzed (about 55%) (Oda *et al.*, 2008). Overall, it appears that genes that are conserved among all strains support growth in diverse environments, and strain-specific genes are retained to take advantage of specific physical and chemical conditions in micro-environments. Figure 3.2 shows the comparative gene inventories of the sixteen strains, illustrating that a significant number of genes are conserved in each genome, while Figure 3.3 shows the portion of genes that absent in one or more strains, illustrating naturally occurring genetic variation among the sixteen strains.

As shown in Table 3.6, the seventeen strains (including strain CGA010) produced significantly different amounts H_2, and we hoped to identify the basis of the phenotypic variation and gain insight into H_2 production. To do this, I first examined the genetic and transcriptomic variations of the genes required for nitrogenase synthesis, as this is the enzyme that reduces nitrogen gas to ammonia and produces H_2 as an obligatory byproduct (Rey *et al.*, 2007). I found that a cluster of 32 genes for nitrogenase assembly, synthesis and activity (orthoMCL1301- 17, orthoMCL3172- 78 and orthoMCL3399-40) is completely conserved among the sixteen strains (Table 3.7). I also found that the genes were highly expressed in all strains under the H_2 producing (NF-high) condition compared to the non-H_2-producing (PM-high) condition, but there were significant differences in expression levels of nitrogenase genes among strains (Figure 3.4 and Table 3.8-3.12). For example, the expression of *nifH* gene, which encodes

for dinitrogenase reductase, was increased as much as 459.00 fold in strain CGA009 when grown under H_2-producing conditions but only 67.38 fold in strain RCH500 when grown under H_2-producing conditions, and the expression of *nifK* gene, which encodes for the beta-chain of dinitrogenase, was increased as much as 378.40 fold in strain CGA009 but only 37.70 fold in strain RCH500.

Furthermore, I compared the expression of nitrogenase genes under the H_2 producing, low light (NF-low) condition to that of the H_2 producing, high light (NF-high) condition (Table 3.12-3.16), and found that most strains (with the exception of strain RCH500) down-regulated the expression of nitrogenase genes, but the magnitudes of the expression changes were not as drastic as those under the NF-high condition compared to the PM-high condition. For example, the expression change of the *nifH* gene was at 0.20 (about 5 fold lower) in strain RSP24 and was at 0.64 (about 1.5 fold lower) in strain AP1, the expression change of the *nifK* gene was at 0.30 (about 3.3 fold lower) in strain RSP24, and at 0.65 (about 1.5 fold lower) in strain AP1. This could reflect that the strains grow more slowly in low light, and thus do not require high levels of nitrogenase synthesis to support growth.

To understand the relationship between the expression of nitrogenase genes and H_2 production, I examined the expression levels of nitrogenase gene and H_2 yields in each strain. I found that changes in nitrogenase gene expression do not consistently reflect changes in H_2 yields (Figure 3.5 A and B). For example, strain CGA009 increased the expression of the *nifH* gene 459.00 fold and the *nifK* gene 378.40 fold and produced 110.7 μmol of H_2/mg protein under the NF-high condition, while strain DCP3 increased the expression level of the *nifH* gene 209.77 fold and the *nifK* gene 228.00 fold (about two times lower than strain CGA009) and produced only 25.6 μmol of H_2/mg protein (about five times lower than strain CGA009) under the same condition. The discrepancy is likely due to the fact that H_2 production is a complex process that derives from the interplay of dozens of metabolic reactions, and thus it is important

to take into account the expression of other genes that might as well be important to H_2 production.

Genes Similarly Regulated in All *Rhodopseudomonas* Strains

To broaden the search for genes that might be important to H_2 production, I examined the collective expression levels of 6,976 orthologous groups and looked for genes that were up-regulated or down-regulated in all *Rhodopseudomonas* strains. I found that, in addition to genes from the nitrogenase cluster, additional genes from 22 orthologous groups were up-regulated in common under the NF-high condition compared to the PM-high condition (Table 3.16-3.19). These include genes involved in nitrogen assimilation (e.g. RPA4209: glutamine synthetase II) and transport (e.g. RPA0275: putative ammonium transporter, RPA2112: putative nitrate transporter component) as well as genes involved in electron transfer (e.g. RPA1928: ferredoxin-like protein [2Fe-2S], RPA2117: putative flavodoxin) and iron acquisition (e.g. RPA2124: *tonB* dependent iron siderophore receptor). This suggests that, when *Rhodopseudomonas* is starved of ammonia, it not only start to fix nitrogen but also attempts to scavenge fixed nitrogen from the environment. Interestingly, there are a number of unknown genes that were highly expressed in all strains (e.g. RPA1134: conserved hypothetical protein, RPA1927: hypothetical protein, RPA2156: hypothetical protein) but little is known about their functions. These unknown genes are good candidates to follow up on and look at the H_2 producing phenotypes of individual mutants.

I also found that genes from 15 orthologous groups were down-regulated in common under the NF-low condition compared to the NF-high condition (Table 3.20-3.23). These include genes with known functions (e.g. as RPA2117: putative flavodoxin, RPA2118: putative ATP-binding protein of ABC transporter, RPA2124: *tonB* dependent iron siderophore receptor, RPA3477: *exbD*, uptake of enterochelin) as well as a number of hypothetical genes. This might reflect the decreasing needs for resources to support growth, as cells grown in low light tend to

have significantly lower growth rates than cells grown in high light. It should be noted that no gene was down-regulated in common under the NF-high condition compared to the PM-high condition, and none was up-regulated in common under the NF-low compared to the NF-high condition. As shown in Figure 3.6 A and B, as the numbers of *Rhodopseudomonas* strains included in the analysis increased, the number of genes that were down-regulated in common under the NF-high condition compared to the PM-high condition and up-regulated in common under the NF-low compared to the NF-high condition decreased and eventually became zero. The lack of similarly regulated genes under those conditions likely reflects the differences in regulatory strategies each strain used in response to environmental changes.

DISCUSSION

In this chapter, I have shown that draft genomes assembled from Illumina GA-II sequencing data, which consisted of relatively shorter sequences but with higher number of reads, and from Illumina HiSeq sequencing data, which consisted of relatively longer sequences but with lower number of reads for genome assembly, are comparable in their quality. Although draft genome assembled from Illumina GA-II sequencing data were composed of a smaller number of scaffolds and had significantly larger scaffolds than draft genomes assembled from Illumina HiSeq data, both sets of draft genomes have more than 90% of predicted protein sequences matched (with very high percentage of sequence identity) to known protein sequences. This implies that there were very few assembly errors that could contribute to false annotation in both sets of draft genomes. In some cases, it may be preferable to have a draft genome consisting of as few scaffolds as possible, for example, when one is interested in inspecting regulatory components intergenic regions. In our case, however, we are interested in examining the presence or absence of orthologous genes and their expression levels, and thus having the least number of scaffolds is not as important as having accurately predicted genes.

I have also shown that the sixteen strains of *Rhodopseudomonas* are closely related, as predicted by the similarity in 16S rRNA sequences. The majority of genes are conserved in all strains and less than 1% of the genes are specific to certain strains. Considering that these strains were isolated from different geographical locations, it is interesting that they have so many core genes in common. Nonetheless, about one third of the genes are absent in one or more strains, reflecting the activity of gene loss (or acquisition), which is a major adaptive mechanism for bacteria to take advantage of physical and chemical conditions in the environment. As expected, a cluster of 32 genes required for synthesis and regulation of nitrogenase is completely conserved in all strains, reflecting the important role of nitrogenase in supporting growth of diazotrophic organisms like *Rhodopseudomonas*. I found that the nitrogenase genes were highly expressed under H_2 producing conditions. This is expected because nitrogenase is a relatively slow enzyme with a turnover time of ~5 s^{-1} (Dixon and Kahn, 2004) and thus needs to be synthesized in the large amounts to meet the demand for fixed nitrogen of growing cells. I also found that the expression of nitrogenase genes were lower under the NF-low condition compared to the NF-high condition, which likely reflects the slower growth of cells in low light condition.

Importantly, I observed that the elevated levels of nitrogenase gene expression do not consistently reflect changes in H_2 yields, as strains that expressed nitrogenase genes at higher levels did not always have higher H_2 yields at the same. This is likely due to the complexity of H_2 production, which derives from the interplay of dozens of metabolic reactions, not merely the expression of nitrogenase genes or the function of nitrogenase enzyme. Therefore, I broadened my analysis for genes important to H_2 production by looking for genes that were similarly up-regulated or down-regulated in all strains under H_2 producing conditions. I found that under the NF-high condition compared to the PM-high condition all strains not only up-regulated genes encoding for nitrogenase synthesis and regulation (as expected) but also genes involved in nitrogen assimilation and transport as well as genes involved in electron transfer and iron

acquisition. These reflect activities that *Rhodopseudomonas* likely needs to engage while growing under nitrogen-fixing conditions. These includ fixing nitrogen, assimilating fixed nitrogen into cell material, and scavenging resources (e.g. iron and fixed nitrogen in the environment) to support growth. In contrast, I found that under the nitrogen-fixing, low light condition compared to the nitrogen-fixing, high light condition, all strains seem to focus more on cutting expenses. This is likely because energy generation is limited in this condition, and thus being industrious and clever in managing resources is important to sustain growth. Also, it is surprising that a number of commonly up-regulated and down-regulated genes are hypothetical (listed in Table 3.24-3.25). These genes may be important to *Rhodopseudomonas* growth under those conditions, and are good candidates to follow up with experiments with mutants to see if there is any defect in growth or H_2 yields.

Lastly, it is important to note that the list of genes identified as important to H_2 production that I presented here likely represent a small number of factors that affect the ability of cells to generate H_2. Therefore, it is important to examine the differences in gene content and gene expression at a systems level, so that genes in peripheral metabolic modules that are nevertheless key drivers of H_2 production will not be missed. This is the topic of the next chapter of this thesis.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Chugani, S., Kim, B.S., Phattarasukol, S., Brittnacher, M.J., Choi, S.H., Harwood, C.S., and Greenberg, E.P. (2012). Strain-dependent diversity in the *Pseudomonas aeruginosa* quorum-sensing regulon. Proc. Natl. Acad. Sci. U.S.A. *109*, E2823–E2831.

Das, D., and Veziroglu, T.N. (2001). Hydrogen production by biological processes: a survey of literature. Int J Hydrogen Energy *26*, 13–28.

Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. *2*, 621–631.

Hirakawa, H., Oda, Y., Phattarasukol, S., Armour, C.D., Castle, J.C., Raymond, C.K., Lappala, C.R., Schaefer, A.L., Harwood, C.S., and Greenberg, E.P. (2011). Activity of the *Rhodopseudomonas palustris p*-coumaroyl-homoserine lactone-responsive transcription factor RpaR. J. Bacteriol. *193*, 2598–2607.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.

Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., *et al.* (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat. Biotechnol. *22*, 55–61.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

Lukashin, A.V., and Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. *26*, 1107–1115.

Mavromatis, K., Ivanova, N.N., Chen, I.-M.A., Szeto, E., Markowitz, V.M., and Kyrpides, N.C. (2009). The DOE-JGI Standard Operating Procedure for the Annotations of Microbial Genomes. Stand Genomic Sci *1*, 63–67.

Noguchi, H., Park, J., and Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res. *34*, 5623–5630.

Oda, Y., Larimer, F.W., Chain, P.S.G., Malfatti, S., Shin, M.V., Vergez, L.M., Hauser, L., Land, M.L., Braatsch, S., Beatty, J.T., *et al.* (2008). Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. Proc. Natl. Acad. Sci. U.S.A. *105*, 18543–18548.

Oda, Y., Samanta, S.K., Rey, F.E., Wu, L., Liu, X., Yan, T., Zhou, J., and Harwood, C.S. (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. J. Bacteriol. *187*, 7784–7794.

Oda, Y., Wanders, W., Huisman, L.A., Meijer, W.G., Gottschal, J.C., and Forney, L.J. (2002). Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments. Appl. Environ. Microbiol. *68*, 3467–3477.

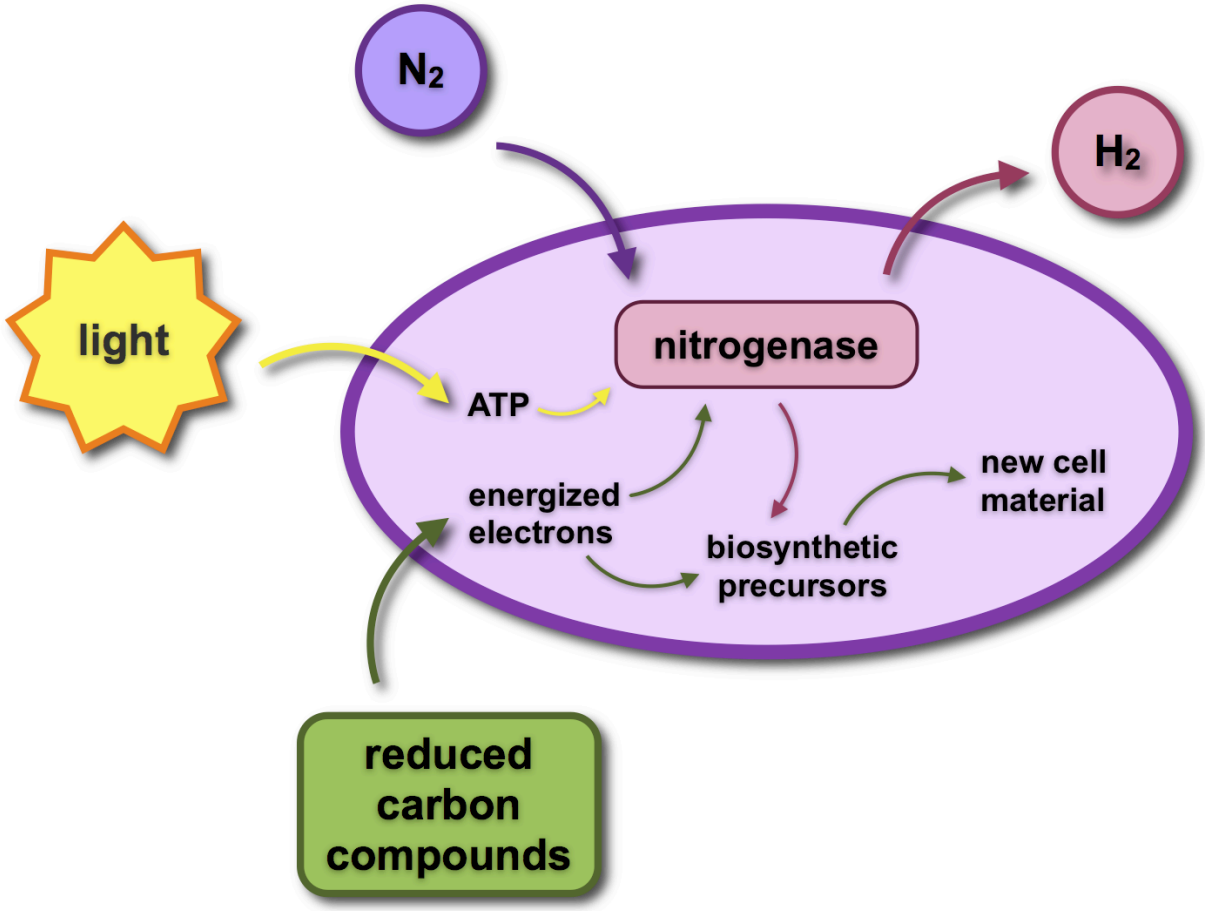
Phattarasukol, S., Radey, M.C., Lappala, C.R., Oda, Y., Hirakawa, H., Brittnacher, M.J., and Harwood, C.S. (2012). Identification of a *p*-coumarate degradation regulon in *Rhodopseudomonas palustris* by Xpression, an integrated tool for prokaryotic RNA-seq data processing. Appl. Environ. Microbiol. *78*, 6812–6818.

Rey, F.E., Heiniger, E.K., and Harwood, C.S. (2007). Redirection of metabolism for biological hydrogen production. Appl. Environ. Microbiol. *73*, 1665–1671.

Rey, F.E., Oda, Y., and Harwood, C.S. (2006). Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. J. Bacteriol. *188*, 6143–6152.

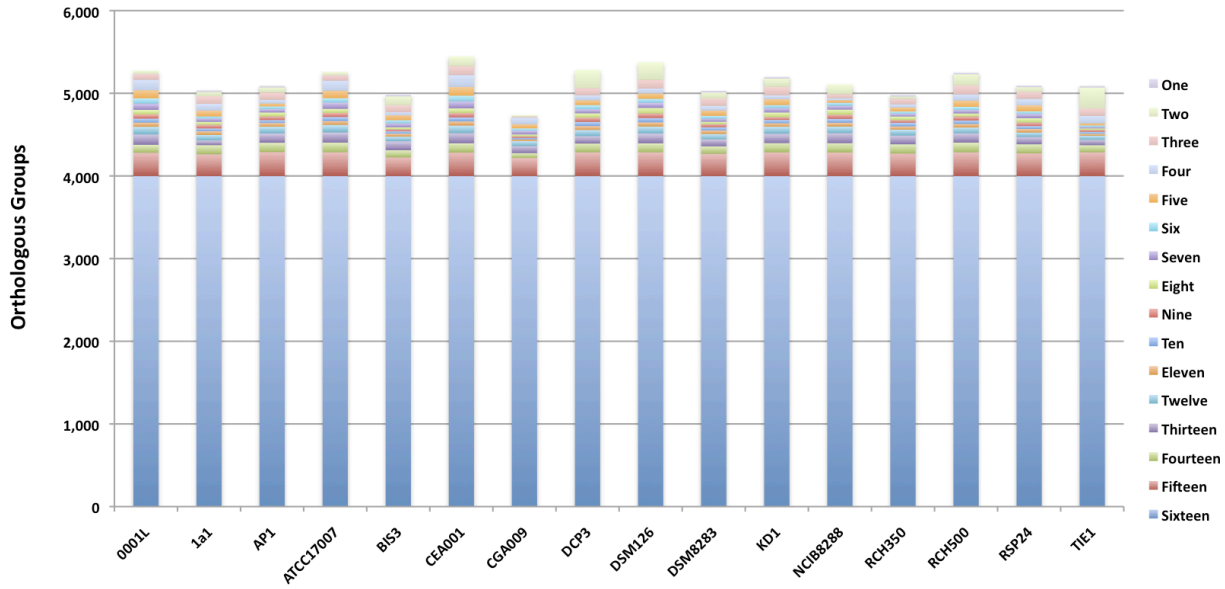
Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

Figure 3.1. H_2 production by *Rhodopseudomonas* requires the integration of dozens of metabolic reactions.



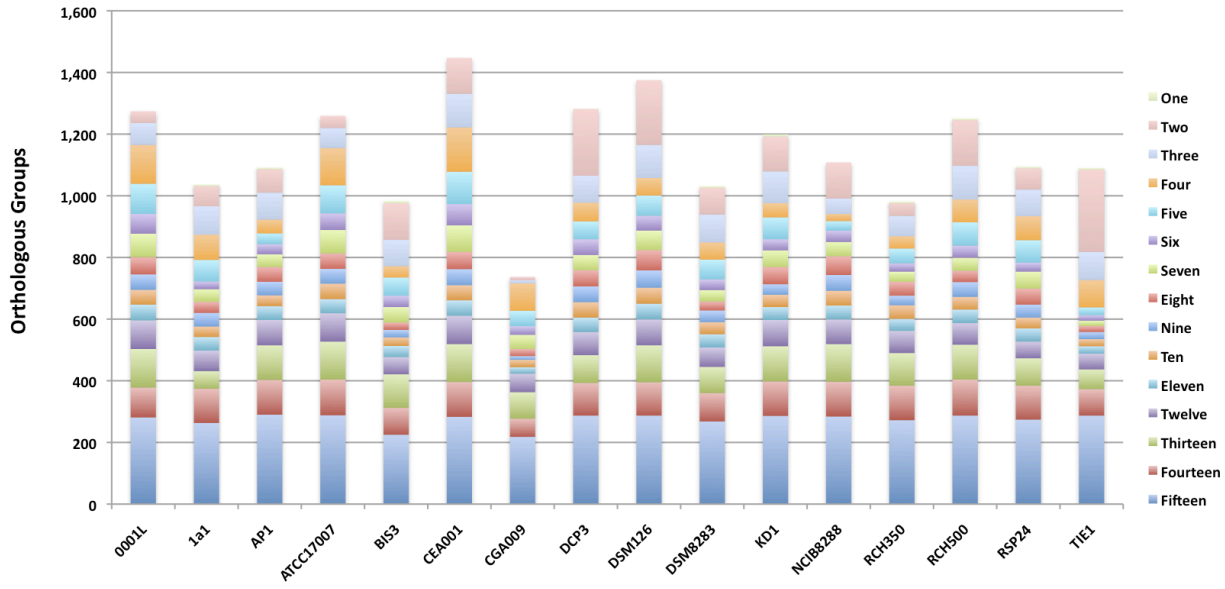
For example, the metabolic modules of photophosporylation to generate ATP from light, carbon metabolism to generate reduced electrons, and nitrogen metabolism to generate nitrogenase, the enzyme that produces H_2.

Figure 3.2. Graph depicted numbers of orthologous groups shared by different combination of *Rhodopseudomonas* strains.



The genome sequences of sixteen closely related strains of *Rhodopseudomonas* reveal that 3,999 orthologous groups are shared by all strains (the blue boxes on the bottom this graph). Other colored boxes depict the number of orthologous groups shared by different combinations of strains, as noted in the color key on the right side of the graph.

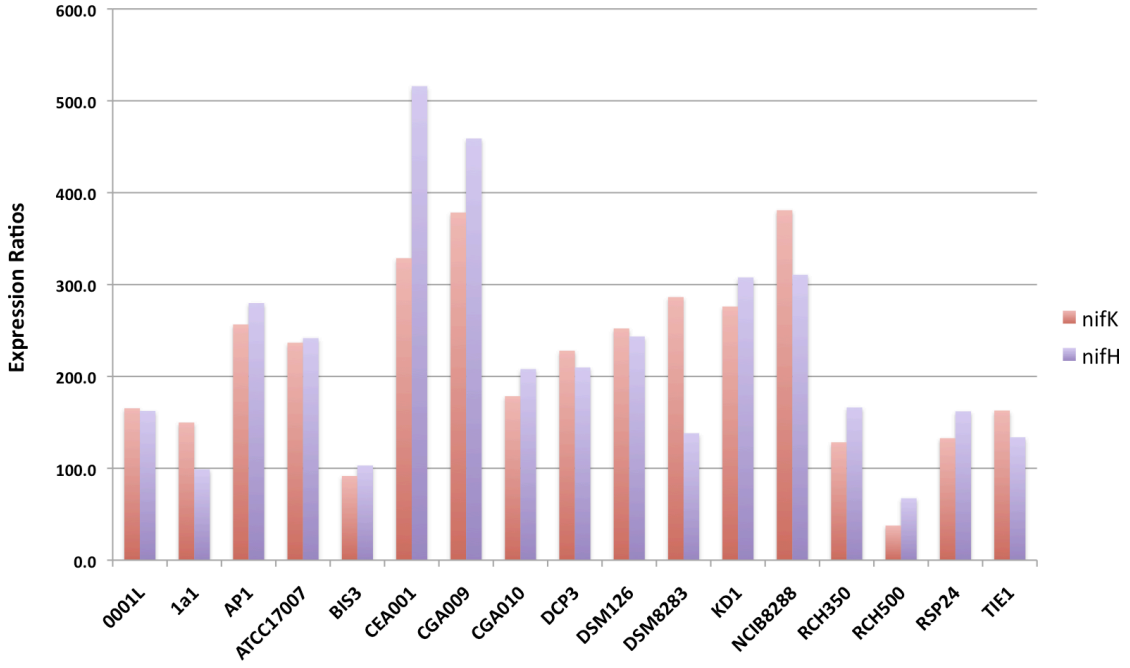
Figure 3.3. Graph depicted numbers of orthologous groups not shared by all but found in different combination of *Rhodopseudomonas* strains.



Orthologous groups shared by fifteen strains are shown in blue, fourteen strains in red, and so on, as noted in the color key on the right side of this graph. The boxes for orthologous groups present in one strain are not visible in this graph because very few genes are strain-specific. See Table 3.5 for a list of these genes.

Figure 3.4. Expression ratios of the *nifHK* genes vary among closely related strains of *Rhodopseudomonas*.

A) Under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.



B) Under the H_2-producing, low light condition (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

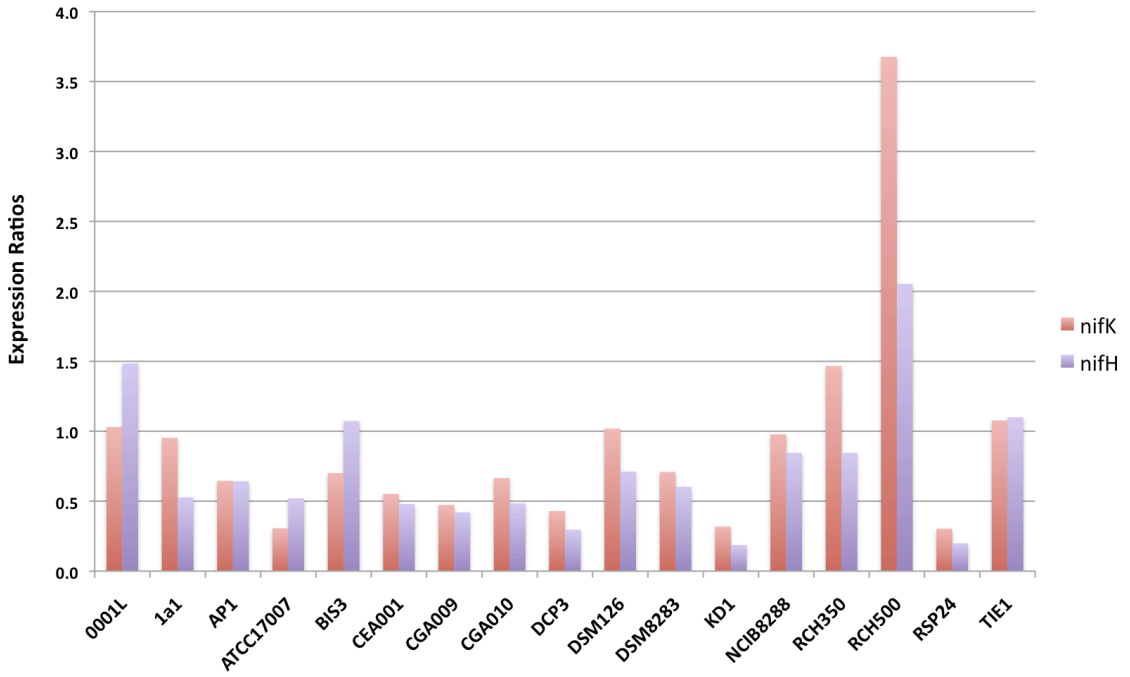
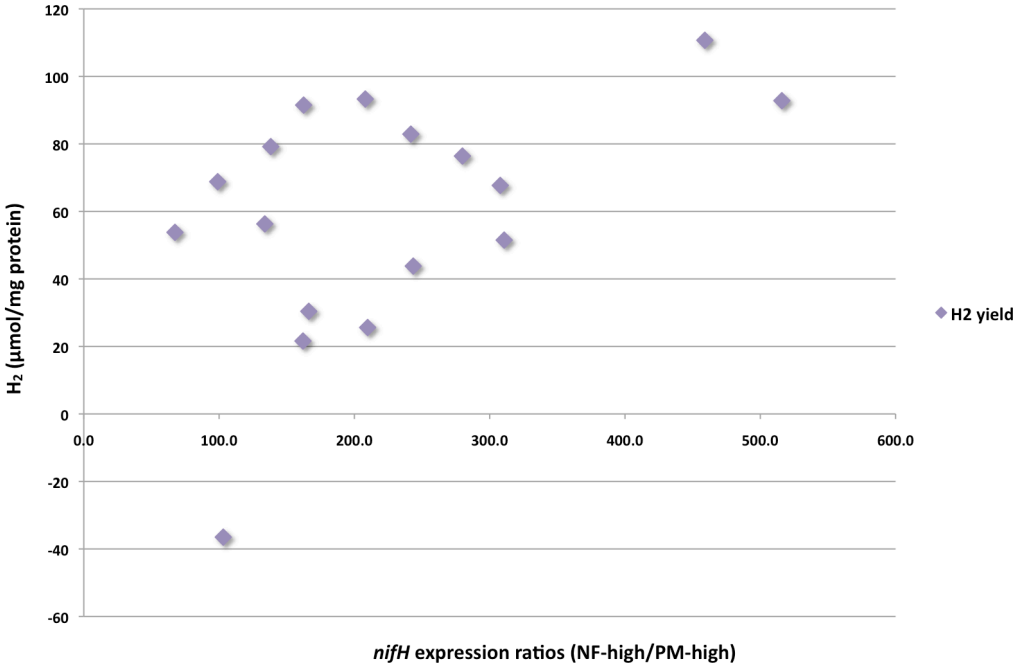


Figure 3.5. Expression ratios of the *nifH* gene does not reflect H_2 yield.

A) The correlation coefficient between H_2 yield and the expression ratios of the *nifH* gene between the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition is 0.49 (*p*-value = 0.04).



B) The correlation coefficient between H_2 yield and the expression ratios of the *nifH* gene between the H_2-producing, low light condition (NF-low) condition compared to the H_2-producing, high light (NF-high) condition is -0.11 (*p*-value = 0.67).

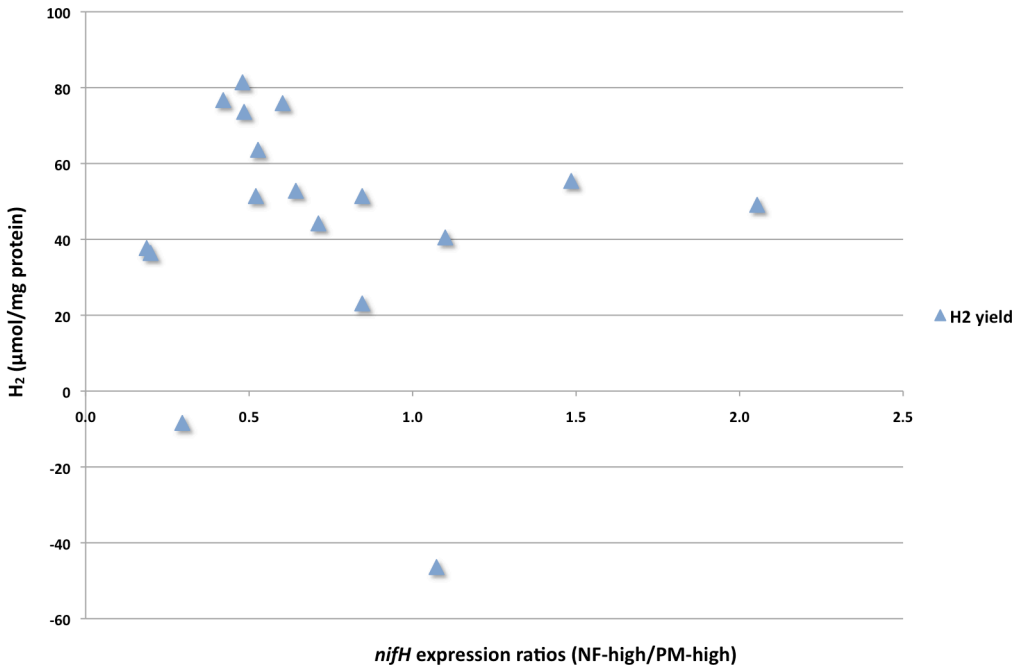
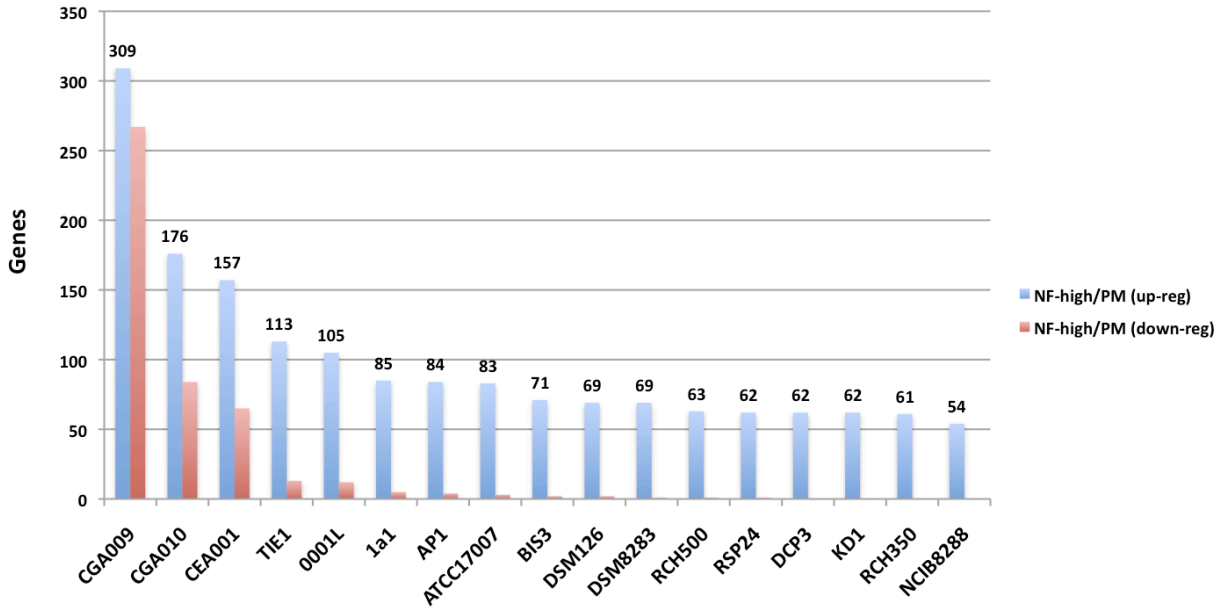
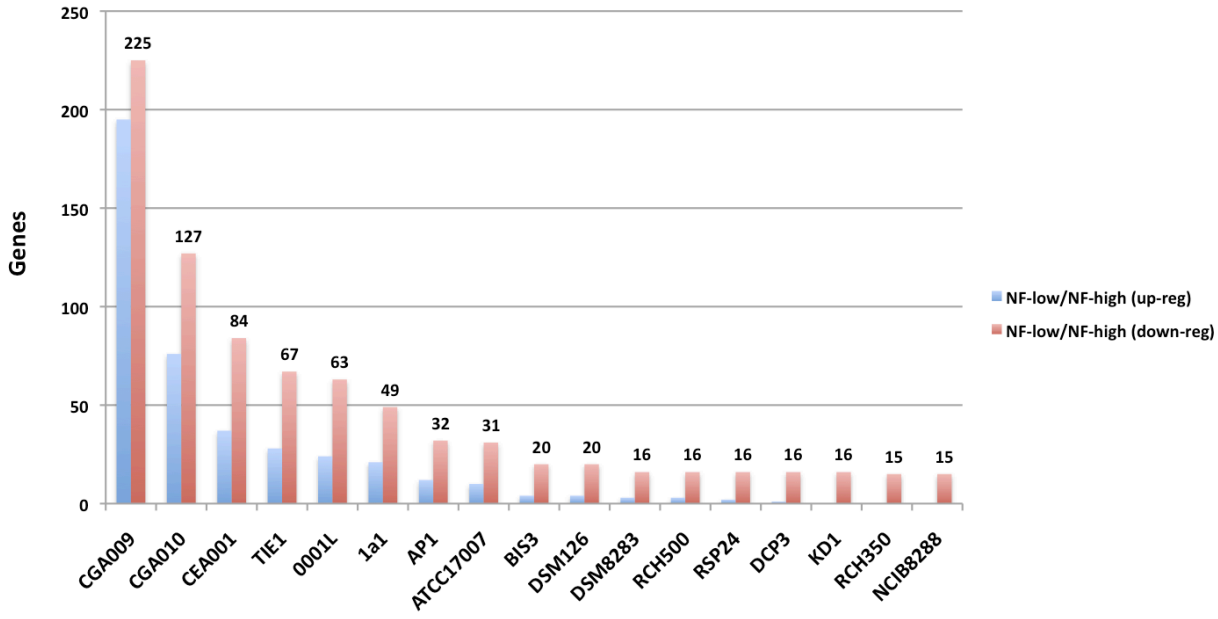


Figure 3.6. Number of genes that were up-regulated or down-regulated in all *Rhodopseudomonas* strains.

A) Under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.



B) Under the H_2-producing, low light condition (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.



As the numbers of strains included in the analysis increases, the number of genes that were up-regulated or down-regulated in all strains decreases.

Table 3.1. Origins of sixteen *Rhodopseudomonas* strains and percentages of 16S rRNA sequence identity relative to strain CGA009.

Name	16S rRNA Identity (%)	Isolation Country
0001L	100.0	Woods Hole, MA, USA
1a1	100.0	Bonn, Germany
AP1	100.0	Appelbergen, Netherlands
ATCC17007	100.0	Unknown
BIS3	100.0	Biesbosch, Netherlands
CEA001	100.0	Ithaca, NY, USA
CGA009	100.0	Ithaca, NY, USA
DCP3	99.9	Biesbosch, Netherlands
DSM126	100.0	Unknown
DSM8283	100.0	Japan
KD1	99.9	Haren, Netherlands
NCIB8288	99.8	Unknown
RCH350	99.9	Woods Hole, MA, USA
RCH500	100.0	Woods Hole, MA, USA
RSP24	100.0	Woods Hole, MA, USA
TIE-1	100.0	Woods Hole, MA, USA

Table 3.2. Statistics of fourteen draft genomes assembled from Illumina GA-II and Illumina HiSeq sequencing data.

Strain	Sequencer	Number of Scaffolds	Size of Largest Scaffolds	Number of Predicted Genes	Number of Matched Proteins	Percentage of Matched Proteins
1a1	GA-II	89	745,665	4,987	4,793	96.11
BIS3	GA-II	172	522,820	5,286	4,848	91.71
CEA001	GA-II	61	810,898	5,014	4,925	98.22
0001L	HiSeq	340	138,925	5,298	5,093	96.13
1a1	HiSeq	372	181,712	5,323	5,014	94.20
AP1	HiSeq	253	188,880	5,313	4,848	91.25
ATCC17007	HiSeq	344	136,943	5,346	5,143	96.20
BIS3	HiSeq	222	257,548	5,259	4,761	90.53
CEA001	HiSeq	860	81,546	5,768	5,512	95.56
DCP3	HiSeq	367	166,652	5,470	4,965	90.77
DSM126	HiSeq	617	105,685	5,464	5,046	92.35
DSM8283	HiSeq	213	178,335	5,135	4,668	90.91
KD1	HiSeq	203	219,230	5,342	4,920	92.10
NCIB8288	HiSeq	208	168,701	5,126	4,846	94.54
RCH350	HiSeq	332	128,740	5,237	4,781	91.29
RCH500	HiSeq	308	161,003	5,356	4,879	91.09
RSP24	HiSeq	276	159,769	5,341	4,929	92.29

Gray rows are draft genomes assembled from Illumina GA-II sequencing data by Somsak Phattarasukol and white rows are draft genomes assembled from Illumina HiSeq sequencing data by Matthew Ready.

Table 3.3. Gene information of fourteen *Rhodopseudomonas* strains as predicted by the Integrated Microbial Genomes Expert Review (IMG/ER) system.

Strain	Genome Size	No. Genes[a]	No. CDSs[b]	% CDSs	No. RNAs[c]	% RNAs	No. rRNA[d]	% GC
0001L	5,452,064	5,622	5,549	98.70	73	1.30	8	65.00
1a1	5,310,758	5,436	5,370	98.79	66	1.21	3	65.00
AP1	5,482,873	5,633	5,558	98.67	75	1.33	5	65.00
ATCC17007	5,493,281	5,676	5,606	98.77	70	1.23	8	65.00
BIS3	5,456,053	5,551	5,485	98.81	66	1.19	3	65.00
CEA001	5,417,108	6,082	6,015	98.90	67	1.10	4	65.00
CGA009[*]	5,467,640	4,918	4,838	98.37	80	1.63	6	65.03
DCP3	5,510,482	5,874	5,809	98.89	65	1.11	3	65.00
DSM126	5,327,492	5,918	5,852	98.88	66	1.12	4	65.00
DSM8283	5,343,628	5,629	5,565	98.86	64	1.14	3	65.00
KD1	5,553,564	5,617	5,551	98.82	66	1.18	3	65.00
NCIB8288	5,349,750	5,352	5,287	98.79	65	1.21	3	65.00
RCH350	5,420,144	5,498	5,432	98.80	66	1.20	3	65.00
RCH500	5,468,780	5,715	5,647	98.81	68	1.19	3	65.00
RSP24	5,453,366	5,525	5,458	98.79	67	1.21	5	65.00
TIE-1[*]	5,744,041	5,377	5318	98.90	59	1.34	6	64.86

[*] Complete genome
[a] Number of all predicted genes, including protein-coding genes and non-protein coding genes.
[b] Number of protein-coding genes.
[c] Number of rRNA and tRNA genes.
[d] Number of 5S, 16S and 23S rRNA genes.

Table 3.4. Number of orthologous groups shared by *Rhodopseudomonas* strains.

Shared By Strains	Number of Orthologous Groups
1	26
2	878
3	436
4	292
5	204
6	105
7	116
8	89
9	72
10	64
11	61
12	98
13	126
14	118
15	292
16	3,999
Total	6,976

Table 3.5. List of strain-specific orthologous genes.

Strain	orthoMCL ID	Product Description
1a1	orthoMCL6932	hypothetical protein
AP1	orthoMCL5918	Ribosomal protein L11 methylase
BIS3	orthoMCL6707	Chemotaxis signal transduction protein
BIS3	orthoMCL6716	hypothetical protein
BIS3	orthoMCL6722	Glycosyltransferases involved in cell wall biogenesis
BIS3	orthoMCL6744	hypothetical protein
BIS3	orthoMCL6745	Predicted ATPase
BIS3	orthoMCL6750	ATP-dependent metalloprotease FtsH(EC:3.4.24.-)
BIS3	orthoMCL6752	hypothetical protein
DSM8283	orthoMCL5701	Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components
DSM8283	orthoMCL6246	Mannosyltransferase OCH1 and related enzymes
DSM8283	orthoMCL6258	hypothetical protein
DSM8283	orthoMCL6259	hypothetical protein
KD1	orthoMCL6212	hypothetical protein
KD1	orthoMCL6213	hypothetical protein
KD1	orthoMCL6223	hypothetical protein
KD1	orthoMCL6225	hypothetical protein
KD1	orthoMCL6226	hypothetical protein
KD1	orthoMCL6229	hypothetical protein
KD1	orthoMCL6230	Site-specific recombinase XerD
RCH350	orthoMCL6168	nucleotide sugar dehydrogenase(EC:1.1.1.-)
RCH500	orthoMCL6159	Protein of unknown function (DUF2971).
RSP24	orthoMCL6086	hypothetical protein
RSP24	orthoMCL6090	Predicted P-loop ATPase and inactivated derivatives
RSP24	orthoMCL6091	Uncharacterized conserved protein
TIE-1	orthoMCL6085	hypothetical protein

Table 3.6. Nitrogenase activities and H_2 yields of seventeen *Rhodopseudomonas* strains grown under H_2-producing, high light (NF-high) and H_2-producing, low light (NF-low) conditions.

Strain	H_2 Yield under NF-high Condition[a]	Nitrogenase Activity under NF-high Condition[b]	H_2 Yield under NF-low Condition[a]	Nitrogenase Activity under NF-low Condition[b]
0001L	91.5	109.9	55.4	67.1
1a1	68.8	124.5	63.6	48.9
AP1	76.4	116.5	52.8	87.3
ATCC17007	82.9	69.8	51.4	28.6
BIS3	-36.5[*]	36.7	-46.4[*]	24.9
CEA001	92.8	69.6	81.4	35.4
CGA009	110.7	107.7	76.7	82.1
CGA010	93.3	97.8	73.6	84.3
DCP3	25.6	67.2	-8.4[*]	35.4
DSM126	43.8	72.8	44.2	70.7
DSM8283	79.2	94.3	75.9	66.8
KD1	67.7	110.2	37.7	57.6
NCIB8288	51.5	96.7	51.4	66.1
RCH350	30.4	52	23.1	41.4
RCH500	53.8	76.1	49.1	83.9
RSP24	21.6	64.3	36.4	37.5
TIE-1	56.3	90.8	40.5	32.1

Data collected by Dr. Yasuhiro Oda and Colin Lappala, a research scientist and research assistant in the Harwood laboratory

[a] H_2 yield in μmol/mg protein.

[b] Nitrogenase activity in nmol C_2H_4 formed/min/mg protein.

[*] H_2 yield is negative because *Rhodopseudomonas* recaptured H_2 that it produced to reuse the electrons in H_2 in metabolic reactions (see Rey *et al.*, 2006).

Table 3.7. A cluster of 32 genes for nitrogenase assembly, synthesis and activity is completely conserved among sixteen *Rhodopseudomonas* strains.

orthoMCL ID	0001L	1a1	CGA009	AP1	ATCC17007	BIS3	CEA001	DCP3	DSM126	DSM8283	KD1	NCIB8288	RCH350	RCH500	RSP24	TIE-1	RPA Number[#]	Product Description
orthoMCL3400	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4607	putative homocitrate synthase
orthoMCL3176	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4609	putative nifU protein
orthoMCL3178	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4613	DUF683
orthoMCL1314	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4614	DUF269
orthoMCL1313	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE
orthoMCL1310	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4621	conserved hypothetical protein
orthoMCL1306	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4622	hypothetical protein
orthoMCL1305	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4623	conserved hypothetical protein
orthoMCL1304	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4624	hypothetical protein
orthoMCL1303	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4625	NifZ domain
orthoMCL1302	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4626	Protein of unknown function from Deinococcus and Synechococcus

orthoMCL1301	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4627	conserved hypothetical protein
orthoMCL1300	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RPA4633	short-chain dehydrogenase

‡ *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.8. Nitrogenase gene expression in strains 0001L, 1a1, CGA009, CGA010 and AP1 under the H_2-producing, high light condition (NF-high) condition and the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	0001L[a]	0001L[b]	0001L[c]	1a1[a]	1a1[b]	1a1[c]	CGA009[a]	CGA009[b]	CGA009[c]	CGA010[a]	CGA010[b]	CGA010[c]	AP1[a]	AP1[b]	AP1[c]	RPA Number[#]	Product Description
orthoMCL3400	3,047	21	127.08	3,152	117	26.29	7,597	30	230.30	2,863	27	95.53	6,780	57	113.05	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	666	6	74.33	1,440	26	49.76	1,658	15	92.28	776	9	64.92	1,506	8	137.18	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	320	5	40.38	1,251	10	96.46	1,500	5	187.88	975	6	108.67	911	5	114.25	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	358	4	51.57	1,686	20	73.43	1,608	5	201.38	1,233	6	137.33	936	5	117.38	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	504	9	42.25	1,183	58	19.44	1,195	19	54.45	657	10	50.77	896	18	42.81	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	727	6	81.11	1,679	33	46.72	1,432	7	143.50	1,029	7	103.20	1,807	14	106.47	RPA4607	putative homocitrate synthase
orthoMCL3176	105	2	21.60	410	6	45.89	260	2	52.60	171	3	29.00	224	3	37.83	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	161	3	27.33	713	16	37.68	897	5	112.50	467	4	67.14	227	4	32.86	RPA4609	putative nifU protein
orthoMCL3178	149	3	25.33	817	27	27.33	656	2	131.80	490	7	49.30	333	13	21.00	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	523	7	52.60	321	9	27.00	987	9	82.50	416	13	26.19	912	8	83.18	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	515	8	47.09	704	19	32.14	758	10	58.54	453	11	32.57	836	5	104.88	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1,566	14	92.29	1,245	23	48.00	2,362	14	139.12	1,064	21	44.46	1,970	10	151.77	RPA4613	DUF683
orthoMCL1314	1,468	32	42.03	1,570	55	27.12	1,786	29	55.91	1,074	40	25.05	1,208	10	93.15	RPA4614	DUF269
orthoMCL1313	1,810	34	49.00	1,649	47	33.04	3,476	32	99.40	1,591	33	44.28	3,546	73	46.70	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	362	4	52.14	286	15	16.06	612	5	76.88	314	6	35.22	813	6	90.67	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	764	9	63.92	822	22	33.00	1,343	9	112.17	677	11	48.57	1,099	8	100.18	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	2,807	14	165.29	3,144	18	149.86	5,673	12	378.40	3,031	14	178.47	4,102	13	256.56	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	2,729	95	27.88	3,082	37	77.13	5,430	75	69.65	2,829	27	94.40	2,884	46	58.92	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	3,409	18	162.48	4,846	46	98.96	7,341	13	459.00	4,366	18	208.05	5,034	15	279.83	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	554	8	50.64	588	44	12.57	721	9	60.33	462	13	29.06	616	9	51.58	RPA4621	conserved hypothetical protein
orthoMCL1306	512	13	32.19	553	17	27.80	972	11	69.64	492	12	33.00	508	6	56.78	RPA4622	hypothetical protein
orthoMCL1305	597	15	33.33	816	33	22.75	843	12	56.40	526	17	26.45	1,167	7	117.00	RPA4623	conserved hypothetical protein
orthoMCL1304	364	8	33.36	765	24	28.44	723	5	90.75	385	6	43.11	712	6	79.44	RPA4624	hypothetical protein
orthoMCL1303	643	9	53.83	764	25	27.39	1,279	10	98.62	738	6	82.33	789	8	72.00	RPA4625	NifZ domain
orthoMCL1302	538	10	41.62	794	29	24.91	1,057	9	88.33	585	8	53.45	682	8	62.27	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	196	3	33.17	387	12	26.00	435	2	87.60	276	3	46.50	375	3	63.00	RPA4627	conserved hypothetical protein
orthoMCL1300	227	3	38.33	434	15	24.28	461	4	66.29	329	2	66.40	387	4	55.71	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	379	6	42.44	893	14	52.71	655	6	73.11	539	4	77.43	765	4	109.71	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	186	4	27.00	859	33	23.94	659	5	82.75	411	7	41.40	477	9	40.00	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	17,889	94	184.45	44,927	1,013	44.22	58,179	79	709.54	35,435	80	426.96	43,630	131	325.62	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	914	185	4.88	914	331	2.75	1,730	150	11.33	1,319	205	6.36	983	146	6.62	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	71	4	10.57	514	37	12.93	584	8	53.36	765	16	40.42	147	6	16.67	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[ǂ] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.9. Nitrogenase gene expression in strains ATCC17007, BIS3, CEA001, DCP3 and DSM126 under the H_2-producing, high light condition (NF-high) condition and the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	ATCC17007[a]	ATCC17007[b]	ATCC17007[c]	BIS3[a]	BIS3[b]	BIS3[c]	CEA001[a]	CEA001[b]	CEA001[c]	DCP3[a]	DCP3[b]	DCP3[c]	DSM126[a]	DSM126[b]	DSM126[c]	RPA Number[#]	Product Description
orthoMCL3400	3,034	20	132.04	2,191	82	25.81	2,824	28	91.19	4,271	38	104.24	3,446	30	104.52	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	777	12	52.00	628	17	31.55	1,205	10	92.92	1,220	16	64.37	1,270	10	97.92	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	596	6	66.56	275	4	39.71	1,180	5	147.88	1,252	4	179.29	1,564	7	156.70	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	720	3	120.50	238	5	30.13	1,381	4	197.71	853	6	95.11	1,406	5	176.13	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	667	9	55.83	729	12	48.80	848	10	65.46	1,184	15	65.94	1,013	11	72.57	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	848	8	77.36	1,071	10	82.62	1,060	4	151.86	1,303	7	130.60	947	7	95.00	RPA4607	putative homocitrate synthase
orthoMCL3176	125	3	21.33	127	3	21.67	282	2	57.00	366	3	61.50	324	4	46.71	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	304	3	51.17	92	4	13.57	568	3	95.17	367	12	24.67	687	4	98.57	RPA4609	putative nifU protein
orthoMCL3178	251	8	23.09	153	6	17.33	554	5	69.63	289	6	32.44	682	5	85.63	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	764	13	47.94	218	8	20.09	1,272	9	106.25	837	13	52.50	956	13	59.94	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	527	8	48.18	288	12	19.40	916	6	102.11	497	9	41.67	593	6	66.22	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1,392	18	66.43	766	13	48.06	1,913	14	112.71	1,092	10	84.23	1,379	17	69.10	RPA4613	DUF683
orthoMCL1314	1,106	17	55.45	1,002	33	27.92	1,557	14	91.76	682	14	40.29	1,517	69	21.11	RPA4614	DUF269
orthoMCL1313	2,299	41	52.32	641	14	37.88	3,333	29	104.25	2,657	103	25.09	2,063	61	32.28	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	413	5	52.00	216	4	31.29	582	4	83.57	516	7	51.90	605	6	67.56	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	1,125	12	75.20	376	9	31.58	1,104	9	92.25	598	12	40.07	1,158	12	77.40	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	3,311	11	236.71	1,097	9	91.67	4,599	11	328.71	2,733	9	228.00	4,537	15	252.22	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	3,141	76	39.80	1,437	47	28.80	5,310	26	183.21	3,301	20	143.65	3,463	15	192.56	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	4,348	15	241.72	1,337	10	103.08	8,252	13	515.94	2,724	10	209.77	4,380	15	243.50	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	606	14	35.82	366	13	23.06	653	11	46.86	405	14	24.00	509	10	39.38	RPA4621	conserved hypothetical protein
orthoMCL1306	840	8	76.64	287	12	19.33	794	14	46.88	424	15	23.72	408	8	37.36	RPA4622	hypothetical protein
orthoMCL1305	669	12	44.80	936	22	37.56	679	10	52.46	1,003	15	55.89	1,596	11	114.21	RPA4623	conserved hypothetical protein
orthoMCL1304	890	5	111.63	481	10	37.23	616	5	77.38	942	14	55.59	805	6	89.78	RPA4624	hypothetical protein
orthoMCL1303	943	7	94.60	487	9	40.83	1,048	6	116.78	475	6	53.11	799	5	100.25	RPA4625	NifZ domain
orthoMCL1302	966	9	80.75	239	31	7.12	886	7	88.90	841	6	93.78	820	9	68.58	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	229	5	29.00	169	4	24.57	314	3	52.83	492	4	70.71	431	2	86.80	RPA4627	conserved hypothetical protein
orthoMCL1300	165	3	28.00	247	7	25.00	282	2	57.00	486	2	97.80	399	3	67.00	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	437	11	31.43	251	3	42.33	857	8	78.18	667	7	67.00	478	3	80.17	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	321	8	29.45	128	5	16.38	630	5	79.13	651	6	72.67	562	6	62.78	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	23,577	57	393.00	16,148	94	166.51	17,934	34	484.78	18,332	40	426.40	38,387	89	417.28	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	594	151	3.88	768	253	3.01	781	131	5.85	601	113	5.21	827	128	6.34	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	115	5	14.75	36	15	2.17	201	7	20.40	251	13	15.88	265	7	26.80	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.10. Nitrogenase gene expression in strains DSM8283, KD1, NCIB8288, RCH350 and RCH500 under the H_2-producing, high light condition (NF-high) condition and the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	DSM8283[a]	DSM8283[b]	DSM8283[c]	KD1[a]	KD1[b]	KD1[c]	NCIB8288[a]	NCIB8288[b]	NCIB8288[c]	RCH350[a]	RCH350[b]	RCH350[c]	RCH500[a]	RCH500[b]	RCH500[c]	RPA Number[d]	Product Description
orthoMCL3400	2,258	15	125.61	4,765	32	136.23	2,440	16	128.58	2,236	82	26.34	709	31	20.94	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	871	7	87.40	2,202	15	122.50	1,198	8	109.18	587	18	28.10	220	10	17.15	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	733	3	122.67	1,600	5	200.38	1,238	5	155.13	664	7	66.70	326	6	36.56	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	703	2	141.20	1,158	4	165.86	1,061	3	177.33	864	4	123.86	460	3	77.17	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	927	12	62.00	956	12	63.93	953	9	79.67	289	10	22.46	123	11	9.00	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	1,817	9	151.67	2,139	8	194.73	840	6	93.67	488	6	54.56	196	7	19.90	RPA4607	putative homocitrate synthase
orthoMCL3176	304	3	51.17	587	6	65.56	260	4	37.57	234	7	23.70	98	3	16.83	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	205	3	34.67	526	4	75.57	517	3	86.67	281	5	35.50	175	5	22.25	RPA4609	putative nifU protein
orthoMCL3178	255	5	32.25	477	4	68.57	425	4	61.14	350	5	44.13	156	2	31.80	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	194	1	49.25	706	9	59.08	686	9	57.42	256	11	18.50	116	7	11.90	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	470	7	47.30	645	8	58.91	500	6	55.89	191	9	16.17	98	7	10.10	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1,160	10	89.46	1,478	13	92.56	961	12	64.27	524	11	37.64	197	11	14.29	RPA4613	DUF683
orthoMCL1314	1,357	28	43.87	1,527	73	20.13	655	13	41.13	587	30	17.88	257	83	3.02	RPA4614	DUF269
orthoMCL1313	1,555	20	67.74	3,235	61	50.59	2,472	99	24.26	796	22	31.96	301	55	5.24	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	484	5	60.88	706	6	78.78	605	6	67.56	472	6	52.78	79	6	9.11	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	1,027	13	64.38	1,093	13	68.50	1,026	11	73.50	505	10	39.08	179	10	14.00	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	2,288	5	286.38	3,586	10	276.08	4,187	8	380.91	1,665	10	128.31	751	17	37.70	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	1,683	36	43.23	3,825	26	132.00	3,554	21	148.21	1,762	13	110.31	687	17	34.50	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	2,208	13	138.19	4,306	11	307.79	4,036	10	310.69	1,826	8	166.27	873	10	67.38	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	606	9	50.75	726	20	31.70	408	10	31.62	414	13	26.06	107	21	4.58	RPA4621	conserved hypothetical protein
orthoMCL1306	480	11	34.50	1,156	20	50.39	421	10	32.62	637	20	27.83	199	34	5.46	RPA4622	hypothetical protein
orthoMCL1305	960	5	120.38	2,113	19	96.18	1,072	8	97.73	728	28	23.58	355	62	5.51	RPA4623	conserved hypothetical protein
orthoMCL1304	750	4	107.57	1,041	11	74.57	952	10	73.46	373	10	28.92	146	10	11.46	RPA4624	hypothetical protein
orthoMCL1303	567	5	71.25	896	8	81.73	565	6	63.11	399	7	40.20	90	8	8.45	RPA4625	NifZ domain
orthoMCL1302	786	4	112.71	923	8	84.18	706	6	78.78	268	16	14.26	178	11	12.93	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	295	2	59.60	618	3	103.50	396	3	66.50	232	3	39.17	96	4	14.14	RPA4627	conserved hypothetical protein
orthoMCL1300	308	2	62.20	595	4	85.43	318	2	64.20	243	2	49.20	92	3	15.83	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	489	4	70.29	794	4	113.86	342	5	43.13	376	4	54.14	101	4	14.86	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	412	7	41.50	810	7	81.30	425	5	53.50	678	8	61.91	234	6	26.33	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	40,476	83	470.69	45,204	119	370.55	15,412	26	531.55	27,248	33	756.97	10,417	101	100.19	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	1,144	214	5.29	935	166	5.55	927	88	10.22	850	246	3.43	635	114	5.45	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	155	5	19.75	379	13	23.88	199	6	22.44	320	7	32.30	294	12	19.80	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.11. Nitrogenase gene expression in strains RSP24 and TIE-1 under the H_2-producing, high light condition (NF-high) condition and the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	RSP24[a]	RSP24[b]	RSP24[c]	TIE-1[a]	TIE-1[b]	TIE-1[c]	RPA Number[‡]	Product Description
orthoMCL3400	3,798	63	57.59	2,277	82	26.82	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	710	13	44.56	1,006	12	67.27	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	746	5	93.63	531	4	76.29	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	1,900	6	211.44	285	3	48.00	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	846	29	26.53	454	13	28.56	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	724	7	72.70	739	8	67.45	RPA4607	putative homocitrate synthase
orthoMCL3176	221	3	37.33	215	6	24.22	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	379	5	47.75	176	8	16.27	RPA4609	putative nifU protein
orthoMCL3178	779	6	86.89	142	5	18.13	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	213	6	24.00	347	7	35.00	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	474	10	36.69	396	6	44.33	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1,479	16	78.00	1,030	10	79.46	RPA4613	DUF683
orthoMCL1314	1,475	52	26.87	1,074	18	51.29	RPA4614	DUF269
orthoMCL1313	1,123	35	29.63	1,595	18	76.10	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	311	5	39.25	435	11	31.29	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	655	10	50.62	419	8	38.36	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	1,856	11	132.79	1,300	5	162.88	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	1,741	55	30.07	1,175	4	168.29	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	2,751	14	162.00	1,870	11	133.79	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	321	18	15.43	391	12	26.27	RPA4621	conserved hypothetical protein
orthoMCL1306	516	20	22.57	445	7	44.80	RPA4622	hypothetical protein
orthoMCL1305	394	12	26.47	1,457	18	69.52	RPA4623	conserved hypothetical protein
orthoMCL1304	226	6	25.44	552	11	39.64	RPA4624	hypothetical protein
orthoMCL1303	841	7	84.40	291	9	24.50	RPA4625	NifZ domain
orthoMCL1302	362	9	30.42	404	9	33.92	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	182	2	37.00	290	4	41.86	RPA4627	conserved hypothetical protein
orthoMCL1300	255	4	36.86	354	4	51.00	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	343	4	49.43	348	4	50.14	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	379	6	42.44	297	8	27.27	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	69,616	206	333.11	12,058	33	335.03	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	977	198	4.88	609	169	3.56	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	205	8	18.91	74	12	5.13	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.12. Nitrogenase gene expression in strains 0001L, 1a1, CGA009, CGA010 and AP1 under the H_2-producing, low light condition (NF-low) condition and H_2-producing, high light condition (NF-high) condition.

orthoMCL ID	0001L[a]	0001L[b]	0001L[c]	1a1[a]	1a1[b]	1a1[c]	CGA009[a]	CGA009[b]	CGA009[c]	CGA010[a]	CGA010[b]	CGA010[c]	AP1[a]	AP1[b]	AP1[c]	RPA Number[#]	Product Description
orthoMCL3400	2,849	3,047	0.94	1,823	3,152	0.58	4,117	7,597	0.54	2,475	2,863	0.86	4,212	6,780	0.62	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	668	666	1.00	665	1,440	0.46	776	1,658	0.47	548	776	0.71	1,027	1,506	0.68	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	548	320	1.71	360	1,251	0.29	659	1,500	0.44	378	975	0.39	730	911	0.80	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	619	358	1.72	541	1,686	0.32	806	1,608	0.50	409	1,233	0.33	559	936	0.60	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	363	504	0.72	581	1,183	0.49	632	1,195	0.53	486	657	0.74	502	896	0.56	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	701	727	0.96	742	1,679	0.44	722	1,432	0.51	649	1,029	0.63	1,085	1,807	0.60	RPA4607	putative homocitrate synthase
orthoMCL3176	153	105	1.44	109	410	0.27	118	260	0.46	107	171	0.63	161	224	0.72	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	197	161	1.22	140	713	0.20	276	897	0.31	125	467	0.27	178	227	0.79	RPA4609	putative nifU protein
orthoMCL3178	153	149	1.03	338	817	0.42	360	656	0.55	184	490	0.38	172	333	0.52	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	454	523	0.87	335	321	1.04	414	987	0.42	399	416	0.96	483	912	0.53	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	384	515	0.75	854	704	1.21	298	758	0.40	413	453	0.91	494	836	0.59	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	1,128	1,566	0.72	1,291	1,245	1.04	1,045	2,362	0.44	1,076	1,064	1.01	1,062	1,970	0.54	RPA4613	DUF683
orthoMCL1314	1,073	1,468	0.73	1,624	1,570	1.03	688	1,786	0.39	923	1,074	0.86	649	1,208	0.54	RPA4614	DUF269
orthoMCL1313	1,415	1,810	0.78	1,491	1,649	0.90	1,428	3,476	0.41	1,288	1,591	0.81	1,939	3,546	0.55	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	229	362	0.64	212	286	0.74	203	612	0.33	230	314	0.74	428	813	0.53	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	614	764	0.80	546	822	0.67	493	1,343	0.37	533	677	0.79	689	1,099	0.63	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	2,893	2,807	1.03	2,995	3,144	0.95	2,684	5,673	0.47	2,016	3,031	0.67	2,646	4,102	0.65	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	3,102	2,729	1.14	2,259	3,082	0.73	2,491	5,430	0.46	1,842	2,829	0.65	1,924	2,884	0.67	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	5,064	3,409	1.49	2,553	4,846	0.53	3,087	7,341	0.42	2,114	4,366	0.48	3,235	5,034	0.64	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	268	554	0.49	353	588	0.60	286	721	0.40	262	462	0.57	311	616	0.51	RPA4621	conserved hypothetical protein
orthoMCL1306	334	512	0.65	382	553	0.69	393	972	0.41	375	492	0.76	285	508	0.56	RPA4622	hypothetical protein
orthoMCL1305	375	597	0.63	539	816	0.66	329	843	0.39	387	526	0.74	701	1,167	0.60	RPA4623	conserved hypothetical protein
orthoMCL1304	238	364	0.66	491	765	0.64	292	723	0.41	257	385	0.67	558	712	0.78	RPA4624	hypothetical protein
orthoMCL1303	600	643	0.93	416	764	0.55	769	1,279	0.60	539	738	0.73	352	789	0.45	RPA4625	NifZ domain
orthoMCL1302	418	538	0.78	424	794	0.54	448	1,057	0.43	416	585	0.71	410	682	0.60	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	177	196	0.90	166	387	0.43	179	435	0.42	165	276	0.60	227	375	0.61	RPA4627	conserved hypothetical protein
orthoMCL1300	202	227	0.89	177	434	0.41	202	461	0.44	231	329	0.70	227	387	0.59	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	439	379	1.16	381	893	0.43	359	655	0.55	401	539	0.75	500	765	0.65	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	267	186	1.43	235	859	0.28	256	659	0.39	186	411	0.46	343	477	0.72	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	21,290	17,889	1.19	23,926	44,927	0.53	20,854	58,179	0.36	17,782	35,435	0.50	25,190	43,630	0.58	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	851	914	0.93	912	914	1.00	2,131	1,730	1.23	1,166	1,319	0.88	723	983	0.74	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	36	71	0.53	120	514	0.24	564	584	0.97	383	765	0.50	78	147	0.54	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.13. Nitrogenase gene expression in strains ATCC17007, BIS3, CEA001, DCP3 and DSM126 under the H_2-producing, low light condition (NF-low) condition and H_2-producing, high light condition (NF-high) condition.

orthoMCL ID	ATCC17007[a]	ATCC17007[b]	ATCC17007[c]	BIS3[a]	BIS3[b]	BIS3[c]	CEA001[a]	CEA001[b]	CEA001[c]	DCP3[a]	DCP3[b]	DCP3[c]	DSM126[a]	DSM126[b]	DSM126[c]	RPA Number[#]	Product Description
orthoMCL3400	2,087	3,034	0.69	2,268	2,191	1.04	2,236	2,824	0.79	1,693	4,271	0.40	3,348	3,446	0.97	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	581	777	0.75	580	628	0.92	716	1,205	0.60	333	1,220	0.27	1,063	1,270	0.84	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	401	596	0.67	273	275	0.99	547	1,180	0.46	282	1,252	0.23	1,032	1,564	0.66	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	276	720	0.39	516	238	2.15	290	1,381	0.21	441	853	0.52	494	1,406	0.35	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	419	667	0.63	597	729	0.82	300	848	0.36	451	1,184	0.38	729	1,013	0.72	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	666	848	0.79	990	1,071	0.92	627	1,060	0.59	341	1,303	0.26	660	947	0.70	RPA4607	putative homocitrate synthase
orthoMCL3176	124	125	0.99	142	127	1.12	126	282	0.45	81	366	0.23	187	324	0.58	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	141	304	0.47	122	92	1.32	152	568	0.27	90	367	0.25	198	687	0.29	RPA4609	putative nifU protein
orthoMCL3178	150	251	0.60	249	153	1.62	119	554	0.22	168	289	0.59	233	682	0.34	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	519	764	0.68	93	218	0.43	712	1,272	0.56	349	837	0.42	749	956	0.78	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	294	527	0.56	141	288	0.49	554	916	0.61	248	497	0.50	572	593	0.96	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	741	1,392	0.53	374	766	0.49	1,079	1,913	0.56	459	1,092	0.42	1,135	1,379	0.82	RPA4613	DUF683
orthoMCL1314	653	1,106	0.59	522	1,002	0.52	1,016	1,557	0.65	290	682	0.43	1,493	1,517	0.98	RPA4614	DUF269
orthoMCL1313	1,486	2,299	0.65	333	641	0.52	1,852	3,333	0.56	1,078	2,657	0.41	2,132	2,063	1.03	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	258	413	0.63	118	216	0.55	322	582	0.56	208	516	0.41	489	605	0.81	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	471	1,125	0.42	293	376	0.78	609	1,104	0.55	294	598	0.49	962	1,158	0.83	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	1,012	3,311	0.31	768	1,097	0.70	2,538	4,599	0.55	1,172	2,733	0.43	4,620	4,537	1.02	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	1,348	3,141	0.43	1,249	1,437	0.87	3,295	5,310	0.62	1,036	3,301	0.31	3,188	3,463	0.92	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	2,262	4,348	0.52	1,435	1,337	1.07	3,958	8,252	0.48	803	2,724	0.30	3,117	4,380	0.71	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	205	606	0.34	157	366	0.43	253	653	0.39	166	405	0.41	343	509	0.68	RPA4621	conserved hypothetical protein
orthoMCL1306	345	840	0.41	186	287	0.65	449	794	0.57	178	424	0.42	277	408	0.68	RPA4622	hypothetical protein
orthoMCL1305	259	669	0.39	618	936	0.66	334	679	0.49	398	1,003	0.40	1,285	1,596	0.81	RPA4623	conserved hypothetical protein
orthoMCL1304	308	890	0.35	351	481	0.73	390	616	0.63	372	942	0.40	723	805	0.90	RPA4624	hypothetical protein
orthoMCL1303	229	943	0.25	458	487	0.94	406	1,048	0.39	169	475	0.36	496	799	0.62	RPA4625	NifZ domain
orthoMCL1302	365	966	0.38	274	239	1.14	432	886	0.49	249	841	0.30	582	820	0.71	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	173	229	0.76	176	169	1.04	167	314	0.54	107	492	0.22	316	431	0.74	RPA4627	conserved hypothetical protein
orthoMCL1300	173	165	1.05	197	247	0.80	133	282	0.48	107	486	0.22	266	399	0.67	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	413	437	0.95	218	251	0.87	467	857	0.55	135	667	0.21	308	478	0.65	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	210	321	0.66	188	128	1.46	237	630	0.38	142	651	0.22	309	562	0.55	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	9,238	23,577	0.39	24,233	16,148	1.50	13,117	17,934	0.73	4,776	18,332	0.26	20,909	38,387	0.54	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	452	594	0.76	775	768	1.01	617	781	0.79	457	601	0.76	805	827	0.97	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	30	115	0.28	51	36	1.38	28	201	0.15	62	251	0.26	67	265	0.26	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.14. Nitrogenase gene expression in strains DSM8283, KD1, NCIB8288, RCH350 and RCH500 under the H_2-producing, low light condition (NF-low) condition and H_2-producing, high light condition (NF-high) condition.

orthoMCL ID	DSM8283[a]	DSM8283[b]	DSM8283[c]	KD1[a]	KD1[b]	KD1[c]	NCIB8288[a]	NCIB8288[b]	NCIB8288[c]	RCH350[a]	RCH350[b]	RCH350[c]	RCH500[a]	RCH500[b]	RCH500[c]	RPA Number[#]	Product Description
orthoMCL3400	1,025	2,258	0.45	2,062	4,765	0.43	2,270	2,440	0.93	3,152	2,236	1.41	2,810	709	3.95	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	368	871	0.42	600	2,202	0.27	1,118	1,198	0.93	621	587	1.06	834	220	3.75	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	386	733	0.53	161	1,600	0.10	1,079	1,238	0.87	329	664	0.50	655	326	2.00	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	263	703	0.38	149	1,158	0.13	635	1,061	0.60	296	864	0.34	452	460	0.98	RPA4605	electron transfer flavoprotein beta chain fixA
orthoMCL3174	255	927	0.28	373	956	0.39	695	953	0.73	306	289	1.06	418	123	3.34	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	600	1,817	0.33	687	2,139	0.32	633	840	0.75	413	488	0.85	569	196	2.87	RPA4607	putative homocitrate synthase
orthoMCL3176	142	304	0.47	57	587	0.10	230	260	0.89	97	234	0.42	204	98	2.05	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	46	205	0.24	51	526	0.10	236	517	0.46	50	281	0.19	118	175	0.68	RPA4609	putative nifU protein
orthoMCL3178	119	255	0.47	101	477	0.22	164	425	0.39	160	350	0.46	122	156	0.79	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	118	194	0.61	164	706	0.24	776	686	1.13	359	256	1.40	307	116	2.61	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	297	470	0.63	163	645	0.26	558	500	1.12	345	191	1.79	285	98	2.85	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	638	1,160	0.55	329	1,478	0.22	959	961	1.00	768	524	1.46	624	197	3.14	RPA4613	DUF683
orthoMCL1314	701	1,357	0.52	393	1,527	0.26	694	655	1.06	989	587	1.68	803	257	3.10	RPA4614	DUF269
orthoMCL1313	794	1,555	0.51	907	3,235	0.28	2,652	2,472	1.07	1,281	796	1.61	1,051	301	3.47	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	172	484	0.36	138	706	0.20	572	605	0.95	383	472	0.81	271	79	3.34	RPA4616	nitrogenase reductase-associated ferredoxin, nifN
orthoMCL1311	453	1,027	0.44	277	1,093	0.26	976	1,026	0.95	489	505	0.97	624	179	3.45	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE

orthoMCL1310	1,621	2,288	0.71	1,137	3,586	0.32	4,092	4,187	0.98	2,442	1,665	1.47	2,769	751	3.68	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	1,102	1,683	0.66	827	3,825	0.22	3,300	3,554	0.93	2,104	1,762	1.19	2,220	687	3.22	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	1,329	2,208	0.60	801	4,306	0.19	3,413	4,036	0.85	1,544	1,826	0.85	1,796	873	2.05	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	171	606	0.29	167	726	0.23	313	408	0.77	256	414	0.62	259	107	2.38	RPA4621	conserved hypothetical protein
orthoMCL1306	171	480	0.36	315	1,156	0.27	324	421	0.77	466	637	0.73	460	199	2.29	RPA4622	hypothetical protein
orthoMCL1305	360	960	0.38	565	2,113	0.27	982	1,072	0.92	679	728	0.93	863	355	2.42	RPA4623	conserved hypothetical protein
orthoMCL1304	219	750	0.29	363	1,041	0.35	963	952	1.01	310	373	0.83	456	146	3.08	RPA4624	hypothetical protein
orthoMCL1303	256	567	0.45	402	896	0.45	456	565	0.81	346	399	0.87	241	90	2.62	RPA4625	NifZ domain
orthoMCL1302	298	786	0.38	444	923	0.48	558	706	0.79	296	268	1.10	437	178	2.43	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	100	295	0.35	172	618	0.28	343	396	0.87	174	232	0.75	241	96	2.46	RPA4627	conserved hypothetical protein
orthoMCL1300	128	308	0.42	138	595	0.24	275	318	0.87	130	243	0.54	221	92	2.36	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	380	489	0.78	96	794	0.12	284	342	0.83	166	376	0.45	197	101	1.92	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	185	412	0.45	50	810	0.07	332	425	0.78	177	678	0.26	219	234	0.94	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	23,098	40,476	0.57	12,484	45,204	0.28	10,078	15,412	0.65	14,985	27,248	0.55	15,901	10,417	1.53	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	980	1,144	0.86	677	935	0.72	634	927	0.68	825	850	0.97	838	635	1.32	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	26	155	0.18	35	379	0.10	77	199	0.40	62	320	0.20	113	294	0.39	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.15. Nitrogenase gene expression in strains RSP24 and TIE-1 under the H_2-producing, low light condition (NF-low) condition and H_2-producing, high light condition (NF-high) condition.

orthoMCL ID	RSP24[a]	RSP24[b]	RSP24[c]	TIE-1[a]	TIE-1[b]	TIE-1[c]	RPA Number[#]	Product Description
orthoMCL3400	696	3,798	0.18	2,441	2,277	1.07	RPA4602	ferredoxin like protein, fixX
orthoMCL3399	176	710	0.25	871	1,006	0.87	RPA4603	nitrogen fixation protein,fixC
orthoMCL3172	131	746	0.18	341	531	0.64	RPA4604	electron transfer flavoprotein alpha chain protein fixB
orthoMCL3173	148	1,900	0.08	312	285	1.09	RPA4605	electron ransfer flavoprotein beta chain fixA
orthoMCL3174	147	846	0.18	421	454	0.93	RPA4606	nitrogenase stabilizer NifW
orthoMCL3175	158	724	0.22	590	739	0.80	RPA4607	putative homocitrate synthase
orthoMCL3176	45	221	0.21	137	215	0.64	RPA4608	nitrogenase cofactor synthesis protein nifS
orthoMCL3177	53	379	0.15	90	176	0.52	RPA4609	putative nifU protein
orthoMCL3178	57	779	0.08	220	142	1.54	RPA4610	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1317	65	213	0.31	428	347	1.23	RPA4611	putative nitrogen fixation protein nifQ
orthoMCL1316	132	474	0.28	486	396	1.23	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB
orthoMCL1315	331	1,479	0.23	1,157	1,030	1.12	RPA4613	DUF683
orthoMCL1314	388	1,475	0.26	966	1,074	0.90	RPA4614	DUF269
orthoMCL1313	264	1,123	0.24	1,413	1,595	0.89	RPA4615	nitrogenase molybdenum-iron protein nifX
orthoMCL1312	70	311	0.23	234	435	0.54	RPA4616	nitrogenase reductase-associated ferredoxin, nifN

orthoMCL1311	157	655	0.24	330	419	0.79	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE
orthoMCL1310	562	1,856	0.30	1,400	1,300	1.08	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK
orthoMCL1309	587	1,741	0.34	1,167	1,175	0.99	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD
orthoMCL1308	544	2,751	0.20	2,057	1,870	1.10	RPA4620	nitrogenase iron protein, nifH
orthoMCL1307	83	321	0.27	196	391	0.51	RPA4621	conserved hypothetical protein
orthoMCL1306	102	516	0.20	238	445	0.54	RPA4622	hypothetical protein
orthoMCL1305	92	394	0.24	781	1,457	0.54	RPA4623	conserved hypothetical protein
orthoMCL1304	89	226	0.40	342	552	0.62	RPA4624	hypothetical protein
orthoMCL1303	127	841	0.15	208	291	0.72	RPA4625	NifZ domain
orthoMCL1302	75	362	0.21	328	404	0.81	RPA4626	Protein of unknown function from Deinococcus and Synechococcus
orthoMCL1301	54	182	0.31	195	290	0.68	RPA4627	conserved hypothetical protein
orthoMCL1300	77	255	0.31	192	354	0.55	RPA4628	Protein of unknown function, HesB/YadR/YfhF
orthoMCL1299	91	343	0.27	167	348	0.48	RPA4629	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1298	68	379	0.19	170	297	0.58	RPA4630	nitrogen fixation protein nifB
orthoMCL1297	9,259	69,616	0.13	13,472	12,058	1.12	RPA4631	ferredoxin 2[4Fe-4S], fdxN
orthoMCL1296	721	977	0.74	628	609	1.03	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA
orthoMCL1295	24	205	0.13	38	74	0.53	RPA4633	short-chain dehydrogenase

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.16. Expression levels in strains 0001L, 1a1, CGA009, CGA010 and AP1 of genes outside the nitrogenase gene cluster that were up-regulated in all *Rhodopseudomonas* strains under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	0001L[a]	0001L[b]	0001L[c]	1a1[a]	1a1[b]	1a1[c]	CGA009[a]	CGA009[b]	CGA009[c]	CGA010[a]	CGA010[b]	CGA010[c]	AP1[a]	AP1[b]	AP1[c]	RPA Number[‡]	Product Description
orthoMCL2061	957	26	33.10	715	54	12.60	1,796	30	54.52	1,187	30	36.06	1,276	20	55.61	RPA0274	GlnK, nitrogen regulatory protein P-II
orthoMCL2062	294	16	15.63	255	32	7.37	401	14	23.76	391	15	21.89	326	8	29.91	RPA0275	putative ammonium transporter AmtB
orthoMCL2521	43	8	4.18	51	7	5.40	44	4	6.71	60	10	4.85	53	10	4.31	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component
orthoMCL0453	2,331	782	2.97	2,528	1,011	2.50	2,656	930	2.85	2,589	746	3.46	3,831	723	5.28	RPA1134	conserved hypothetical protein
orthoMCL2794	61	4	9.14	109	8	10.18	106	4	15.57	50	6	5.89	172	10	13.46	RPA1774	OmpA/MotB domain, possible porin
orthoMCL0250	1,792	19	81.59	1,197	29	37.50	1,857	8	169.09	1,392	15	77.50	1,056	12	70.60	RPA1927	hypothetical protein
orthoMCL3203	441	12	29.60	601	20	26.26	815	11	58.43	610	12	40.87	822	11	58.93	RPA1928	ferredoxin-like protein [2Fe-2S]
orthoMCL2430	121	15	6.89	60	8	5.73	231	13	14.63	176	11	12.79	85	4	12.57	RPA2112	putative nitrate transporter component, nrtA
orthoMCL2432	64	6	7.44	19	4	3.14	85	8	8.00	69	8	6.55	115	44	2.51	RPA2114	putative nitrate transport system ATP-binding protein
orthoMCL2433	348	39	8.36	95	16	5.16	367	49	7.12	368	60	5.89	165	27	5.60	RPA2115	putative cyanate lyase
orthoMCL2435	4,881	419	11.57	1,329	32	38.06	3,499	737	4.73	6,298	708	8.86	2,471	85	28.11	RPA2117	putative flavodoxin
orthoMCL2441	222	46	4.59	30	13	2.06	289	61	4.56	470	98	4.68	241	30	7.39	RPA2123	conserved unknown protein
orthoMCL2442	875	99	8.61	178	8	16.45	1,645	319	5.12	2,237	262	8.45	715	13	44.88	RPA2124	tonB dependent iron siderophore receptor
orthoMCL2366	593	26	20.55	672	38	16.46	1,757	29	55.00	1,414	29	44.28	698	13	43.81	RPA2156	hypothetical protein
orthoMCL1580	27	5	3.75	37	8	3.64	74	6	8.56	44	7	4.70	51	8	4.91	RPA2409	possible AmiR antitermination protein

orthoMCL0492	358	20	15.70	1,101	189	5.75	1,024	95	10.48	666	113	5.77	1,263	130	9.52	RPA2463	putative cysteine desulfurase, nifS homolog
orthoMCL0794	52	7	5.50	25	7	2.80	53	8	5.09	62	9	5.42	90	9	7.75	RPA2677	putative substrate-binding protein, subunit of ABC transporter
orthoMCL3684	17	4	2.86	46	18	2.33	19	3	3.67	28	5	3.88	35	4	5.43	RPA3666	possible ATP-binding component of ABC transporter
orthoMCL3687	200	18	9.67	153	21	6.50	161	11	11.71	256	12	17.27	232	9	19.58	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor
orthoMCL0975	433	32	12.46	470	86	5.31	739	43	16.13	831	29	26.06	647	30	19.70	RPA4209	glutamine synthetase II
orthoMCL1294	377	11	27.14	481	16	25.47	637	10	49.23	604	15	33.72	344	5	43.38	RPA4634	hypothetical protein
orthoMCL0247	51	4	7.71	55	6	6.44	162	4	23.57	52	5	6.88	115	3	19.67	RPA4714	hypothetical protein
orthoMCL1477	247	19	11.36	1,044	40	24.35	773	10	59.69	639	25	22.93	709	9	59.33	RPA4827	conserved hypothetical protein

* Expression levels of genes encoding for nitrogenase synthesis and regulation are not shown.
[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
‡ *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.17. Expression levels in strains ATCC17007, BIS3, CEA001, DCP3 and DSM126 of genes outside the nitrogenase gene cluster that were up-regulated in all *Rhodopseudomonas* strains under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	ATCC17007[a]	ATCC17007[b]	ATCC17007[c]	BIS3[a]	BIS3[b]	BIS3[c]	CEA001[a]	CEA001[b]	CEA001[c]	DCP3[a]	DCP3[b]	DCP3[c]	DSM126[a]	DSM126[b]	DSM126[c]	RPA Number[#]	Product Description
orthoMCL2061	873	19	39.82	654	37	16.43	1,377	26	47.59	791	26	27.38	1,074	23	41.42	RPA0274	GlnK, nitrogen regulatory protein P-II
orthoMCL2062	297	15	16.67	241	24	9.04	362	17	18.25	196	10	15.31	240	12	16.20	RPA0275	putative ammonium transporter AmtB
orthoMCL2521	31	6	3.78	35	8	3.45	56	6	6.56	41	8	4.00	35	7	3.80	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component
orthoMCL0453	3,748	923	4.05	5,076	1,189	4.26	3,484	822	4.23	2,298	569	4.02	4,245	695	6.09	RPA1134	conserved hypothetical protein
orthoMCL2794	76	6	8.78	119	15	6.78	72	2	15.00	278	18	13.38	133	9	11.33	RPA1774	OmpA/MotB domain, possible porin
orthoMCL0250	2,659	20	115.74	774	10	59.77	3,234	17	161.85	2,171	17	108.70	1,001	9	83.67	RPA1927	hypothetical protein
orthoMCL3203	634	16	33.53	354	19	16.23	1,299	11	93.00	683	24	25.41	1,745	12	116.53	RPA1928	ferredoxin-like protein [2Fe-2S]
orthoMCL2430	119	15	6.78	49	14	3.06	250	13	15.81	72	12	5.00	111	17	5.70	RPA2112	putative nitrate transporter component, nrtA
orthoMCL2432	45	6	5.33	69	9	6.00	71	6	8.22	30	3	5.50	35	5	4.75	RPA2114	putative nitrate transport system ATP-binding protein
orthoMCL2433	295	24	11.04	181	22	7.36	258	26	9.00	73	17	3.80	194	25	7.04	RPA2115	putative cyanate lyase
orthoMCL2435	2,378	41	54.11	2,147	175	12.08	2,547	449	5.64	449	33	12.56	2,203	86	24.79	RPA2117	putative flavodoxin
orthoMCL2441	146	15	8.28	64	29	2.09	186	30	5.73	80	16	4.37	328	18	15.76	RPA2123	conserved unknown protein
orthoMCL2442	1,056	15	58.83	47	13	3.13	1,014	77	12.71	298	31	8.85	947	12	63.33	RPA2124	tonB dependent iron siderophore receptor
orthoMCL2366	904	38	22.12	184	25	6.68	1,144	28	37.00	510	14	30.18	930	10	71.77	RPA2156	hypothetical protein
orthoMCL1580	40	6	4.78	50	9	4.42	25	4	4.00	4,614	580	7.92	34	5	4.63	RPA2409	possible AmiR antitermination protein

orthoMCL0492	574	27	19.23	297	22	12.00	1,566	158	9.75	1,241	104	11.63	1,178	138	8.38	RPA2463	putative cysteine desulfurase, nifS homolog
orthoMCL0794	93	13	6.00	40	11	3.07	103	11	7.57	31	7	3.40	75	8	7.09	RPA2677	putative substrate-binding protein, subunit of ABC transporter
orthoMCL3684	25	5	3.50	73	5	9.50	20	3	3.83	34	9	3.08	22	5	3.13	RPA3666	possible ATP-binding component of ABC transporter
orthoMCL3687	259	14	15.41	431	14	25.53	220	11	15.93	67	7	7.00	153	10	12.00	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor
orthoMCL0975	369	29	11.63	376	45	7.90	609	40	14.23	324	25	11.68	666	29	20.91	RPA4209	glutamine synthetase II
orthoMCL1294	454	6	50.78	240	14	14.29	1,159	6	129.11	220	7	22.30	617	17	31.00	RPA4634	hypothetical protein
orthoMCL0247	42	4	6.43	21	4	3.43	167	7	17.00	203	28	6.65	111	2	22.80	RPA4714	hypothetical protein
orthoMCL1477	359	18	17.24	135	11	9.86	2,390	15	132.94	2,286	21	95.38	969	13	60.75	RPA4827	conserved hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.18. Expression levels in strains DSM8283, KD1, NCIB8288, RCH350 and RCH500 of genes outside the nitrogenase gene cluster that were up-regulated in all *Rhodopseudomonas* strains under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	DSM8283[a]	DSM8283[b]	DSM8283[c]	KD1[a]	KD1[b]	KD1[c]	NCIB8288[a]	NCIB8288[b]	NCIB8288[c]	RCH350[a]	RCH350[b]	RCH350[c]	RCH500[a]	RCH500[b]	RCH500[c]	RPA Number[‡]	Product Description
orthoMCL2061	676	22	27.16	1,233	38	30.15	1,710	24	63.44	774	23	29.88	985	25	35.29	RPA0274	GlnK, nitrogen regulatory protein P-II
orthoMCL2062	177	10	13.85	326	17	16.45	437	12	29.33	161	16	8.63	266	8	24.45	RPA0275	putative ammonium transporter AmtB
orthoMCL2521	51	10	4.15	64	9	5.58	42	7	4.50	28	7	3.10	38	11	2.93	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component
orthoMCL0453	3,263	1,211	2.69	5,523	1,908	2.89	3,707	863	4.28	4,294	885	4.84	2,147	574	3.73	RPA1134	conserved hypothetical protein
orthoMCL2794	176	14	10.53	275	16	14.63	193	8	17.82	367	18	17.62	132	18	6.43	RPA1774	OmpA/MotB domain, possible porin
orthoMCL0250	1,002	13	62.81	1,710	13	107.06	749	10	57.85	616	7	61.90	143	8	13.27	RPA1927	hypothetical protein
orthoMCL3203	427	19	19.55	1,417	10	109.23	1,645	13	103.00	745	12	49.87	305	11	22.00	RPA1928	ferredoxin-like protein [2Fe-2S]
orthoMCL2430	86	7	8.90	125	16	6.74	171	14	10.24	41	6	4.89	42	5	5.63	RPA2112	putative nitrate transporter component, nrtA
orthoMCL2432	72	5	9.38	57	2	12.00	60	17	3.15	16	8	1.73	22	4	3.57	RPA2114	putative nitrate transport system ATP-binding protein
orthoMCL2433	138	16	7.42	144	23	5.65	211	119	1.75	63	28	2.13	162	23	6.35	RPA2115	putative cyanate lyase
orthoMCL2435	927	37	23.25	840	72	11.24	3,557	1,598	2.22	1,817	295	6.11	3,378	99	33.15	RPA2117	putative flavodoxin
orthoMCL2441	288	11	20.79	150	18	7.29	841	473	1.77	190	35	5.08	223	21	9.42	RPA2123	conserved unknown protein
orthoMCL2442	1,071	19	48.82	672	21	28.13	2,690	1,471	1.83	1,014	126	7.88	2,102	44	44.79	RPA2124	tonB dependent iron siderophore receptor
orthoMCL2366	836	28	27.06	1,069	21	44.67	688	20	30.04	1,117	22	44.80	278	16	14.79	RPA2156	hypothetical protein
orthoMCL1580	40	5	5.38	57	15	3.33	60	5	7.88	19	8	2.00	34	8	3.36	RPA2409	possible AmiR antitermination protein

orthoMCL0492	897	19	40.91	543	18	26.00	468	77	5.89	386	63	5.89	201	83	2.37	RPA2463	putative cysteine desulfurase, nifS homolog
orthoMCL0794	46	8	4.45	112	17	5.75	104	10	8.23	40	10	3.31	158	17	8.05	RPA2677	putative substrate-binding protein, subunit of ABC transporter
orthoMCL3684	24	6	3.00	49	14	3.06	38	6	4.56	29	10	2.46	22	6	2.78	RPA3666	possible ATP-binding component of ABC transporter
orthoMCL3687	135	8	12.55	143	10	11.23	252	9	21.25	137	7	14.00	157	5	20.00	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor
orthoMCL0975	519	25	18.64	592	32	17.00	923	28	29.87	449	35	11.89	533	29	16.75	RPA4209	glutamine synthetase II
orthoMCL1294	133	7	13.60	426	14	25.24	606	12	40.60	664	15	37.06	289	8	26.55	RPA4634	hypothetical protein
orthoMCL0247	46	10	3.77	142	5	18.13	96	11	7.07	57	11	4.29	33	6	4.00	RPA4714	hypothetical protein
orthoMCL1477	715	23	27.62	2,485	35	65.47	669	17	33.60	913	14	53.88	807	28	26.13	RPA4827	conserved hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.19. Expression levels in strains RSP24 and TIE-1 of genes outside the nitrogenase gene cluster that were up-regulated in all *Rhodopseudomonas* strains under the H_2-producing, high light condition (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	RSP24[a]	RSP24[b]	RSP24[c]	TIE-1[a]	TIE-1[b]	TIE-1[c]	RPA Number[#]	Product Description
orthoMCL2061	1,020	27	34.10	564	37	14.18	RPA0274	GlnK, nitrogen regulatory protein P-II
orthoMCL2062	293	13	18.50	188	25	6.82	RPA0275	putative ammonium transporter AmtB
orthoMCL2521	31	7	3.40	27	7	3.00	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component
orthoMCL0453	3,949	1,093	3.61	4,256	1,289	3.30	RPA1134	conserved hypothetical protein
orthoMCL2794	194	8	17.91	157	26	5.52	RPA1774	OmpA/MotB domain, possible porin
orthoMCL0250	588	10	45.46	514	6	57.44	RPA1927	hypothetical protein
orthoMCL3203	579	11	41.57	432	10	33.46	RPA1928	ferredoxin-like protein [2Fe-2S]
orthoMCL2430	104	9	8.92	22	7	2.50	RPA2112	putative nitrate transporter component, nrtA
orthoMCL2432	29	3	5.33	38	6	4.56	RPA2114	putative nitrate transport system ATP-binding protein
orthoMCL2433	199	26	6.97	167	27	5.67	RPA2115	putative cyanate lyase
orthoMCL2435	5,580	151	36.25	6,194	464	13.27	RPA2117	putative flavodoxin
orthoMCL2441	33	3	6.00	860	224	3.80	RPA2123	conserved unknown protein
orthoMCL2442	241	7	24.40	1,502	196	7.56	RPA2124	tonB dependent iron siderophore receptor
orthoMCL2366	664	28	21.52	547	28	17.74	RPA2156	hypothetical protein
orthoMCL1580	26	8	2.64	33	8	3.27	RPA2409	possible AmiR antitermination protein

orthoMCL0492	1,101	148	7.31	547	89	5.98	RPA2463	putative cysteine desulfurase, nifS homolog
orthoMCL0794	61	15	3.56	30	4	4.71	RPA2677	putative substrate-binding protein, subunit of ABC transporter
orthoMCL3684	52	6	6.11	21	7	2.40	RPA3666	possible ATP-binding component of ABC transporter
orthoMCL3687	281	11	20.29	68	10	5.46	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor
orthoMCL0975	584	33	16.31	309	37	7.80	RPA4209	glutamine synthetase II
orthoMCL1294	855	13	53.63	225	25	8.14	RPA4634	hypothetical protein
orthoMCL0247	53	4	8.00	72	8	6.82	RPA4714	hypothetical protein
orthoMCL1477	1,788	17	89.55	319	30	9.76	RPA4827	conserved hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the non-H_2-producing, high light condition (PM-high) condition.
[c] Ratio of the NF-high and PM-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.20. Expression levels in strains 0001L, 1a1, CGA009, CGA010 and AP1 of genes that were down-regulated in all *Rhodopseudomonas* strains under the H_2-producing, low light (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

orthoMCL ID	0001L[a]	0001L[b]	0001L[c]	1a1[a]	1a1[b]	1a1[c]	CGA009[a]	CGA009[b]	CGA009[c]	CGA010[a]	CGA010[b]	CGA010[c]	AP1[a]	AP1[b]	AP1[c]	RPA Number[‡]	Product Description
orthoMCL1728	5	114	0.07	1	176	0.02	3	62	0.09	9	162	0.07	27	100	0.29	RPA1874	hypothetical protein
orthoMCL2434	90	1,510	0.06	10	348	0.04	72	2,422	0.03	119	2,470	0.05	129	1,284	0.10	RPA2116	hypothetical protein
orthoMCL2435	387	4,881	0.08	14	1,329	0.01	109	3,499	0.03	363	6,298	0.06	344	2,471	0.14	RPA2117	putative flavodoxin
orthoMCL2436	32	434	0.08	2	151	0.03	29	562	0.06	29	581	0.05	56	316	0.18	RPA2118	putative ATP-binding protein of ABC transporter
orthoMCL2437	10	178	0.07	2	64	0.07	6	179	0.05	9	253	0.05	14	97	0.17	RPA2119	putative permease protein of ABC transporter
orthoMCL2438	19	171	0.13	4	69	0.10	8	214	0.05	17	275	0.07	17	109	0.18	RPA2120	putative hemin binding protein
orthoMCL2439	90	827	0.11	6	155	0.06	39	1,034	0.04	107	1,341	0.08	84	446	0.19	RPA2121	conserved unknown protein
orthoMCL2442	21	875	0.03	7	178	0.06	13	1,645	0.01	30	2,237	0.01	59	715	0.09	RPA2124	tonB dependent iron siderophore receptor
orthoMCL0061	42	306	0.15	9	62	0.18	26	147	0.19	27	147	0.20	22	79	0.30	RPA2130	DUF81
orthoMCL1552	5	78	0.10	5	73	0.11	8	70	0.15	10	111	0.11	21	65	0.35	RPA2311	hypothetical protein
orthoMCL1554	5	69	0.11	1	73	0.05	5	58	0.13	8	123	0.09	17	59	0.32	RPA2313	unknown protein
orthoMCL3575	19	102	0.21	20	162	0.14	27	140	0.21	22	159	0.15	39	116	0.35	RPA3476	possible energy transducer TonB
orthoMCL3576	34	197	0.19	20	238	0.10	25	247	0.11	43	293	0.16	50	168	0.31	RPA3477	exbD, uptake of enterochelin
orthoMCL3579	64	396	0.17	11	264	0.05	113	753	0.15	142	969	0.15	72	201	0.37	RPA3481	hypothetical protein
orthoMCL1394	24	172	0.15	20	205	0.11	30	346	0.09	17	275	0.07	23	132	0.19	RPA4402	hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.21. Expression levels in strains ATCC17007, BIS3, CEA001, DCP3 and DSM126 of genes that were down-regulated in all *Rhodopseudomonas* strains under the H_2-producing, low light (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

orthoMCL ID	ATCC17007[a]	ATCC17007[b]	ATCC17007[c]	BIS3[a]	BIS3[b]	BIS3[c]	CEA001[a]	CEA001[b]	CEA001[c]	DCP3[a]	DCP3[b]	DCP3[c]	DSM126[a]	DSM126[b]	DSM126[c]	RPA Number[‡]	Product Description
orthoMCL1728	5	76	0.10	20	50	0.43	5	53	0.14	6	111	0.08	2	146	0.03	RPA1874	hypothetical protein
orthoMCL2434	30	820	0.04	411	4,943	0.08	52	1,000	0.05	17	231	0.09	17	731	0.03	RPA2116	hypothetical protein
orthoMCL2435	83	2,378	0.04	293	2,147	0.14	211	2,547	0.08	24	449	0.06	48	2,203	0.02	RPA2117	putative flavodoxin
orthoMCL2436	10	276	0.05	71	328	0.22	17	332	0.06	11	234	0.06	10	473	0.03	RPA2118	putative ATP-binding protein of ABC transporter
orthoMCL2437	5	158	0.05	35	207	0.18	6	110	0.08	3	62	0.09	4	199	0.03	RPA2119	putative permease protein of ABC transporter
orthoMCL2438	4	104	0.07	73	360	0.21	4	93	0.07	4	84	0.08	5	188	0.04	RPA2120	putative hemin binding protein
orthoMCL2439	22	515	0.05	220	1,237	0.18	59	868	0.07	13	235	0.07	21	327	0.07	RPA2121	conserved unknown protein
orthoMCL2442	18	1,056	0.02	16	47	0.38	16	1,014	0.02	8	298	0.04	10	947	0.01	RPA2124	tonB dependent iron siderophore receptor
orthoMCL0061	10	73	0.17	80	256	0.32	27	148	0.20	10	70	0.18	20	393	0.06	RPA2130	DUF81
orthoMCL1552	6	60	0.14	25	61	0.44	3	49	0.12	6	87	0.10	7	92	0.11	RPA2311	hypothetical protein
orthoMCL1554	6	58	0.15	22	63	0.38	6	47	0.18	6	86	0.10	3	78	0.07	RPA2313	unknown protein
orthoMCL3575	19	64	0.33	44	139	0.33	21	101	0.23	9	63	0.18	15	141	0.13	RPA3476	possible energy transducer TonB
orthoMCL3576	28	125	0.24	267	901	0.30	34	178	0.20	17	128	0.15	22	306	0.08	RPA3477	exbD, uptake of enterochelin
orthoMCL3579	76	736	0.11	59	237	0.26	64	324	0.20	9	115	0.10	25	349	0.08	RPA3481	hypothetical protein
orthoMCL1394	18	135	0.15	47	218	0.23	28	136	0.22	36	373	0.10	22	137	0.18	RPA4402	hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.22. Expression levels in strains DSM8283, KD1, NCIB8288, RCH350 and RCH500 of genes that were down-regulated in all *Rhodopseudomonas* strains under the H_2-producing, low light (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

orthoMCL ID	DSM8283[a]	DSM8283[b]	DSM8283[c]	KD1[a]	KD1[b]	KD1[c]	NCIB8288[a]	NCIB8288[b]	NCIB8288[c]	RCH350[a]	RCH350[b]	RCH350[c]	RCH500[a]	RCH500[b]	RCH500[c]	RPA Number[‡]	Product Description
orthoMCL1728	36	252	0.15	5	95	0.08	7	182	0.05	16	121	0.15	45	210	0.23	RPA1874	hypothetical protein
orthoMCL2434	22	305	0.08	13	499	0.03	59	1,320	0.05	60	546	0.11	216	1,475	0.15	RPA2116	hypothetical protein
orthoMCL2435	52	927	0.06	22	840	0.03	165	3,557	0.05	254	1,817	0.14	694	3,378	0.21	RPA2117	putative flavodoxin
orthoMCL2436	8	153	0.07	13	449	0.04	28	641	0.05	29	356	0.09	112	733	0.16	RPA2118	putative ATP-binding protein of ABC transporter
orthoMCL2437	5	92	0.08	2	81	0.06	10	247	0.05	10	107	0.12	24	215	0.12	RPA2119	putative permease protein of ABC transporter
orthoMCL2438	6	75	0.12	4	135	0.05	12	244	0.06	14	135	0.12	35	284	0.13	RPA2120	putative hemin binding protein
orthoMCL2439	19	301	0.07	11	166	0.08	30	678	0.05	21	152	0.15	91	644	0.15	RPA2121	conserved unknown protein
orthoMCL2442	15	1,071	0.02	15	672	0.03	16	2,690	0.01	20	1,014	0.02	55	2,102	0.03	RPA2124	tonB dependent iron siderophore receptor
orthoMCL0061	16	133	0.14	15	104	0.17	36	518	0.07	26	144	0.20	96	569	0.17	RPA2130	DUF81
orthoMCL1552	5	21	0.33	7	37	0.25	13	140	0.11	19	69	0.31	13	111	0.14	RPA2311	hypothetical protein
orthoMCL1554	7	22	0.40	1	24	0.15	8	119	0.09	13	66	0.23	14	91	0.18	RPA2313	unknown protein
orthoMCL3575	109	320	0.35	22	179	0.14	36	220	0.17	47	137	0.36	116	494	0.24	RPA3476	possible energy transducer TonB
orthoMCL3576	119	342	0.35	30	224	0.15	46	349	0.14	43	172	0.26	111	615	0.18	RPA3477	exbD, uptake of enterochelin
orthoMCL3579	8	75	0.14	8	214	0.05	59	444	0.14	53	266	0.21	130	370	0.36	RPA3481	hypothetical protein
orthoMCL1394	20	146	0.15	28	256	0.12	34	224	0.16	37	166	0.24	93	768	0.12	RPA4402	hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.23. Expression levels in strains RSP24 and TIE-1 of genes that were down-regulated in all *Rhodopseudomonas* strains under the H_2-producing, low light (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

orthoMCL ID	RSP24[a]	RSP24[b]	RSP24[c]	TIE-1[a]	TIE-1[b]	TIE-1[c]	RPA Number[‡]	Product Description
orthoMCL1728	3	72	0.08	26	339	0.08	RPA1874	hypothetical protein
orthoMCL2434	14	1,627	0.01	41	2,106	0.02	RPA2116	hypothetical protein
orthoMCL2435	47	5,580	0.01	115	6,194	0.02	RPA2117	putative flavodoxin
orthoMCL2436	6	233	0.04	22	643	0.04	RPA2118	putative ATP-binding protein of ABC transporter
orthoMCL2437	3	161	0.04	6	275	0.03	RPA2119	putative permease protein of ABC transporter
orthoMCL2438	3	154	0.04	8	299	0.04	RPA2120	putative hemin binding protein
orthoMCL2439	8	271	0.04	18	519	0.04	RPA2121	conserved unknown protein
orthoMCL2442	4	241	0.03	9	1,502	0.01	RPA2124	tonB dependent iron siderophore receptor
orthoMCL0061	25	129	0.21	12	221	0.07	RPA2130	DUF81
orthoMCL1552	3	33	0.17	10	98	0.13	RPA2311	hypothetical protein
orthoMCL1554	3	60	0.10	4	90	0.08	RPA2313	unknown protein
orthoMCL3575	35	196	0.19	28	170	0.18	RPA3476	possible energy transducer TonB
orthoMCL3576	25	213	0.13	38	553	0.07	RPA3477	exbD, uptake of enterochelin
orthoMCL3579	11	238	0.06	25	846	0.03	RPA3481	hypothetical protein
orthoMCL1394	19	168	0.13	23	174	0.15	RPA4402	hypothetical protein

[a] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, low light condition (NF-low) condition.
[b] Number of RNA-seq reads (RPKM) in the indicated genes under the H_2-producing, high light condition (NF-high) condition.
[c] Ratio of the NF-low and NF-high gene expression. To avoid the devided-by-zero problem, '3' is added to the expression levels of all orthologous groups before calculation.
[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.24. Hypothetical genes that were up-regulated in all *Rhodopseudomonas* strains under the H_2-producing, high light (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.

orthoMCL ID	RPA Number[‡]	Product Description
orthoMCL0453	RPA1134	conserved hypothetical protein
orthoMCL0250	RPA1927	hypothetical protein
orthoMCL2441	RPA2123	conserved unknown protein
orthoMCL2366	RPA2156	hypothetical protein
orthoMCL1294	RPA4634	hypothetical protein
orthoMCL0247	RPA4714	hypothetical protein
orthoMCL1477	RPA4827	conserved hypothetical protein

[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

Table 3.25. Hypothetical genes that were down-regulated in all *Rhodopseudomonas* strains under the H_2-producing, low light (NF-low) condition compared to the H_2-producing, high light (NF-high) condition.

orthoMCL ID	RPA Number[‡]	Product Description
orthoMCL1728	RPA1874	hypothetical protein
orthoMCL2434	RPA2116	hypothetical protein
orthoMCL2439	RPA2121	conserved unknown protein
orthoMCL1552	RPA2311	hypothetical protein
orthoMCL1554	RPA2313	unknown protein
orthoMCL3579	RPA3481	hypothetical protein
orthoMCL1394	RPA4402	hypothetical protein

[‡] *Rhodopseudomonas* strain CGA009's gene numbering.

CHAPTER 4

Construction of Co-expression Networks of Diverse Strains of *Rhodopseudomonas* to Identify Genes Associated with Hydrogen Production

INTRODUCTION

The advent of RNA-seq technology provides the opportunity to study gene functions and interactions at a systems level (Croucher and Thomson, 2010; Wang *et al.*, 2009). We can now monitor the transcription levels of each gene in a genome simultaneously, at low cost and without needing to know the identities of all the nucleotide sequences being queried ahead of time. This large amount of data enables us to create a comprehensive picture of the gene expression activity inside cells. Genes participating in a common biochemical pathway or metabolic process often exhibit similar expression patterns. Thus, in large-scale gene expression data analysis, it is common to cluster genes according to the similarity in their expression patterns (Eisen *et al.*, 1998; Wen *et al.*, 1998). Doing so can lead to the discovery of new genes that are related to a phenotype and can help with the characterization of unknown gene, by inferring their functions from their association with known genes with similar expression patterns (Mani *et al.*, 2008; Slavov and Dawson, 2009).

Co-expression network analysis is a computational method for grouping genes according to their expression similarities (Stuart *et al.*, 2003). Co-expression networks are defined in mathematical terms as undirected graphs, where nodes correspond to genes and edges represent the strength of co-expression relationships between genes (Figure 4.1). Co-expression networks can be constructed as unweighted or weighted. In unweighted networks, the strength of a co-expression relationship is binary; genes in the networks either have a relationship or do not. In weighted networks, the strength of a co-expression relationship is weighted; some genes in the networks have stronger relationships than the others. Both unweighted and weighted networks have advantages. In practice, weighted networks are generally preferable as they preserve the continuous property of gene expression data (Fuller *et al.*, 2011).

Weighted Gene Co-expression Network Analysis (WGCNA) is a relatively new method for constructing co-expression networks (Zhang and Horvath, 2005). It has been used in a number

of eukaryotic studies to study complex traits (Fuller *et al.*, 2007; Ghazalpour *et al.*, 2006; Park *et al.*, 2011; Presson *et al.*, 2008; Saris *et al.*, 2009) and identify disease-associated genes and therapeutic targets (Haas *et al.*, 2012; MacLennan *et al.*, 2009; Plaisier *et al.*, 2009). Also, it has been used to compare cross-species gene expression in mammals (Oldham *et al.*, 2006) and examine phenotypic divergence in lake whitefish (Filteau *et al.*, 2013). WGCNA is useful for finding groups of highly correlated genes (modules), and relating modules to one another as well as to phenotypes.

I was interested in finding groups of genes from *Rhodopseudomonas* that work in concert when cells are producing hydrogen gas (H_2). H_2 has received serious consideration as an alternative fuel to petroleum because it is a clean-burning fuel that can be produced biologically (Das and Veziroglu, 2001). *Rhodopseudomonas* is a H_2-producing bacterial genus that is widely distributed in natural environments (Larimer *et al.*, 2004). When grown with nitrogen gas as a sole source of nitrogen, it uses nitrogenase to convert nitrogen gas to ammonia (a form of nitrogen that is readily assimilated by cells) and produces H_2 as an obligatory by-product (Dixon and Kahn, 2004), as shown below.

$$N_2 + 8e^- + 8H^+ + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP$$

H_2 production is a complex process. Figure 4.2 illustrates the metabolic modules (photophosporylation to generate ATP from light, carbon metabolism to generate reduced electrons, and nitrogenase) required for H_2 production (McKinlay and Harwood, 2010). As shown in Table 4.1, different strains of *Rhodopseudomonas* produce significantly different amounts H_2, and we hoped to identify the basis of the phenotypic variation and gain insight into H_2 production. In the past we have used microarray analysis to identify a relatively small number of genes that are key players in terms of their quantitative contribution to H_2 (Rey *et al.*, 2007). This approach has likely missed genes in peripheral metabolic modules (e.g. oxygen stress, nitrogen acquisition, reductant supply and iron acquisition) that are nevertheless

important for H_2 production (Oda *et al.*, 2005). Therefore, I was interested in determining whether the application of WGCNA to multiple strains of this genus would increase the sensitivity of our analysis and reveal additional genes that contribute to H_2 production.

WGCNA was developed to analyze eukaryotic microarray data, and to date very few studies have applied WGCNA to bacterial RNA-seq data (Fang *et al.*, 2013; Wang *et al.*, 2013). In this chapter, I describe how I adapted WGCNA for constructing co-expression networks from bacterial RNA-seq data and used the results to infer additional genes that are important for H_2 production by *Rhodopseudomonas* when light is limited and when light is plentiful. An important aspect of this study is that, in constructing the co-expression networks, I included RNA-seq data from seven different *Rhodopseudomonas* strains. This presented an additional challenge as genes from different bacterial strains have different genomic locations. Sequences of orthologous genes, furthermore, are often not identical. Consequently, additional analyses are required to create gene-to-gene association among the strains and quantify the expression levels of these genes.

MATERIALS AND METHODS

Bacterial Strains, Growth Conditions and Phenotypes

I selected seven *Rhodopseudomonas* strains whose genomes were completely sequenced and assembled for analysis. I chose strain CGA009 (Larimer *et al.*, 2004) and TIE-1 to serve as anchor strains as they are closely related (sharing 100% 16S-rRNA sequences). I also included strain CGA010, which is strain CGA009 whose *hupV* gene was repaired so that uptake hydrogenase becomes functional (Rey *et al.*, 2006). Moreover, I added strain DX-1, which is the more distant of the three strains (sharing 99.4% of 16S-rRNA sequences) as well as strains BisB18, BisB5, and BisA53 (sharing 97.3%, 97.5%, 97.8% of 16S-rRNA sequences, respectively)

(Figure 4.3). Approximately, the percentage of shared genes between pairs of genomes varies from 67% to 77%, and the average percent amino acid identity between orthologous genes from pairs of genomes varies from 75% to 87% (Oda *et al.*, 2008). Links to the publicly available genomic data for the strains are listed in Table 4.2.

All laboratory experiments, explained briefly here, were performed by Dr. Yasuhiro Oda and Colin Lappala, a research scientist and research assistant in the Harwood laboratory. *Rhodopseudomonas* strains were grown in three different conditions, which were 1) non-H_2-producing, high light, (PM[3]-high), 2) H_2-producing (nitrogen-fixing), high light (NF-high) and 3) H_2-producing (nitrogen-fixing), low light (NF-low). The goal was to compare among strains the levels of gene expression in cells when H_2 was not produced (condition 1), when H_2 was produced while light was plentiful (condition 2) and when H_2 was produced while light was limited (condition 3). Cells were grown anaerobically in light at 30°C in a defined minimal medium (Oda *et al.*, 2005) with 20 mM acetate (as the carbon source and electron donor), 10 μl VCl_3 and Wolfe's vitamin solution (5 ml/l). Ammonium sulfate was added to the non-H_2-producing (non-nitrogen-fixing) medium, but was not added to the H_2-producing (nitrogen-fixing) medium. Cultures were illuminated with 60W incandescent lamps for high-light conditions or 15W incandescent lamps for low-light conditions. The medium was sufficiently depleted of nickel such that uptake hydrogenase, a nickel-contaning enzyme, was not expected to be present and complicate H_2 measurments (Rey *et al.*, 2006). H_2 yield (μmol H_2/mg protein) was measured by gas chromatography with a thermal conductivity detector and nitrogenase activity (nmol C_2H_4 formed/min/mg protein) was also measured by gas chromatography using the acetylene reduction assay as explained in (Oda *et al.*, 2005; Rey *et al.*, 2007).

Orthologous Gene and Gene Expression Analysis

[3] Non-H_2-producing medium is traditionally referred as photosynthetic mineral medium or PM.

Genes from *Rhodopseudomonas* strains as diverse as the seven that we consider here often have different genomic locations and the sequences of orthologous genes are often not identical. It was necessary to create gene-to-gene associations by identifying orthologous genes in these strains. This was accomplished with help from Dr. Frank Larimer, a computational scientist at Oak Ridge National Laboratory, who used OrthoMCL (Li *et al.*, 2003) to analyze the seven *Rhodopseudomonas* genomes. The process is explained briefly here. First, proteins shared among the seven strains were searched against the KEGG database using an e-value cutoff of 1e-05. Next, putative orthologous relationships were identified from reciprocal best hits and the relationship information was converted into a graph. Lastly, the Markov Cluster algorithm was applied to the graph to identify groups of orthologous genes, which will be referred as orthologous genes for the remaining of this chapter.

The next step was to determine the level of orthologous gene expression, which is defined as the sum of expression levels of all genes in an orthologous group. First, I wrote software to define the presence or absence of orthologous group in each strain. Next, I used Xpression (Phattarasukol *et al.*, 2012), please see Chapter 2), to process *Rhodopseudomonas* RNA-seq data and calculate the expression levels of individual genes. Lastly, I wrote Python code to map the expression levels of individual genes in Xpression outputs to orthologous groups and compute the levels of orthologous gene expression in each strain.

RNA-seq Data

Two sets of *Rhodopseudomonas* RNA-seq data were available for analysis: one from seven strains and another from seventeen closely related strains. Although the 17-strain RNA-seq data were preferable to the 7-strain RNA-seq data, close inspection revealed that the 17-strain RNA-seq data were much noisier than the 7-strain RNA-seq data for reasons that we do not clearly understand. The expression levels of genes that are known to be co-transcribed (e.g. the expression levels of *nifKDH* genes) were noticeably more different from each other in the 17-

strain data than in the 7-strain data. This aberration would be exacerbated if the data were used for calculating gene expression ratios when constructing co-expression networks. I tried a number of different techniques to alleviate the problems. For example, I set the basal level of gene expression to a fixed number so that technical noise in the RNA-seq data would be dampened, I classified the strains according to the expression level of *fixABC* genes to detect and remove outlier strains, I remapped the 17-strain RNA-seq data to the complete genome of *R. palustris* strain CGA009 so that all RNA-seq reads were aligned to the same reference, and I log-transformed the level of gene expression before constructing co-expression networks. None of these approaches, however, helped fix or alleviate the problems. Therefore, I focused on constructing co-expression networks from the 7-strain RNA-seq data.

Adapting WGCNA for Constructing Networks from Bacterial RNA-seq Data

In initial work, for reasons that are unclear, I was able to obtain more biologically relevant results if I constructed co-expression networks using gene expression ratios, rather than numbers of reads per gene. The ratios were obtained from the number of RNA-seq reads in each gene expressed under two different environmental conditions. For example, the numbers of RNA-seq reads in the H_2-producing, high-light (NF-high) condition divided by those in the non-H_2-producing, high-light (PM-high) condition. There are two inherent problems however, when working with expression ratios. The first problem is how to deal with high estimated error in expression ratios, which happens when a gene has a low level of expression in one growth condition and a very low expression level in another growth condition, or vice versa. For example, an expression ratio of 10 would have high estimated error if it were calculated from an absolute expression level of 10 in one growth condition and 1 in another growth condition. The second problem is how to avoid the unintended grouping of up-regulated and down-regulated genes during analysis, since expression ratios of up-regulated and down-regulated genes can be mathematically close. For example, a 2-fold up-regulated gene (its expression ratio is 2) might

be grouped with a 2-fold down-regulated gene (its expression ratio of 0.5) into the same module. We found that genes that were not biologically related were sometimes grouped together due to that reason. To solve all these problems, I adapted the original WGCNA method and constructed co-expression networks from bacterial RNA-seq data as follows. Note that only key steps are explained and please see (Fuller *et al.*, 2011; Langfelder and Horvath, 2008; Zhang and Horvath, 2005) for comprehensive details of WGCNA.

Step 1: Load orthologous expression data from files.

```
NF_Expr = read.delim("7_strain_ortholog_expression_nf.tab");
PM_Expr = read.delim("7_strain_ortholog_expression_pm.tab");
```

Step 2: Discard orthologous genes whose expression levels are less than 10.

```
f_less_than_10 = function(x) {all(x<10)};

NF_Expr = NF_Expr[-which(sapply(NF_Expr, f_less_than_10))];
PM_Expr = PM_Expr[-which(sapply(PM_Expr, f_less_than_10))];
```

Step 3: Discard orthologous genes that are not present in both data frames, table-like data structures in R programming language.

```
f_index_of_NF_genes = function(s) {which(names(NF_Expr)==s)};
f_index_of_PM_genes = function(s) {which(names(PM_Expr)==s)};

NF_Expr = NF_Expr[-sapply(setdiff(names(NF_Expr), names(PM_Expr)),
f_index_of_NF_ortho)];
PM_Expr = PM_Expr[-sapply(setdiff(names(PM_Expr), names(NF_Expr)),
f_index_of_PM_ortho)];
```

Step 4: Use DESeq (Anders and Huber, 2010) to estimate errors and identify ortholog whose expression levels were significantly changed.

```
DESeq_res = nbinomTest(cds, 'PM', 'NF')
```

Step 5: Select orthologous genes that meet the following criteria: 1) having expression levels of 10 or more, and 2) having expression ratios of 2.0 fold or more (*i.e.,* up-regulated), or 0.5 fold or less (*i.e.,* down-regulated) under two conditions.

```
upreg_ortho[[1]] = subset(DESeq_res, (foldChange>=2 & baseMean>=10),
select='id');
downreg_ortho[[1]] = subset(DESeq_res, (foldChange<=0.5 &
baseMean>=10), select='id');
```

Step 6: Select orthologous genes that have smaller fold expression changes if they meet the following criteria: 1) having expression levels of 10 or more, and 2) the expression changes were statistically significant (*i.e., p*-value is of 0.05 or less). Note that by applying the criteria in Step 5 and 6, the problem with unmeaningful expression can be alleviated.

```
upreg_ortho[[2]] = subset(DESeq_res, (foldChange>1 & padj<=0.05 &
baseMean>=10), select='id');
downreg_ortho[[2]] = subset(DESeq_res, (foldChange<1 & padj<=0.05 &
baseMean>=10), select='id');
```

Step 7: Construct co-expression networks for up-regulated and down-regulated orthologous genes separately. This prevents the unintended clustering of up-regulated and down-regulated genes. Note that for the remaining of this document only the construction of up-regulated co-expression networks is explained.

Step 8: Combine up-regulated orthologous genes from various strains.

```
for (i in upreg_ortho) {
   all_upreg_ortho = union(all_upreg_ortho, unlist(i));
}
```

Step 9: Select expression data that belong to the up-regulated orthologous genes.

```
NF_Expr = NF_Expr[sapply(all_upreg_ortho, f_index_of_NF_ortho)];
PM_Expr = PM_Expr[sapply(all_upreg_ortho, f_index_of_PM_ortho)];
```

Step 10: Create an expression-ratio data frame for network construction. Note that '1' is added to the expression data to prevent the divided-by-zero problem.

```
datExpr = (NF_Expr+1)/(PM_Expr+1);
```

Step 11: Calculate the similarity between two expression ratio profiles, which is the absolute value of the Pearson correlation coefficient of the two profiles, and store them in an adjacency matrix.

```
adjacency = adjacency(datExpr);
```

Step 12: Transform the adjacency matrix into a topological overlap matrix (TOM) (Ravasz *et al.*, 2002). Doing so helps minimize effects of noise and spurious associations (Li and Horvath, 2007; Yip and Horvath, 2007).

```
TOM = TOMsimilarity(adjacency);
```

Step 13: Compute the dissimilarity TOM.

```
dissTOM = 1-TOM;
```

Step 14: Use the dissimilarity TOM as input of average-linkage hierarchical clustering analysis. A hierarchical clustering tree (dendrogram) of genes was produced.

```
geneTree = flashClust(as.dist(dissTOM), method = "average");
```

Step 15: Cut the dendogram to identify modules, which are clusters of highly interconnected orthologous genes.

```
dynamicMods = cutreeDynamic(dendro = geneTree, distM = dissTOM,
deepSplit = 2, pamRespectsDendro = FALSE, minClusterSize = minModuleSize);
```

Step 16: Calculate a module eigengene for each module. The module eigengene is the first principal component of the standardized expression profiles of a given module.

```
MEList = moduleEigengenes(datExpr, colors = dynamicColors);
```

Step 17: Merge modules whose expression profiles are very similar.

```
merge = mergeCloseModules(datExpr, dynamicColors, cutHeight =
MEDissThres, verbose = 3);
```

Step 18: Load phenotypic data from a file.

```
phenoData = read.csv("7_strain_phenotype.csv");
```

Step 19: Calculate the ratio of phenotypic changes between two conditions.

```
phenoChanges = phenoData[,7]/phenoData[,8];
```

Step 20: Calculate Pearson correlation coefficients between the module eigengenes and the phenotypic changes to identify modules significantly associated with the phenotypic changes.

```
modulePhenoCor = cor(MEs, phenoChanges, use = "p");
```

Step 21: Calculate Student asymptotic *p*-values to determine the statistical significant of the association of the module eigengenes and the phenotypic changes.

```
modulePhenoPvalue = corPvalueStudent(modulePhenoCor, nSamples);
```

RESULTS

Co-expression Networks

Out of 5,807 orthologous groups identified in the seven *Rhodopseudomonas* strain genomes, genes from 750 orthologous groups were up-regulated and genes from 732 orthologous groups were down-regulated when comparing RNA-seq data in the H_2-producing, high light (NF-high) condition to the non-H_2-producing, high light (PM-high) condition. Moreover, genes from 794 orthologous groups were up-regulated and genes from 700 orthologous groups were down-regulated when comparing RNA-seq data in the H_2-producing, low light (NF-low) condition to the H_2-producing, high light (NF-high) condition. Consequently, four co-expression networks were created from the expression ratios of these orthologous genes: 1) up-regulated NF-high/PM-high, 2) down-regulated NF-high/PM-high, 3) up-regulated NF-Low/NF-high and 4) down-regulated NF-low/NF-high.

In the up-regulated NF-high/PM-high and down-regulated NF-high/PM-high networks, a total of 10 and 12 modules were identified, respectively. Meanwhile, in the up-regulated NF-low/NF-high and down-regulated NF-low/NF-high networks a total of 9 and 10 modules were

detected, respectively (Table 4.3 and Figure 4.4-4.7). The largest modules of the four networks contained genes from 120-180 orthologous groups while the smallest modules contained genes from 30-50 orthologous groups (Table 4.3). The networks consisted of relatively small numbers of large modules, which implies that many genes were expressed in similar fashions (approximately, in 9-12 patterns of expression ratio change) when H_2 was produced.

Modules Associated with Phenotypic Changes

The next step was to identify modules that are important to phenotypes, which was done by considering whether any of the module eigengenes (*i.e.*, weighted average of the orthologous gene expression ratios in a module) were associated with phenotypic changes. I started by considering the correlation coefficients between module eigengenes and changes in nitrogenase activity of strains grown in high light as I aimed to use this result as a positive control. Nitrogenase is comprised of a dinitrogenase catalytic component consisting of two polypeptides (*nifDK*) and a dinitrogenase reductase component (*nifH*), which transfers electrons to the site of nitrogen reduction in the dinitrogenase component. Both *NifH* and *NifDK* are metalloproteins with iron sulfur centers. In addition, the catalytic subunit includes an iron-molybdenum cofactor at its active site that is synthesized by a dedicated set of assembly proteins (Seefeldt *et al.*, 2009). In all, about 25 genes have been experimentally verified as being needed for the synthesis of active nitrogenase (Oda *et al.*, 2005); therefore, changes in nitrogenase activity were expected to be associated with modules containing up-regulated nitrogenase synthesis genes. I found that the module eigengenes of the magenta, red and brown modules in the up-regulated NF-high/PM-high network were associated to changes in nitrogenase activity (Figure 4.8). I hoped to find strong and significant associations between module eigengenes and phenotypic changes and thus set cutoffs at $r \geq 0.8$ and p-value ≤ 0.05[4]. Although none of the correlation coefficients and their corresponding p-values met my cutoffs, closer inspection revealed that the

[4] While $p < 0.05$ is widely used in the scientific literature, this cutoff may lead to false positive results.

results were biologically informative. The brown module contains almost all of the 32 genes in the *Rhodopseudomonas* nitrogenase gene cluster and the red module includes the remaining genes (Figure 4.9 and Table 4.4). The brown module also includes a gene known to be a regulator of nitrogen fixation (orthoMCL2498: GlnK2) and genes that are likely involved in iron storage (orthoMCL0794: bacterioferritin) and iron-sulfur center synthesis (*sufCDS*), two functions that likely contribute to the effective synthesis of active nitrogenase. In addition, it includes superoxide dismutase (orthoMCL1706) that could be important for protecting nitrogenase, an extremely oxygen-sensitive enzyme, from superoxide anion. The magenta module includes fewer genes that have an obvious association to nitrogenase activity, but a few stand out. These include a putative molybdate transport gene and an iron-uptake regulatory gene. Since cells grown under nitrogen-fixing conditions (NF) are starved for readily usable nitrogen, it is not surprising that the magenta, brown and red modules all include genes that encode for putative amino acid transport systems. Moreover, the red module also includes genes for light-harvesting peptides, which are known to function to increase the efficiency of photophosphorylation and thus energy generation by *Rhodopseudomonas*. This in turn likely helps nitrogenase, an energy intensive enzyme, to function effectively. All of these biologically meaningful results confirmed that my adapted WGCNA approach is capable of identifying modules that are important to phenotypes without prior biological knowledge.

I next calculated the correlation coefficients between module eigengenes in the up-regulated and down-regulated NF-high/PM-high network and changes in H_2 yields by strains grown in high light. In the up-regulated NF-high/PM-high network, I found that the module eigengenes of the magenta, yellow and red modules are positively associated while the module eigengene of the black module is negatively associated with changes in H_2 yields (Figure 4.10). Again, although none of the correlation coefficients and their corresponding *p*-values met my statistical cutoffs, closer inspection revealed some interesting results. The yellow module, which was less significantly associated with changes in nitrogenase activity than with changes in H_2 yields,

includes a group of contiguous genes annotated as amino acid transport systems that may be part of a nitrogen starvation response. Uridylate kinase and uridine monophosphate kinase genes may be involved in conversion of UMP to ATP. This purine salvage pathway may be helpful in supplying ATP needed by nitrogenase for H_2 production. Other genes may be involved in transfer of electrons from substrates that may donate electrons to nitrogenase. For example, unknown proteins (orthoMCL2768 and 2767) have the signatures of c-type cytochromes. The yellow module also includes a number of hypothetical genes. Interestingly, the nitrogenase-related brown module was not highly associated with changes in H_2 yields even though nitrogenase is the enzyme that produces H_2. One possible explanation is that nitrogenase gene expression is essential but not a limiting factor in H_2 production by *Rhodopseudomonas*. Moreover, I found that the module eigengene of the grey module in the down-regulated NF-high/PM-high network was the most associated with hydrogen production (Figure 4.11). The grey module contained only one gene (Table 4.5), which is annotated as an outer membrane receptor gene and interestingly present only in strains TIE-1 and BisB18.

Lastly, I considered the correlation coefficients between module eigengenes in the up-regulated and down-regulated NF-low/NF-high network to ratios of H_2 yields by strains grown in low light compared to high light. In the up-regulated NF-low/NF-high network, I found that the module eigengene of the brown module was significantly positively associated (r = 0.77, p-value = 0.04) while the module eigengenes of the blue module was significantly negatively associated with the ratios of H_2 yields (r = -0.76, p-value = 0.05) (Figure 4.12). The brown module includes light-harvesting genes and genes for bacteriochlorphyll and carotenoid biosynthesis (Table 4.6). It also includes a relatively large number of genes of unknown function as well as a relatively large number of genes that are present in strains other than strain CGA009. The blue module includes genes that have little obvious relationship to H_2 production. It is possible that here, as with genes in other modules, we have identified new genes associated with H_2 production. One can have most confidence in identifying genes that comprise an operon

(for example RPA3308-13) as good candidates to follow up on to look at the phenotypes of individual mutants. It is also important to note that cells grown in low light tend to have significantly lower growth rates than cells grown in high light and genes that are expressed differentially as a function of growth rate may indirectly affect H_2 yields in unpredictable ways. Moreover, I found that the module eigengenes of the pink, magenta and turquoise modules in the down-regulated NF-low/NF-high network were positively associated with the ratios of H_2 yields (Figure 4.13). These modules contains a number of candidate genes in which up-regulation in high light is associated with increased H_2 yields (Table 4.7). It should be noted that cells exposed to high light tend to grow faster, so growth rate rather than the amount of light per se could be the actual condition leading to the increase in H_2 production. The pink module includes genes such as RPA2119-2121 and RPA2123-6. Several of these genes have unknown functions, but others in this cluster hint at a system for hemin transport. The turquoise module include encodes genes for synthesis and uptake of the siderophore rhizobactin (RPA2378-2390) (Larimer *et al.*, 2004) as well as genes for cytochrome bd-quinol oxidase (RPA4794).

DISCUSSION

The data that I have presented suggest that WGCNA can be adapted to construct co-expression networks from bacterial RNA-seq data and yield biologically meaningful results. Although including data from strains as different as the seven *Rhodopseudomonas* strains presented additional challenges, the problem can be overcome by identifying orthologous genes in different strains and then using that information for determining the expression levels of genes whose genomic locations and sequences are different. I constructed co-expression networks from RNA-seq data of *Rhodopseudomonas* strains grown under two sets of two different conditions. The first set of conditions was nitrogen-fixing with high light (NF-high), a growth condition known to lead to H_2 production, and non-nitrogen fixing with high light (PM-high), a growth condition in which nitrogen fixation and therefore H_2 production does not occur.

For the second set of conditions, cells were grown under nitrogen-fixing conditions with low light (NF-low) and nitrogen-fixing conditions with high light (NF-high). Although H_2 is produced in both NF-low and NF-high growth conditions, phenotypic data showed that the amount of H_2 produced varied depending on light intensity. I used my modified WGCNA protocol to group correlated expression ratios of orthologous genes into modules. I also determined whether any of the weighted averages of the orthologous gene expression ratios of those modules were associated with changes in nitrogenase activity and changes in H_2 yields. In addition, I inspected members of associated modules for candidate genes that might play a role in H_2 production. The results of my analysis showed that modules containing nitrogenase genes are associated to nitrogenase activity. This is expected and provides a proof of concept of my method. The results also indicate that nitrogenase gene expression is essential but may not be a limiting factor for H_2 production. Moreover, light-harvesting gene expression is very important to H_2 production by *Rhodopseudomonas* in growth conditions where light is limited as well as where light is plentiful. The finding is expected as the organism harness energy from light for driving energetically unfavorable reactions, including H_2 production, to support growth. I also identified a number of new genes that have little obvious relationship but might be important to H_2 production. These genes cannot be readily identified from a simple transcriptome analysis of H_2-producing cells of a single strain or a comparative transcriptome analysis of a few strains of *Rhodopseudomonas*, which underlines the benefits of using my adapted WGCNA method to construct co-expression networks.

It should be noted that the statistical significance of my results from the association between module eigengenes and phenotypic changes did not meet traditional cutoffs at $p < 0.05$, except those from the association between module eigengenes in the NF-low/NF-high network to ratios of H_2 yields by strains grown in low light compared to high light. As a result, I cannot exclude the possibility that the module eigengenes were associated to the phenotypic changes by chance, and that genes identified as being important were not biologically related to H_2 production after

all. The problem with statistical significance stems from two sources. First, the sample size of seven strains is too small. It is known that the significance level of a correlation coefficient changes according to the size of the sample from which it is computed. Therefore, it is preferable to have a large collection of strains, ideally 100 strains or more, for analysis in order to increase the level of statistical significance (*i.e.,* to reduce to probability that the identified relationship between variables occurred purely by accident but, in fact, no such relationship exists in reality). Second, the seven *Rhodopseudomonas* strains are too diverse. One of the benefits of studying closely related strains of bacteria is to take advantage of the natural genetic variation in the bacterial population. The key, however, is to assemble a set of strains with an appropriate amount of variation. If there is too little variation among strains, their expression profiles will be too similar, meaning that important and unimportant genes will not be discernable in a co-expression network. If there is too much variation, their expression profiles will be too different, causing modules to become non-homogenous groups of genes and effectively diminishing the ability to resolve module-phenotype relationships on a fine scale. The seven *Rhodopseudomonas* strains are not closely related, having overall levels of 16S-rRNA sequence identities of 97.3%. This degree of divergence is close to a cutoff for being classified as a species and it is likely that there is too much divergence to take advantage of the natural genetic variation in the bacterial population. Lastly, it is important to note that even if changes in the expression levels of certain genes are significantly and strongly correlated to phenotypic changes, one must be aware that such correlation does not necessarily equate to causation, as it is possible for two variables to be correlated but not have one variable cause another. Recently, significant research interest has shifted to the use of Bayesian networks to study causal interaction networks of biological. Bayesian networks, in mathematical terms, are directed acyclic graphs, where nodes represent random variables (e.g. genes under study) and directed edges represent probabilistic beliefs that one node affects the behavior of another. To identify genes that are important for H_2 production, we can integrate the data that we have generated

(*i.e.*, naturally occurring genetic variation, transcriptomic variation and H_2 yields in the closely related strains of *Rhodopseudomonas*) and construct Bayesian networks using methods that have been designed specifically for this purpose (Schadt *et al.*, 2005; Zhu *et al.*, 2004; 2008). Doing so may lead to the discovery of key nodes that maximally impacts H_2 production, as has been successfully done in studies involving higher organisms.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Croucher, N.J., and Thomson, N.R. (2010). Studying bacterial transcriptomes using RNA-seq. Curr. Opin. Microbiol. *13*, 619–624.

Das, D., and Veziroglu, T.N. (2001). Hydrogen production by biological processes: a survey of literature. Int J Hydrogen Energy *26*, 13–28.

Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. *2*, 621–631.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. *95*, 14863–14868.

Fang, G., Passalacqua, K.D., Hocking, J., Llopis, P.M., Gerstein, M., Bergman, N.H., and Jacobs-Wagner, C. (2013). Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. BMC Genomics *14*, 450.

Filteau, M., Pavey, S.A., St-Cyr, J., and Bernatchez, L. (2013). Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. Mol. Biol. Evol. *30*, 1384–1396.

Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusis, A.J., and Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm. Genome *18*, 463–472.

Fuller, T., Langfelder, P., Presson, A., and Horvath, S. (2011). Review of Weighted Gene Coexpression Network Analysis. In Handbook of Statistical Bioinformatics, H.H.-S. Lu, B. Schölkopf, and H. Zhao, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 369–388.

Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., Brozell, A., Schadt, E.E., Drake, T.A., Lusis, A.J., *et al.* (2006). Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet. *2*, e130.

Haas, B.E., Horvath, S., Pietiläinen, K.H., Cantor, R.M., Nikkola, E., Weissglas-Volkov, D., Rissanen, A., Civelek, M., Cruz-Bautista, I., Riba, L., *et al.* (2012). Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. BMC Med Genomics *5*, 61.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., *et al.* (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat. Biotechnol. *22*, 55–61.

Li, A., and Horvath, S. (2007). Network neighborhood analysis with the multi-node topological overlap measure. Bioinformatics *23*, 222–231.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for

eukaryotic genomes. Genome Res. *13*, 2178–2189.

MacLennan, N.K., Dong, J., Aten, J.E., Horvath, S., Rahib, L., Ornelas, L., Dipple, K.M., and McCabe, E.R.B. (2009). Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. Mol. Genet. Metab. *98*, 203–214.

Mani, K.M., Lefebvre, C., Wang, K., Lim, W.K., Basso, K., Dalla-Favera, R., and Califano, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol. Syst. Biol. *4*, 169.

McKinlay, J.B., and Harwood, C.S. (2010). Photobiological production of hydrogen gas as a biofuel. Curr. Opin. Biotechnol. *21*, 244–251.

Oda, Y., Larimer, F.W., Chain, P.S.G., Malfatti, S., Shin, M.V., Vergez, L.M., Hauser, L., Land, M.L., Braatsch, S., Beatty, J.T., *et al.* (2008). Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. Proc. Natl. Acad. Sci. U.S.A. *105*, 18543–18548.

Oda, Y., Samanta, S.K., Rey, F.E., Wu, L., Liu, X., Yan, T., Zhou, J., and Harwood, C.S. (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. J. Bacteriol. *187*, 7784–7794.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc. Natl. Acad. Sci. U.S.A. *103*, 17973–17978.

Park, C.C., Gale, G.D., de Jong, S., Ghazalpour, A., Bennett, B.J., Farber, C.R., Langfelder, P., Lin, A., Khan, A.H., Eskin, E., *et al.* (2011). Gene networks associated with conditional fear in mice identified using a systems genetics approach. BMC Syst Biol *5*, 43.

Phattarasukol, S., Radey, M.C., Lappala, C.R., Oda, Y., Hirakawa, H., Brittnacher, M.J., and Harwood, C.S. (2012). Identification of a p-coumarate degradation regulon in *Rhodopseudomonas palustris* by Xpression, an integrated tool for prokaryotic RNA-seq data processing. Appl. Environ. Microbiol. *78*, 6812–6818.

Plaisier, C.L., Horvath, S., Huertas-Vazquez, A., Cruz-Bautista, I., Herrera, M.F., Tusie-Luna, T., Aguilar-Salinas, C., and Pajukanta, P. (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet. *5*, e1000642.

Presson, A.P., Sobel, E.M., Papp, J.C., Suarez, C.J., Whistler, T., Rajeevan, M.S., Vernon, S.D., and Horvath, S. (2008). Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Syst Biol *2*, 95.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Rey, F.E., Heiniger, E.K., and Harwood, C.S. (2007). Redirection of metabolism for biological hydrogen production. Appl. Environ. Microbiol. *73*, 1665–1671.

Rey, F.E., Oda, Y., and Harwood, C.S. (2006). Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. J. Bacteriol. *188*,

6143–6152.

Saris, C.G.J., Horvath, S., van Vught, P.W.J., van Es, M.A., Blauw, H.M., Fuller, T.F., Langfelder, P., DeYoung, J., Wokke, J.H.J., Veldink, J.H., *et al.* (2009). Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. BMC Genomics *10*, 405.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C., *et al.* (2005). An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. *37*, 710–717.

Seefeldt, L.C., Hoffman, B.M., and Dean, D.R. (2009). Mechanism of Mo-dependent nitrogenase. Annu. Rev. Biochem. *78*, 701–722.

Slavov, N., and Dawson, K.A. (2009). Correlation signature of the macroscopic states of the gene regulatory network in cancer. Proc. Natl. Acad. Sci. U.S.A. *106*, 4079–4084.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249–255.

Wang, J., Wu, G., Chen, L., and Zhang, W. (2013). Cross-species transcriptional network analysis reveals conservation and variation in response to metal stress in cyanobacteria. BMC Genomics *14*, 112.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L., and Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development. Proc. Natl. Acad. Sci. U.S.A. *95*, 334–339.

Yip, A.M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics *8*, 22.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.

Zhu, J., Lum, P.Y., Lamb, J., GuhaThakurta, D., Edwards, S.W., Thieringer, R., Berger, J.P., Wu, M.S., Thompson, J., Sachs, A.B., *et al.* (2004). An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet. Genome Res. *105*, 363–374.

Zhu, J., Zhang, B., Smith, E.N., Drees, B., Brem, R.B., Kruglyak, L., Bumgarner, R.E., and Schadt, E.E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat. Genet. *40*, 854–861.

Figure 4.1. A co-expression network is an undirected graph, where nodes correspond to genes and edges represent the strength of co-expression relationships between genes.

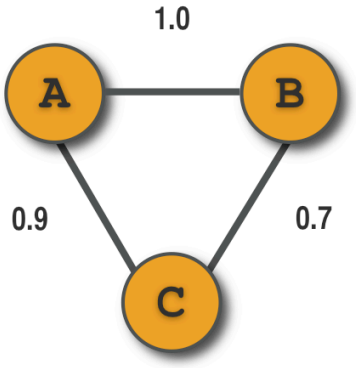
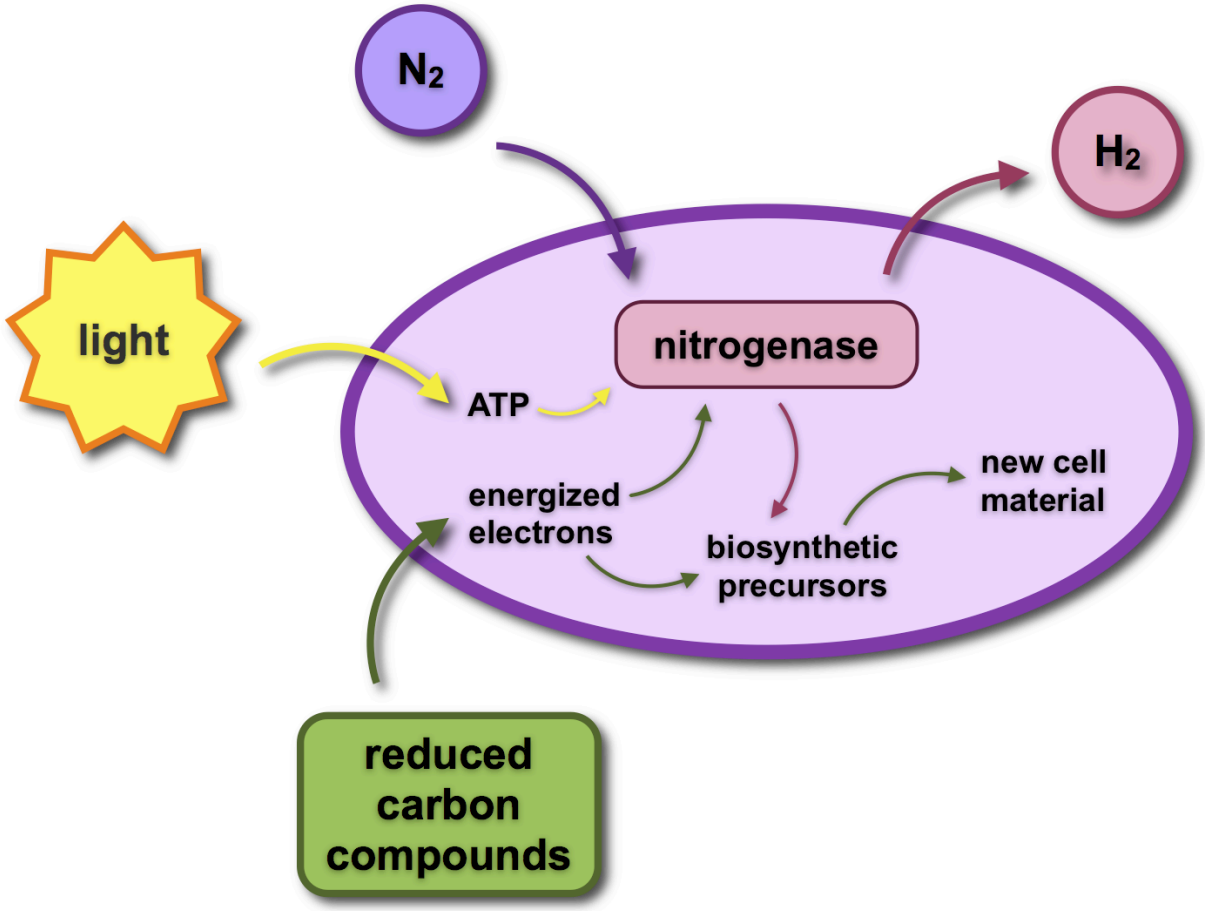
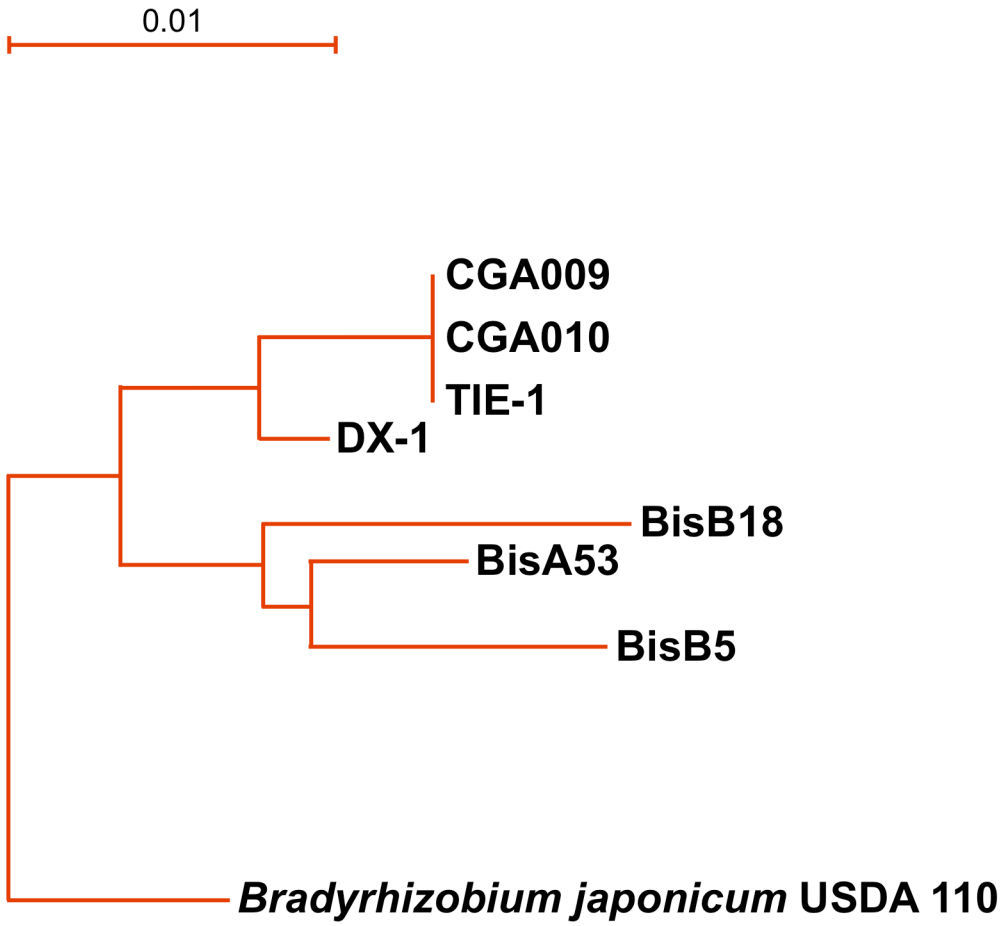


Figure 4.2. H_2 production by *Rhodopseudomonas* requires the integration of dozens of metabolic reactions.



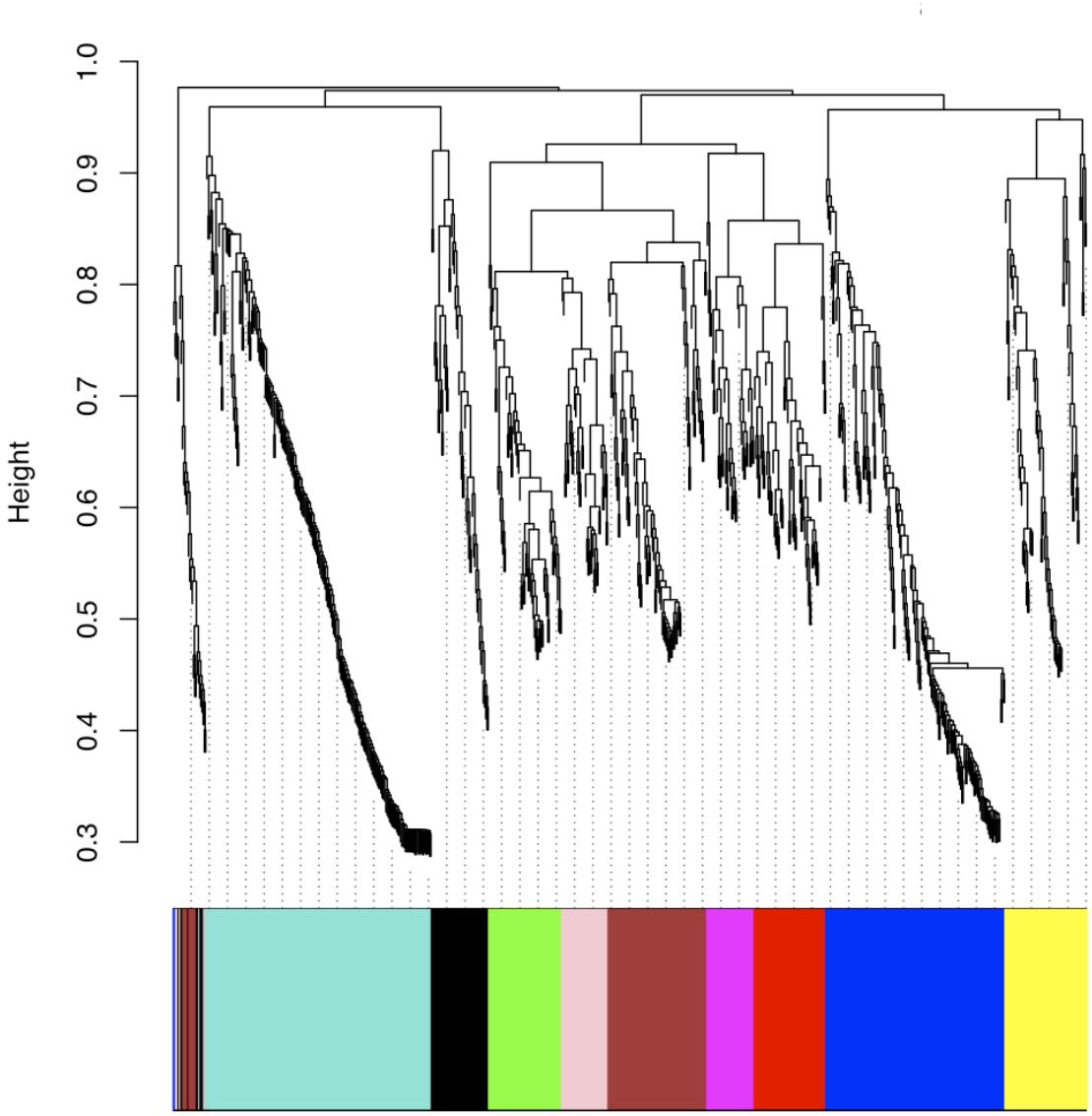
These include, for example, the metabolic modules of photophosporylation to generate ATP from light, carbon metabolism to generate reduced electrons, and nitrogen metabolism to generate nitrogenase, the enzyme that produces H_2.

Figure 4.3. Phylogenetic relationships of seven *Rhodopseudomonas* strains based on 16S-rRNA sequences.



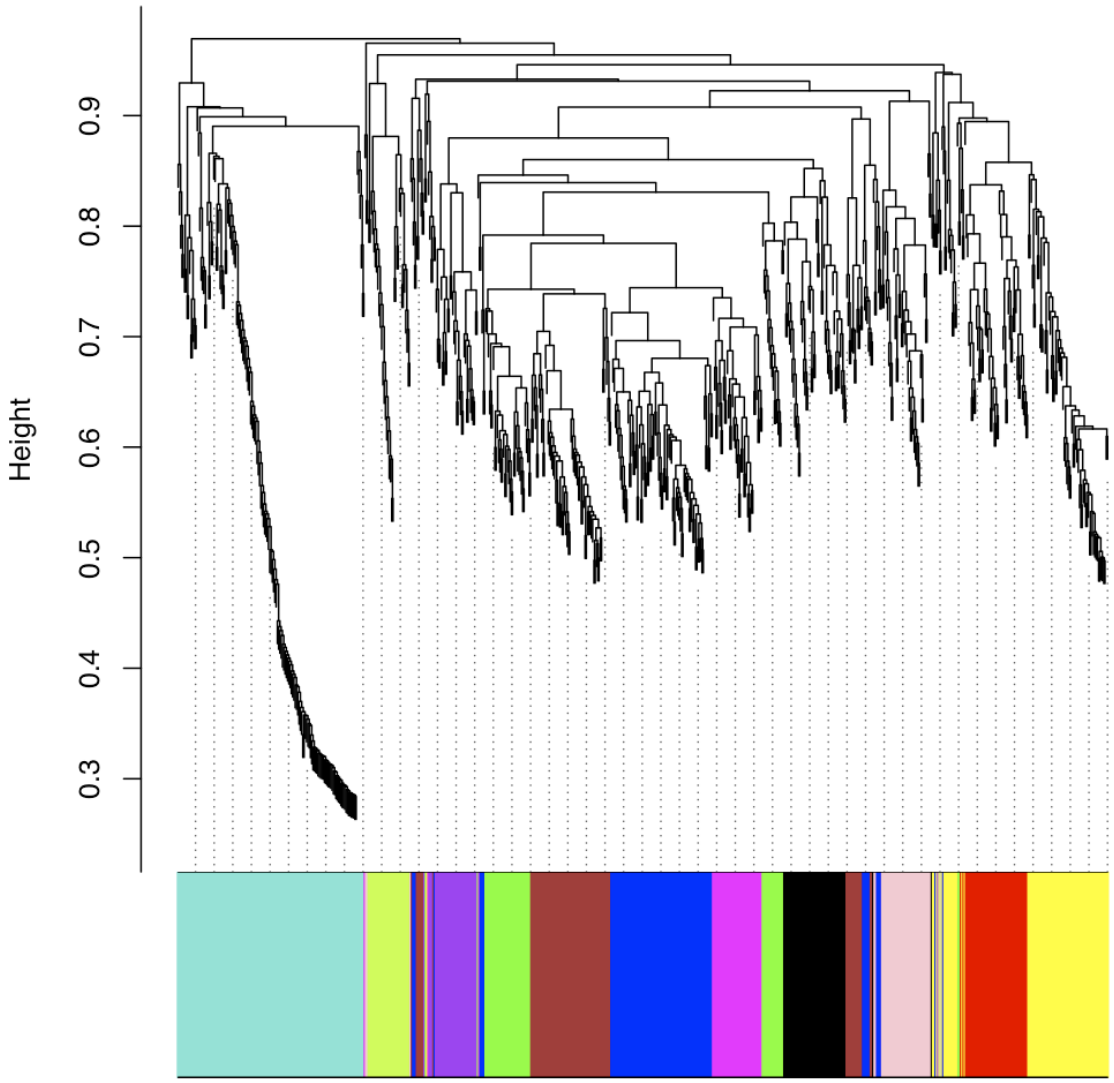
Adapted from (Oda *et al.*, 2008).

Figure 4.4. Dendrogram of 750 orthologous genes that were up-regulated in the H_2-producing, high-light (NF-high) condition compared to the non-H_2-producing, high light (PM-high) condition.



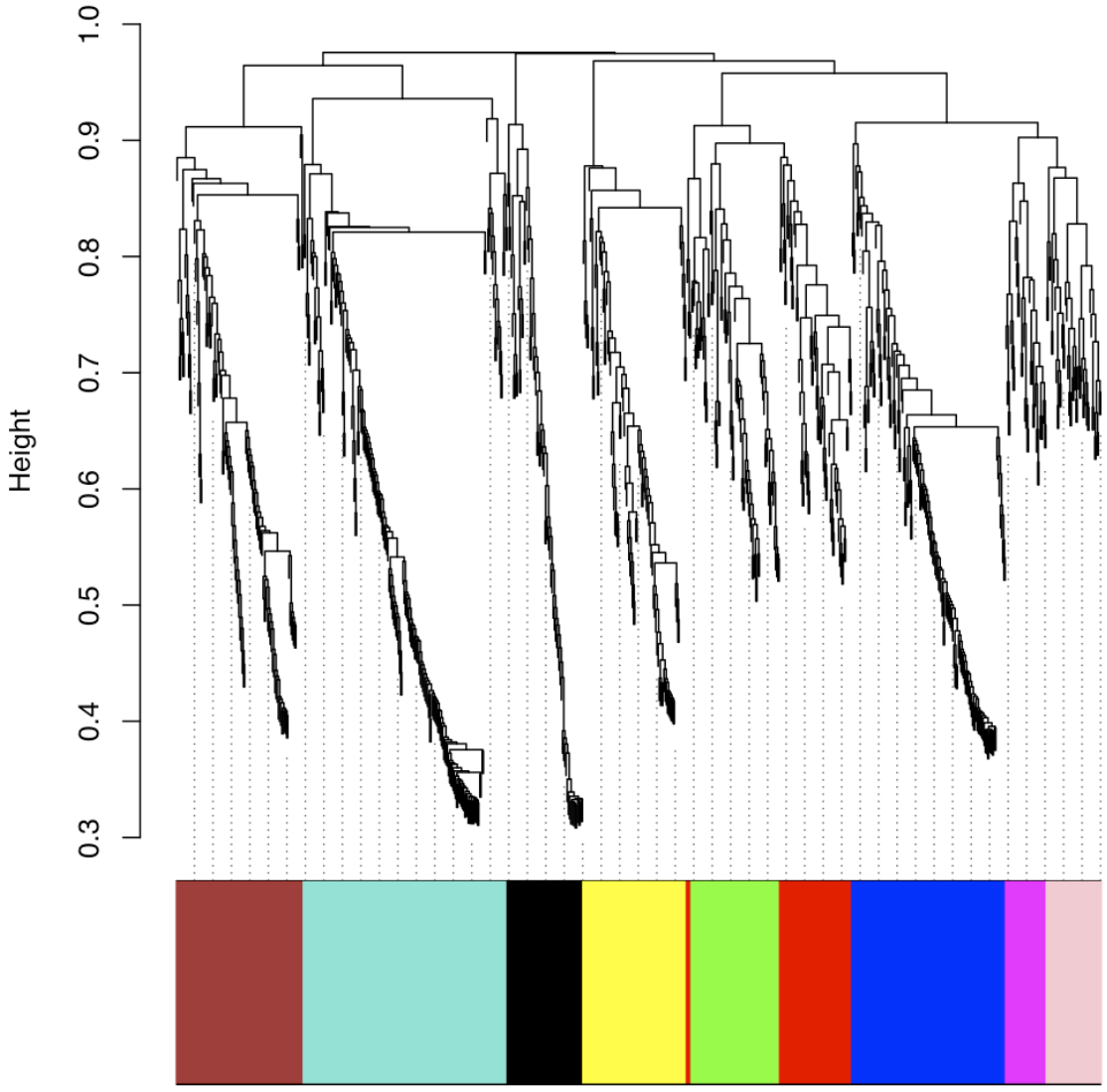
Ten modules were identified and labeled with colors as shown below the dendrogram.

Figure 4.5. Dendrogram of 732 orthologous genes that were down-regulated in the H_2-producing, high-light (NF-high) condition compared to the non-H_2-producing (PM-high) condition.



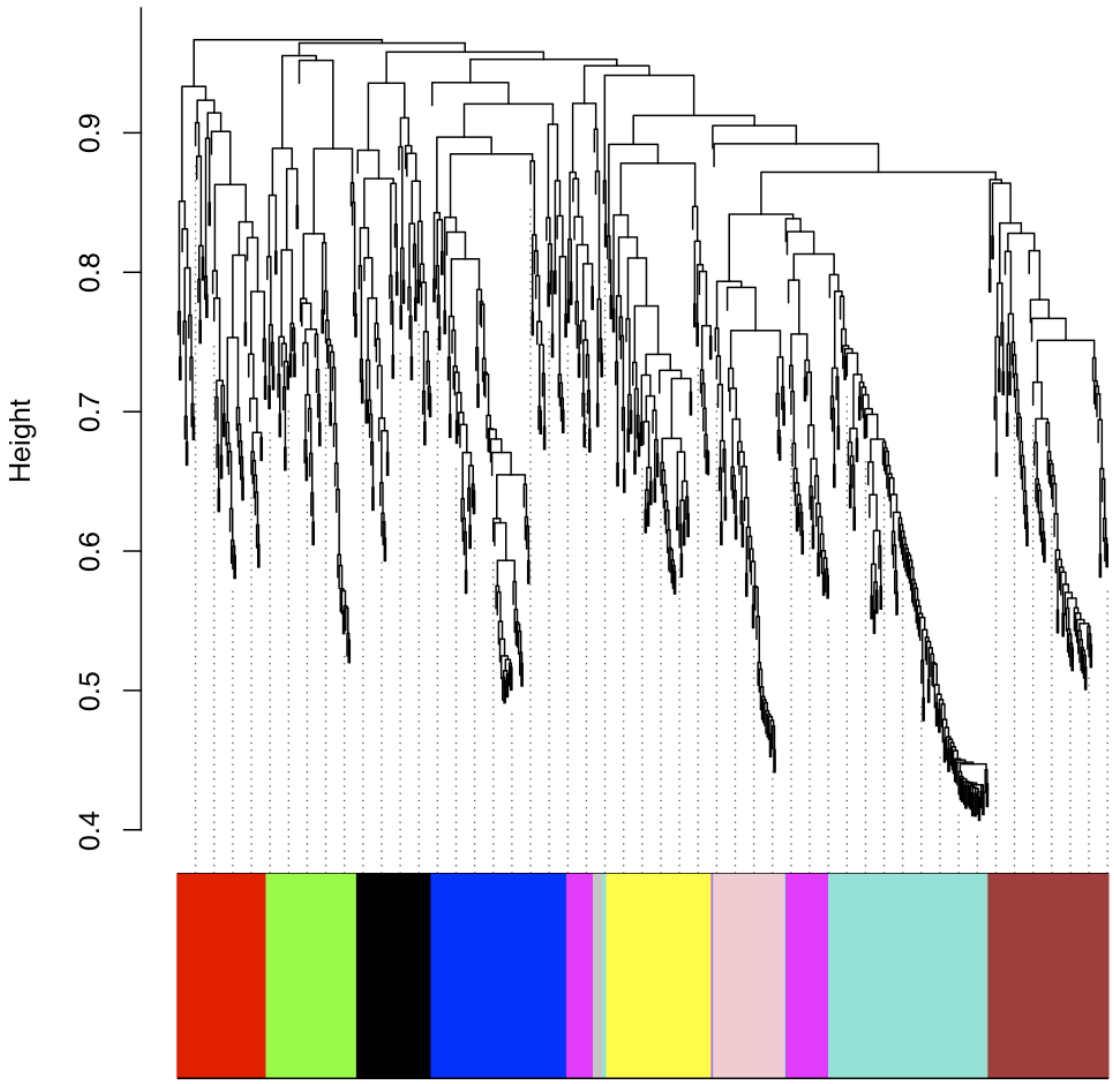
Twelve modules were identified and labeled with colors as shown below the dendrogram.

Figure 4.6. Dendrogram of 794 orthologous genes that were up-regulated in the H_2-producing, low-light (NF-low) condition compared to the H_2-producing, high-light (NF-high) condition.



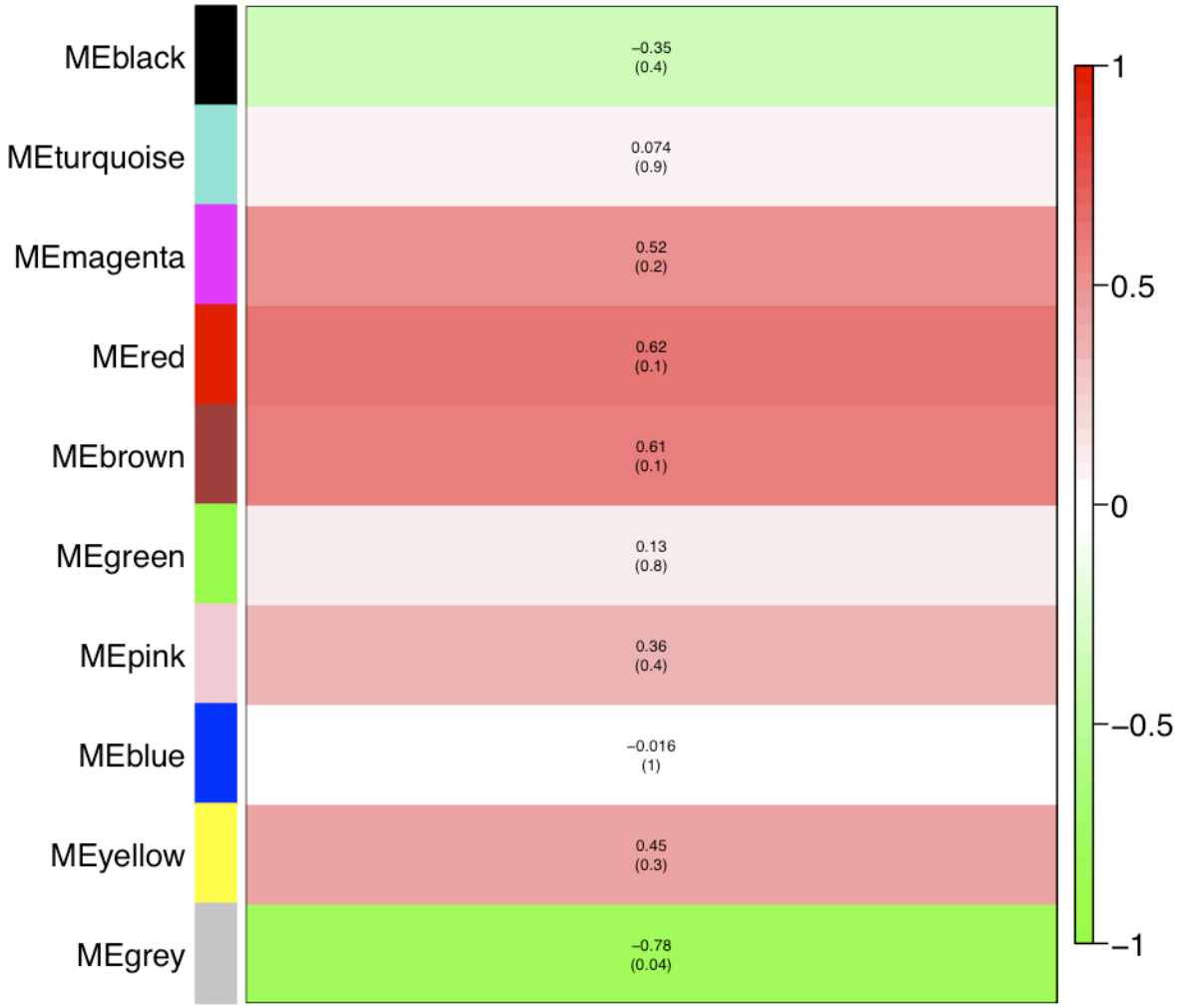
Nine modules were identified and labeled with colors as shown below the dendrogram.

Figure 4.7. Dendrogram of 700 orthologous genes that were down-regulated in the H_2-producing, low-light (NF-low) condition compared to the H_2-producing, high-light (NF-high) condition.



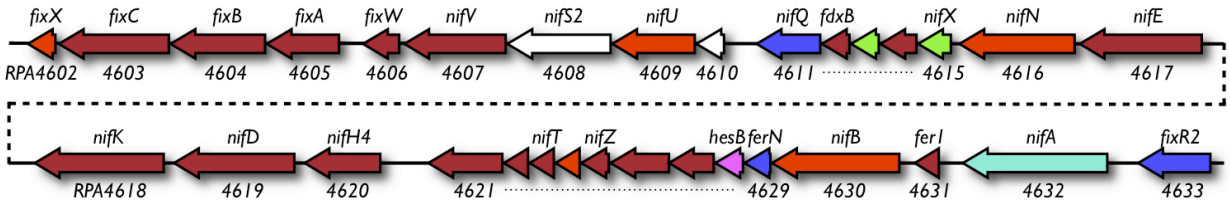
Ten modules were identified and labeled with colors as shown in the below the dendrogram.

Figure 4.8. The association between module eigengenes in the up-regulated H_2-producing, high-light versus non-H_2-producing, high light (NF-high/PM-high) networks and the level of nitrogenase activity in the H_2-producing, high light condition.



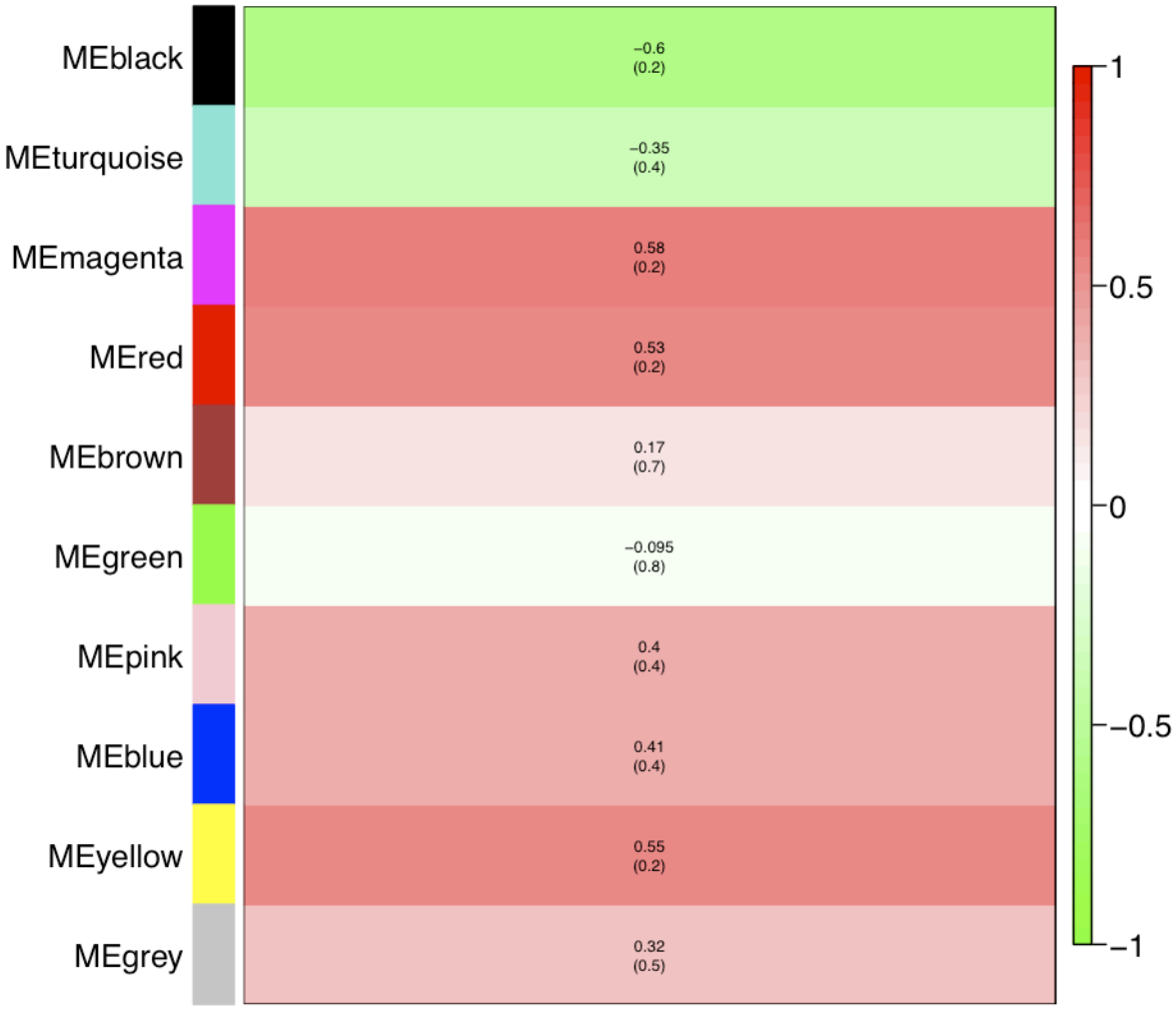
Each row corresponds to a module eigengene identified by its color. Each cell contains the corresponding correlation coefficient and *p*-value, and is color-coded: positive correlations are denoted in red and negative correlation in green.

Figure 4.9. Organization of molybdenum nitrogenase gene cluster in *Rhodopseudomonas* strain CGA009.



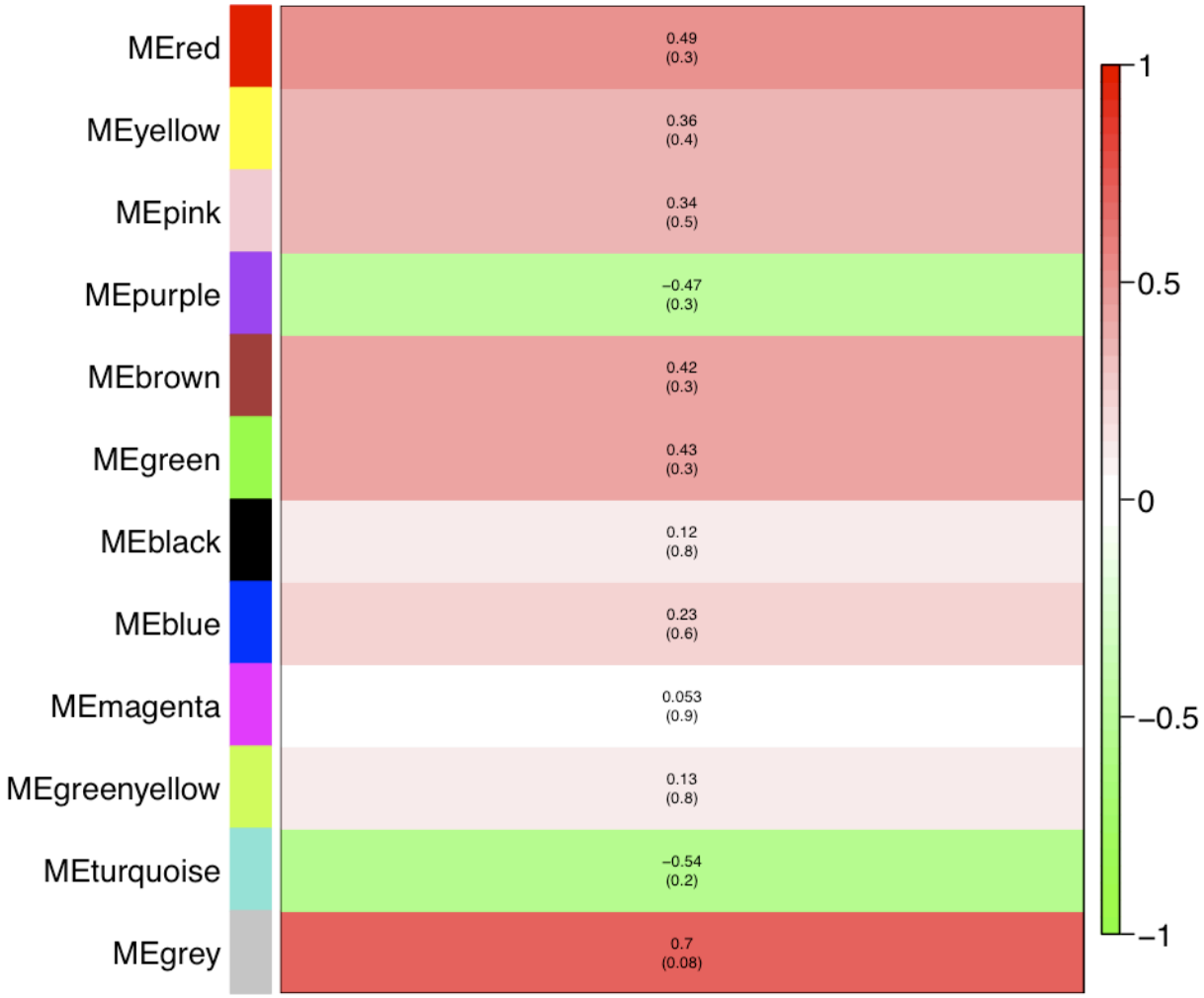
Genes are colored to match the module to which they belong.

Figure 4.10. The association between module eigengenes in the up-regulated H_2-producing, high light versus non-H_2-producing, high light (NF-high/PM-high) networks and the H_2 yields in the H_2-producing, high light condition.



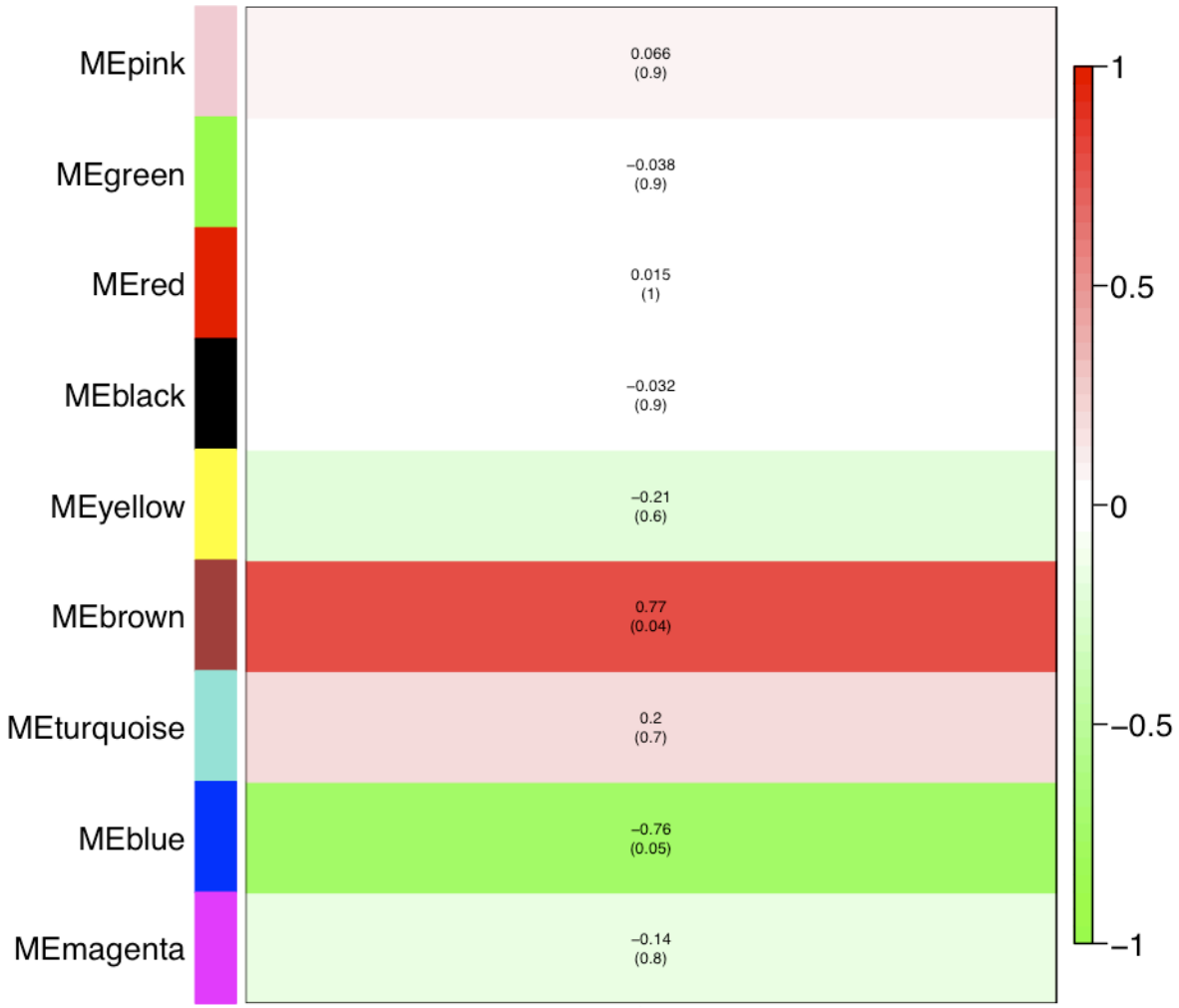
Each row corresponds to a module eigengene identified by its color. Each cell contains the corresponding correlation coefficient and *p*-value, and is color-coded: positive correlations are denoted in red and negative correlation in green.

Figure 4.11. The association between module eigengenes in the down-regulated H_2-producing, high light versus non-H_2-producing, high light (NF-high/PM-high) networks and the H_2 yields in the H_2-producing, high light condition.



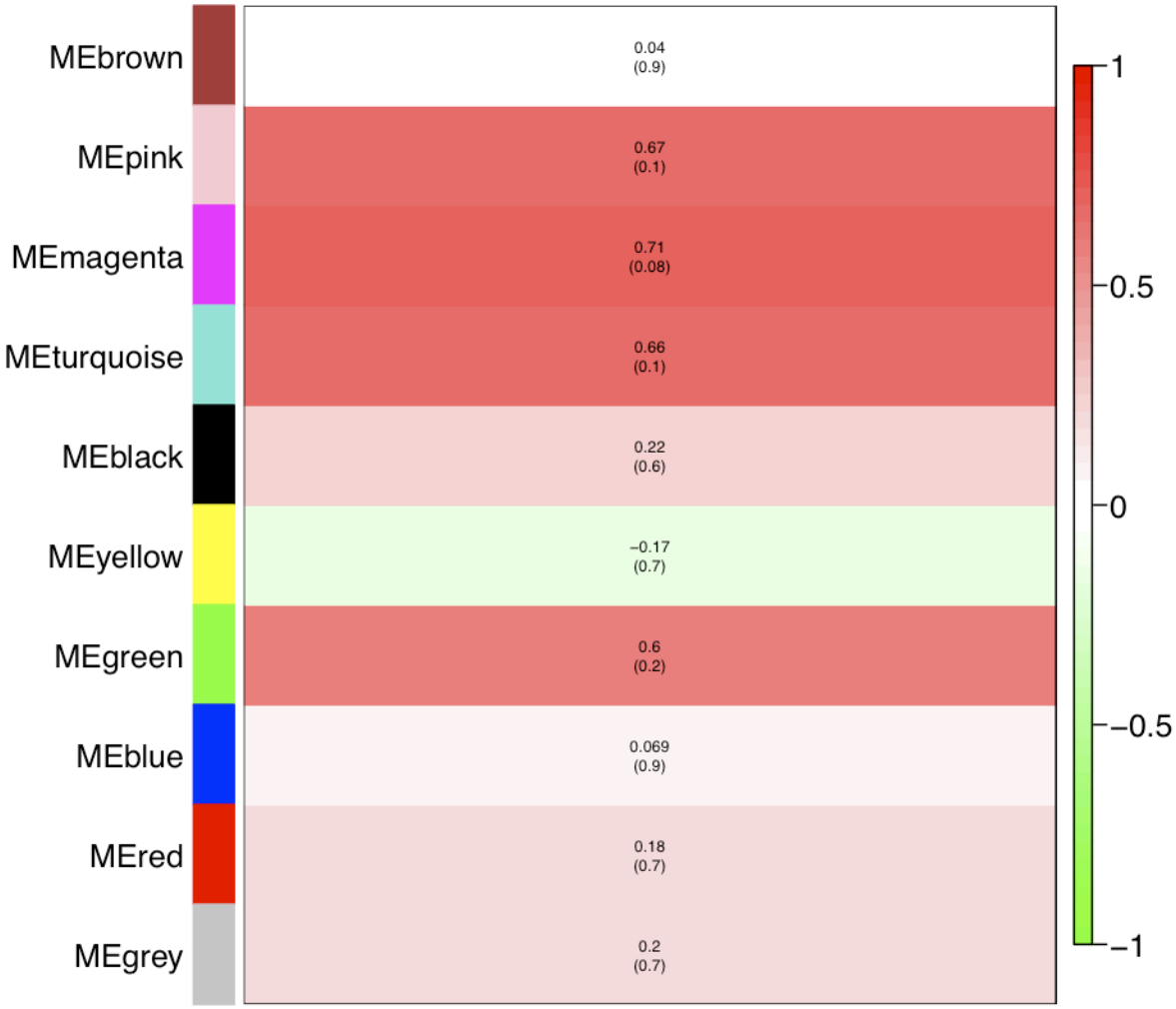
Each row corresponds to a module eigengene identified by its color. Each cell contains the corresponding correlation coefficient and *p*-value, and is color-coded: positive correlations are denoted in red and negative correlation in green.

Figure 4.12. The association between module eigengenes in the up-regulated H_2-producing, low light versus H_2-producing, high light (NF-low/NF-high) networks and the ratios of H_2yields in the NF-low to NF-high condition.



Each row corresponds to a module eigengene identified by its color. Each cell contains the corresponding correlation coefficient and *p*-value, and is color-coded: positive correlations are denoted in red and negative correlation in green.

Figure 4.13. The association between module eigengenes in the down-regulated H_2-producing, low light versus H_2-producing, high light (NF-low/NF-high) networks and the ratios of H_2 yields in the NF-low to NF-high condition.



Each row corresponds to a module eigengene identified by its color. Each cell contains the corresponding correlation coefficient and *p*-value, and is color-coded: positive correlations are denoted in red and negative correlation in green.

Table 4.1. Nitrogenase activities and H_2 yields of seven *Rhodopseudomonas* strains under H_2-producing, high light (NF-high) and H_2-producing, low light (NF-low) conditions.

Strain	Nitrogenase Activity under NF-high Condition [a]	H_2 Yield under NF-high Condition [b]	H_2 Yield under NF-low Condition [b]
CGA009	101.2	101.6	84.9
CGA010	97.8	93.3	73.6
TIE-1	81.4	62.2	40.5
DX-1	72.5	93.3	0.1
BisB18	73.9	25.4	37.6
BisB5	82.1	50.5	48.4
BisA53	39.9	80.1	71.9

[a] Nitrogenase activity in nmol C_2H_4 formed/min/mg protein.
[b] H_2 yield in μmol/mg protein.

Table 4.2. Seven *Rhodopseudomonas* strains have 16S-rRNA sequence identities of 97.3% or greater.

Strain	16S-rRNA Identity	Link
CGA009	100%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=62901
CGA010	100%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=62901
TIE-1	100%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=58995
DX-1	99.4%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=43327
BisB18	97.3%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=58443
BisB5	97.5%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=58441
BisA53	97.8%	http://www.ncbi.nlm.nih.gov/genome/508?project_id=58445

Table 4.3. Statistics of the NF-high/PM-high and the NF-low/NF-high co-expression networks.

Networks	No. Nodes	No. Modules	Size of the Largest	Size of the Smallest
NF-high/PM-high (up-regulated)	750	10	186	39
NF-high/PM-high (down-regulated)	732	12	147	33
NF-low/NF-high (up-regulated)	794	9	175	35
NF-low/NF-high (down-regulated)	700	10	125	53

Table 4.4. List of module members in the up-regulated NF-high/PM-high co-expression networks.

orthoMCL ID	RPA Number	Product Description	Module Color
orthoMCL2719	NA	NA	black
orthoMCL3741	NA	NA	black
orthoMCL4711	NA	NA	black
orthoMCL4891	NA	NA	black
orthoMCL4898	NA	NA	black
orthoMCL4954	NA	NA	black
orthoMCL5115	NA	NA	black
orthoMCL5129	NA	NA	black
orthoMCL5133	NA	NA	black
orthoMCL5144	NA	NA	black
orthoMCL5146	NA	NA	black
orthoMCL5151	NA	NA	black
orthoMCL5159	NA	NA	black
orthoMCL5184	NA	NA	black
orthoMCL5241	NA	NA	black
orthoMCL0052	RPA0139	methyl-accepting chemotaxis receptor/sensory transducer	black
orthoMCL3644	RPA0141	chemotaxis signal transduction/oligomerization protein CheW1-1	black
orthoMCL3643	RPA0142	multidomain chemotaxis histidine kinase CheA1 (Hpt, CheA, & CheW domains)	black
orthoMCL3642	RPA0143	response regulator receiver, CheY1	black
orthoMCL3641	RPA0144	Sulfate transporter/antisigma-factor antagonist domain	black
orthoMCL2604	RPA0149	possible ABC-type iron-siderophore transport system ATP-binding protein	black
orthoMCL4100	RPA0153	possible outer membrane receptor for Fe transport	black
orthoMCL3627	RPA0418	Glutamine amidotransferase class-I	black
orthoMCL3621	RPA0514	putative efflux pump protein FarB	black
orthoMCL3620	RPA0515	possible efflux pump protein FarA	black
orthoMCL4638	RPA0587	putative cationic amino acid transporter	black
orthoMCL3132	RPA0657	benzoyl-CoA reductase subunit	black
orthoMCL3596	RPA0745	possible outer membrane protein precursor	black
orthoMCL3593	RPA0757	hypothetical protein	black
orthoMCL2017	RPA1055	quinolinate synthetase A	black
orthoMCL3063	RPA1214	putative ABC transporter ATP-binding protein	black
orthoMCL3042	RPA1374	Sigma-54 dependent, Vanadium nitrogenase transcriptional regulator, VnfA	black
orthoMCL0001	RPA1491	light harvesting protein B-800-850, beta chain E (antenna pigment protein, beta chain E) (LH II-E beta)	black
orthoMCL0043	RPA1494	unknown protein	black
orthoMCL3011	RPA1719	Protein of unknown function UPF0153	black
orthoMCL1670	RPA1774	OmpA/MotB domain, possible porin	black
orthoMCL1669	RPA1776	possible lipid transfer protein	black
orthoMCL0087	RPA1964	hypothetical protein	black
orthoMCL2941	RPA2289	hypothetical protein	black
orthoMCL3456	RPA2299	possible two-component transcriptional regulator, LuxR family	black

orthoMCL3433	RPA2415	Acetamidase/Formamidase	black
orthoMCL1401	RPA2464	sufB, needed for fhuF Fe-S center stability	black
orthoMCL0077	RPA2790	pH adaptation K efflux system component	black
orthoMCL1078	RPA3052	phosphoribosylglycinamide formyltransferase	black
orthoMCL0886	RPA3426	conserved hypothetical protein	black
orthoMCL2863	RPA3439	putative hydrolase	black
orthoMCL0656	RPA3807	putative permease of ABC transporter (high-affinity branched-chain amino acid transport)	black
orthoMCL0654	RPA3809	ATP-binding component of ABC transporter (putatively for branched chain amino acids)	black
orthoMCL3291	RPA4080	hypothetical protein	black
orthoMCL5540	RPA4090	conserved hypothetical protein	black
orthoMCL5537	RPA4093	hypothetical protein	black
orthoMCL5529	RPA4105	putative hlyD family multidrug secretion protein	black
orthoMCL2748	RPA4635	ferrous iron transport protein B	black
orthoMCL3241	RPA4684	methyl-accepting chemotaxis receptor/sensory transducer	black
orthoMCL0192	RPA4727	glycogen phosphorylase	black
orthoMCL2689	NA	NA	blue
orthoMCL2695	NA	NA	blue
orthoMCL2697	NA	NA	blue
orthoMCL2698	NA	NA	blue
orthoMCL2701	NA	NA	blue
orthoMCL2722	NA	NA	blue
orthoMCL3199	NA	NA	blue
orthoMCL3211	NA	NA	blue
orthoMCL3692	NA	NA	blue
orthoMCL4136	NA	NA	blue
orthoMCL4257	NA	NA	blue
orthoMCL4665	NA	NA	blue
orthoMCL4666	NA	NA	blue
orthoMCL4683	NA	NA	blue
orthoMCL4776	NA	NA	blue
orthoMCL5279	NA	NA	blue
orthoMCL5299	NA	NA	blue
orthoMCL5331	NA	NA	blue
orthoMCL2603	RPA0150	putative ABC transporter, iron, hemin permease homolog	blue
orthoMCL4082	RPA0601	probable DMT superfamily transporter	blue
orthoMCL2269	RPA0602	Permeases of the drug/metabolite transporter (DMT) superfamily	blue
orthoMCL3588	RPA0764	conserved hypothetical protein	blue
orthoMCL3115	RPA0765	putative outer membrane receptor for iron transport	blue
orthoMCL2180	RPA0783	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL)	blue
orthoMCL2149	RPA0828	transcriptional regulator, MarR family	blue
orthoMCL3111	RPA0829	organic hydroperoxide resistance protein	blue
orthoMCL2143	RPA0836	cytochrome c oxidase subunit III	blue
orthoMCL3098	RPA0962	hydrogenase small chain	blue
orthoMCL3097	RPA0963	hydrogenase large chain	blue
orthoMCL3096	RPA0964	Ni/Fe-hydrogenase 1 B-type cytochrome subunit	blue

orthoMCL3095	RPA0965	hydrogenase maturation protein hupD	blue
orthoMCL4071	RPA0966	putative membrane-bound hydrogenase component hupE	blue
orthoMCL3094	RPA0967	hydrogenase expression/formation protein hupF	blue
orthoMCL3092	RPA0969	hydrogenase expression/formation protein hupH	blue
orthoMCL3091	RPA0970	putative rubredoxin hupI	blue
orthoMCL3090	RPA0971	putative hydrogenase expression/formation protein hupJ	blue
orthoMCL3089	RPA0972	putative hydrogenase expression/formation protein hupK	blue
orthoMCL3088	RPA0973	hydrogenase formation/expression protein hypA	blue
orthoMCL3087	RPA0974	hydrogenase expression/formation protein hypB	blue
orthoMCL3086	RPA0975	hydrogenase maturation protein hypF	blue
orthoMCL3085	RPA0976	putative hypC	blue
orthoMCL3084	RPA0977	hydrogenase expression/formation protein hypD	blue
orthoMCL3083	RPA0978	hydrogenase expression/formation protein hypE	blue
orthoMCL3082	RPA0979	two component sigma-54-dependent hydrogenase transcriptional regulator hoxA, Fis family	blue
orthoMCL3081	RPA0980	sensor histidine kinase with a PAS domain	blue
orthoMCL3079	RPA0991	possible transcriptional regulator, Crp/Fnr family	blue
orthoMCL2050	RPA1000	Nitrogenase-associated protein:Arsenate reductase and related	blue
orthoMCL0097	RPA1016	ubiquinol-cytochrome-c reductase, Rieske iron-sulfur protein	blue
orthoMCL3032	RPA1428	possible lipoprotein	blue
orthoMCL3531	RPA1467	transcriptional regulator, GntR family	blue
orthoMCL4013	RPA1470	putative dipeptide ABC transporter	blue
orthoMCL1703	RPA1703	putative acetyl-CoA acyltransferase	blue
orthoMCL1702	RPA1704	probable transcriptional regulator, TetR family	blue
orthoMCL1675	RPA1765	putative enoyl-CoA hydratase	blue
orthoMCL1674	RPA1769	transcriptional regulator, LysR family	blue
orthoMCL2989	RPA1791	branched-chain amino acid transport system ATP-binding protein	blue
orthoMCL4573	RPA1874	hypothetical protein	blue
orthoMCL4572	RPA1875	possible uncharacterized iron-regulated membrane protein	blue
orthoMCL1604	RPA1913	conserved hypothetical protein	blue
orthoMCL1574	RPA1962	unknown protein	blue
orthoMCL4564	RPA2053	conserved hypothetical protein	blue
orthoMCL2958	RPA2116	hypothetical protein	blue
orthoMCL2957	RPA2117	putative flavodoxin	blue
orthoMCL1500	RPA2118	putative ATP-binding protein of ABC transporter	blue
orthoMCL1499	RPA2119	putative permease protein of ABC transporter	blue
orthoMCL1498	RPA2120	putative hemin binding protein	blue
orthoMCL1496	RPA2122	putative oxygen independent coproporphyrinogen III oxidase	blue
orthoMCL1495	RPA2123	conserved unknown protein	blue
orthoMCL1494	RPA2124	tonB dependent iron siderophore receptor	blue
orthoMCL1493	RPA2125	conserved unknown protein	blue
orthoMCL1492	RPA2126	conserved unknown protein	blue
orthoMCL3466	RPA2127	putative exbB, uptake of enterochelin	blue
orthoMCL1491	RPA2128	biopolymer transport protein ExbD/TolR	blue
orthoMCL3965	RPA2129	possible energy transducer TonB, C-terminal region	blue
orthoMCL4551	RPA2130	DUF81	blue
orthoMCL1477	RPA2151	Permeases of the drug/metabolite transporter (DMT) superfamily	blue

orthoMCL3463	RPA2162	possible serine protease/outer membrane autotransporter	blue
orthoMCL1463	RPA2175	Transglutaminase-like domain	blue
orthoMCL1462	RPA2176	DUF403	blue
orthoMCL1461	RPA2177	DUF404	blue
orthoMCL4543	RPA2288	Streptomyces cyclase/dehydrase	blue
orthoMCL2940	RPA2290	conserved hypothetical protein	blue
orthoMCL3953	RPA2295	unknown protein	blue
orthoMCL4540	RPA2307	possible tonB-dependent receptor precursor	blue
orthoMCL4539	RPA2308	possible periplasmic iron siderophore binding protein of ABC transporter	blue
orthoMCL4538	RPA2311	hypothetical protein	blue
orthoMCL4537	RPA2312	hypothetical protein	blue
orthoMCL4536	RPA2313	unknown protein	blue
orthoMCL3940	RPA2377	conserved hypothetical protein	blue
orthoMCL3939	RPA2378	putative tonB-dependent receptor protein	blue
orthoMCL3938	RPA2379	probable acetyltransferase	blue
orthoMCL3937	RPA2380	probable tonB dependent iron siderophore receptor	blue
orthoMCL3936	RPA2381	probable FecR, iron siderophore sensor protein	blue
orthoMCL3935	RPA2382	putative iron(III) ABC transporter, ATP-binding protein	blue
orthoMCL3934	RPA2383	putative iron(III) ABC transporter, permease protein	blue
orthoMCL3932	RPA2385	putative ABC transporter, periplasmic Fe+3 siderophore binding protein	blue
orthoMCL3931	RPA2387	conserved hypothetical protein	blue
orthoMCL3930	RPA2388	possible acyl-CoA ligase for activation during siderophore synthesis	blue
orthoMCL3929	RPA2389	possible Rhizobactin siderophore biosynthesis protein RhsF	blue
orthoMCL3928	RPA2390	possible Rhizobactin siderophore biosynthesis protein rhbC	blue
orthoMCL3440	RPA2407	hypothetical protein	blue
orthoMCL1380	RPA2510	conserved hypothetical protein	blue
orthoMCL1319	RPA2649	conserved unknown protein	blue
orthoMCL4491	RPA2708a	hypothetical membrane protein	blue
orthoMCL1249	RPA2764	CoA Binding Domain	blue
orthoMCL3394	RPA2923	conserved hypothetical protein	blue
orthoMCL1113	RPA2967	glutamine synthetase I	blue
orthoMCL1046	RPA3094	putative tetracycline-efflux transporter	blue
orthoMCL0996	RPA3164	possible chitooligosaccharide deacetylase	blue
orthoMCL0992	RPA3173	putative protein	blue
orthoMCL0967	RPA3227	30S ribosomal protein S11	blue
orthoMCL0965	RPA3229	Adenylate kinase	blue
orthoMCL0964	RPA3230	secretion protein SecY	blue
orthoMCL0937	RPA3267	RNA polymerase beta' subunit	blue
orthoMCL0934	RPA3270	50S ribosomal protein L10	blue
orthoMCL4435	RPA3355	putative exopolysaccharide polymerization protein	blue
orthoMCL4434	RPA3356	unknown protein	blue
orthoMCL4432	RPA3359	O-antigen polymerase	blue
orthoMCL2868	RPA3399	cold shock DNA binding protein	blue
orthoMCL4426	RPA3400	hypothetical protein	blue
orthoMCL3353	RPA3414	putative hydroxamate-type ferrisiderophore receptor	blue

orthoMCL3351	RPA3453	putative enoyl-CoA hydratase	blue
orthoMCL3343	RPA3476	possible energy transducer TonB	blue
orthoMCL3342	RPA3477	exbD, uptake of enterochelin	blue
orthoMCL3341	RPA3478	possible exbB, uptake of enterochelin	blue
orthoMCL2856	RPA3479	2OG-Fe(II) oxygenase superfamily:Prolyl 4-hydroxylase, alpha subunit	blue
orthoMCL3340	RPA3481	hypothetical protein	blue
orthoMCL0799	RPA3576	thiamin phosphate pyrophosphorylase	blue
orthoMCL0754	RPA3662	urease beta subunit	blue
orthoMCL0749	RPA3667	possible permease of ABC transporter	blue
orthoMCL0720	RPA3702	methionine synthase	blue
orthoMCL0636	RPA3849	glycine cleavage system protein H	blue
orthoMCL2827	RPA3866	conserved unknown protein	blue
orthoMCL0603	RPA3909	flagellar P-ring protein FlgI	blue
orthoMCL0602	RPA3910	conserved hypothetical protein	blue
orthoMCL4388	RPA4131	Helix-turn-helix protein, CopG family	blue
orthoMCL2793	RPA4209	glutamine synthetase II	blue
orthoMCL2792	RPA4211	glutaminase A	blue
orthoMCL0401	RPA4345	hypothetical protein	blue
orthoMCL2766	RPA4470	DUF336	blue
orthoMCL2765	RPA4471	conserved hypothetical protein	blue
orthoMCL0305	RPA4495	possible protease htpX homolog	blue
orthoMCL0280	RPA4568	putative enoyl-acyl carrier protein reductase	blue
orthoMCL0273	RPA4601	conserved hypothetical protein	blue
orthoMCL0263	RPA4611	putative nitrogen fixation protein nifQ	blue
orthoMCL0247	RPA4629	ferredoxin 2[4Fe-4S], fdxN	blue
orthoMCL0243	RPA4633	short-chain dehydrogenase	blue
orthoMCL0138	RPA4792	RNA polymerase ECF-type sigma factor	blue
orthoMCL0129	RPA4812	possible glutathione S-transferase	blue
orthoMCL3770	NA	NA	brown
orthoMCL2606	RPA0090	hypothetical protein	brown
orthoMCL0053	RPA0091	hypothetical protein	brown
orthoMCL3183	RPA0114	hypothetical protein	brown
orthoMCL2498	RPA0274	GlnK, nitrogen regulatory protein P-II	brown
orthoMCL2387	RPA0424	transcriptional regulator, FUR family	brown
orthoMCL3594	RPA0755	putative oligopeptide ABC transporter, ATP-binding component	brown
orthoMCL3592	RPA0758	putative oligopeptide ABC transporter, ATP-binding component	brown
orthoMCL3116	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component	brown
orthoMCL2146	RPA0832	cytochrome c oxidase subunit I	brown
orthoMCL2092	RPA0906	putative N-formylglutamate amidohydrolase	brown
orthoMCL4063	RPA0989	putative ATP-binding component of ABC transporter	brown
orthoMCL1985	RPA1114	conserved unknown protein	brown
orthoMCL1903	RPA1210	conserved hypothetical protein	brown
orthoMCL1902	RPA1211	hypothetical protein	brown
orthoMCL1901	RPA1212	Alpha/beta hydrolase fold	brown
orthoMCL3059	RPA1218	putative ABC transporter periplasmic protein	brown

orthoMCL1874	RPA1270	conserved hypothetical protein	brown
orthoMCL0095	RPA1420	putative inner membrane component for iron transport	brown
orthoMCL1855	RPA1421	possible efflux protein	brown
orthoMCL1854	RPA1422	unknown protein	brown
orthoMCL4011	RPA1472	putative dipeptide transport permease protein	brown
orthoMCL5740	RPA1564	possible urea/short-chain amide transport system substrate-binding protein	brown
orthoMCL4004	RPA1642	putative diguanylate cyclase (GGDEF)	brown
orthoMCL1723	RPA1662	hypothetical protein	brown
orthoMCL1706	RPA1693	superoxide dismutase	brown
orthoMCL1684	RPA1748	putative branched-chain amino acid transport system substrate-binding protein	brown
orthoMCL2991	RPA1788	possible 4-hydroxybenzoyl-CoA thioesterase	brown
orthoMCL1622	RPA1879	Choloylglycine hydrolase	brown
orthoMCL5718	RPA1926	possible transcriptional regulator, XRE family	brown
orthoMCL1594	RPA1927	hypothetical protein	brown
orthoMCL1593	RPA1928	ferredoxin-like protein [2Fe-2S]	brown
orthoMCL1474	RPA2156	hypothetical protein	brown
orthoMCL3457	RPA2298	hypothetical protein	brown
orthoMCL3439	RPA2408	putative aliphatic amidase expression-regulating protein, AmiC	brown
orthoMCL1400	RPA2465	sufC, related to ABC transporter ATP-binding protein	brown
orthoMCL1399	RPA2466	sufD, needed for fhuF Fe-S center production/stability	brown
orthoMCL1398	RPA2467	sufS, putative selenosysteine lyase	brown
orthoMCL1397	RPA2468	DUF59	brown
orthoMCL1396	RPA2470	Protein of unknown function, HesB/YadR/YfhF	brown
orthoMCL1389	RPA2498	possible ABC transporter, permease protein	brown
orthoMCL4510	RPA2499	possible ABC transporter, periplasmic protein	brown
orthoMCL4509	RPA2500	possible amidase	brown
orthoMCL3913	RPA2550	hypothetical protein	brown
orthoMCL3904	RPA2677	putative substrate-binding protein, subunit of ABC transporter	brown
orthoMCL3903	RPA2678	putative permease protein, subunit of ABC transporter	brown
orthoMCL2920	RPA2756	hypothetical protein	brown
orthoMCL1224	RPA2815	possible outer membrane protein	brown
orthoMCL4483	RPA2859	hypothetical protein	brown
orthoMCL2880	RPA3199	unknown protein	brown
orthoMCL5601	RPA3211	TPR repeat	brown
orthoMCL3369	RPA3257	probable transcriptional regulator, AraC family	brown
orthoMCL5597	RPA3277	putative exbD, uptake of enterochelin	brown
orthoMCL4450	RPA3331	hypothetical protein	brown
orthoMCL4449	RPA3332	hypothetical protein	brown
orthoMCL0911	RPA3372	hypothetical protein	brown
orthoMCL5584	RPA3373	hypothetical protein	brown
orthoMCL3859	RPA3384	possible ABC related periplasmic binding protein	brown
orthoMCL5572	RPA3596	hypothetical protein	brown
orthoMCL0794	RPA3600	bacterioferritin	brown
orthoMCL0685	RPA3767	phenylacetic acid degradation protein paaB	brown
orthoMCL0060	RPA3788	putative ABC transporter, ATP-binding protein	brown

orthoMCL3829	RPA3860	hypothetical protein	brown
orthoMCL0525	RPA4077	ATPase, ParA type	brown
orthoMCL3283	RPA4162	putative taurine ABC transport system permease protein	brown
orthoMCL4385	RPA4164	possible aliphatic sulfonate binding protein of ABC transporter system	brown
orthoMCL0476	RPA4219	conserved hypothetical protein	brown
orthoMCL3266	RPA4416	conserved hypothetical protein	brown
orthoMCL4371	RPA4429	Uncharacterized iron-regulated membrane protein DUF337	brown
orthoMCL0271	RPA4603	nitrogen fixation protein,fixC	brown
orthoMCL0270	RPA4604	electron transfer flavoprotein alpha chain protein fixB	brown
orthoMCL0269	RPA4605	electron trnasfer flavoprotein beta chain fixA	brown
orthoMCL0268	RPA4606	nitrogenase stabilizer NifW	brown
orthoMCL0267	RPA4607	putative homocitrate synthase	brown
orthoMCL0262	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB	brown
orthoMCL0260	RPA4614	DUF269	brown
orthoMCL0257	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE	brown
orthoMCL0256	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK	brown
orthoMCL0255	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD	brown
orthoMCL0254	RPA4620	nitrogenase iron protein, nifH	brown
orthoMCL0253	RPA4621	conserved hypothetical protein	brown
orthoMCL2749	RPA4622	hypothetical protein	brown
orthoMCL0252	RPA4623	conserved hypothetical protein	brown
orthoMCL0250	RPA4625	NifZ domain	brown
orthoMCL3245	RPA4626	Protein of unknown function from Deinococcus and Synechococcus	brown
orthoMCL0249	RPA4627	conserved hypothetical protein	brown
orthoMCL0245	RPA4631	ferredoxin 2[4Fe-4S], fdxN	brown
orthoMCL0031	RPA4634	hypothetical protein	brown
orthoMCL2747	RPA4636	FeoA family	brown
orthoMCL2733	RPA4740	putative 4-carboxymuconolactone decarboxylase	brown
orthoMCL0122	RPA4827	conserved hypothetical protein	brown
orthoMCL2723	RPA4828	conserved hypothetical protein	brown
orthoMCL3208	NA	NA	green
orthoMCL3710	NA	NA	green
orthoMCL3711	NA	NA	green
orthoMCL4191	NA	NA	green
orthoMCL4200	NA	NA	green
orthoMCL4664	NA	NA	green
orthoMCL4676	NA	NA	green
orthoMCL4681	NA	NA	green
orthoMCL4733	NA	NA	green
orthoMCL4841	NA	NA	green
orthoMCL2541	RPA0220	putative cation efflux system protein	green
orthoMCL3164	RPA0440	conserved hypothetical protein	green
orthoMCL4644	RPA0472	OmpA-like transmembrane domain	green
orthoMCL3619	RPA0516	transcriptional regulator, MarR family	green
orthoMCL2300	RPA0556	HNH endonuclease:HNH nuclease	green
orthoMCL2187	RPA0756	putative amidase	green

orthoMCL4069	RPA0983	possible phthalate dioxygenase	green
orthoMCL4068	RPA0984	putative glutamine synthetase-like protein	green
orthoMCL4067	RPA0985	putative branched-chain amino acid transport system substrate-binding protein	green
orthoMCL1933	RPA1173	cold shock DNA binding protein	green
orthoMCL4601	RPA1411	possible enoyl-CoA hydratase/isomerase	green
orthoMCL4015	RPA1468	conserved hypothetical protein	green
orthoMCL5742	RPA1562	transcriptional regulator, LysR family	green
orthoMCL1651	RPA1822	methyl-accepting chemotaxis receptor/sensory transducer	green
orthoMCL3986	RPA1850	methyl-accepting chemotaxis receptor/sensory transducer	green
orthoMCL1535	RPA2023	conserved hypothetical protein	green
orthoMCL1475	RPA2155	dihydroxy-acid dehydratase	green
orthoMCL4530	RPA2337	hypothetical protein	green
orthoMCL4529	RPA2338	unknown protein	green
orthoMCL4528	RPA2339	transcriptional regulator, FUR family	green
orthoMCL5683	RPA2340	putative membrane protein	green
orthoMCL1373	RPA2523	putative lactoylglutathione lyase	green
orthoMCL1096	RPA3031	possible Acetyltransferase (GNAT) family	green
orthoMCL3371	RPA3201	formate/nitrate transporter	green
orthoMCL4453	RPA3328	possible transcriptional regulator, XRE family	green
orthoMCL4446	RPA3335	hypothetical protein	green
orthoMCL3866	RPA3349	putative exopolysaccharide biosynthesis protein	green
orthoMCL3865	RPA3350	putative aminotransferase	green
orthoMCL3864	RPA3352	glycosyltransferase	green
orthoMCL3863	RPA3353	putative serine/threonine protein phosphatase	green
orthoMCL4436	RPA3354	hypothetical protein	green
orthoMCL4425	RPA3434	Universal stress protein (Usp)	green
orthoMCL2855	RPA3480	putative outer membrane receptor for iron transport	green
orthoMCL4415	RPA3587	hypothetical protein	green
orthoMCL0721	RPA3701	putative 5,10-methylenetetrahydrofolate reductase	green
orthoMCL0679	RPA3774	possible D-amino-acid dehydrogenase	green
orthoMCL5517	RPA4150	conserved hypothetical protein	green
orthoMCL0517	RPA4152	periplasmic iron binding protein FbpA precursor	green
orthoMCL5515	RPA4169	Bacterial regulatory protein, GntR family	green
orthoMCL0488	RPA4196	TPR repeat	green
orthoMCL4376	RPA4296	unknown protein	green
orthoMCL3800	RPA4344	hypothetical protein	green
orthoMCL0360	RPA4394	isocitrate lyase	green
orthoMCL0359	RPA4395	transcriptional regulator, XRE family	green
orthoMCL4370	RPA4430	putative TonB-dependent receptor	green
orthoMCL4362	RPA4523	response regulator receiver (CheY-like protein)	green
orthoMCL0261	RPA4613	DUF683	green
orthoMCL0259	RPA4615	nitrogenase molybdenum-iron protein nifX	green
orthoMCL3777	RPA4640	hypothetical protein	green
orthoMCL2743	RPA4696	conserved unknown protein	green
orthoMCL4314	NA	NA	grey
orthoMCL2199	RPA0732	putative NAD-dependent formate dehydrogenase gamma subunit	grey

orthoMCL3954	RPA2293	hypothetical protein	grey
orthoMCL4119	RPA0102	putative ABC transporter oligopeptide-binding protein	magenta
orthoMCL2553	RPA0207	unknown protein	magenta
orthoMCL0102	RPA0763	possible methyltransferases	magenta
orthoMCL2100	RPA0890	hypothetical protein	magenta
orthoMCL2048	RPA1002	putative molybdenum transport system protein	magenta
orthoMCL4059	RPA1059	probable outer membrane protein, TonB-dependent receptor	magenta
orthoMCL4056	RPA1082	conserved hypothetical protein	magenta
orthoMCL4054	RPA1084	possible minor curlin subunit precursor (fimbrin sef17 minor subunit).	magenta
orthoMCL3062	RPA1215	putative ABC transporter, permease protein	magenta
orthoMCL3529	RPA1476	putative periplasmic solute-binding protein	magenta
orthoMCL1764	RPA1604	conserved hypothetical protein	magenta
orthoMCL1729	RPA1650	putative ABC transporter, ATP-binding protein	magenta
orthoMCL1660	RPA1809	hypothetical protein	magenta
orthoMCL3500	RPA1847	conserved hypothetical protein	magenta
orthoMCL4567	RPA1957	alkanal monooxygenase (LuxA-like protein)	magenta
orthoMCL3486	RPA1960	putative RND efflux membrane protein	magenta
orthoMCL4562	RPA2074	response regulator receiver (CheY-like protein)	magenta
orthoMCL3469	RPA2113	possible nitrate transport system permease protein	magenta
orthoMCL2945	RPA2244	conserved hypothetical protein	magenta
orthoMCL4541	RPA2296	conserved hypothetical protein	magenta
orthoMCL3441	RPA2391	RNA polymerase ECF-type sigma factor, possible FecI	magenta
orthoMCL2924	RPA2706	manganese transport protein	magenta
orthoMCL3397	RPA2877	putative glycosyltransferase family protein	magenta
orthoMCL3882	RPA2987	conserved hypothetical protein	magenta
orthoMCL1094	RPA3034	unknown protein	magenta
orthoMCL1010	RPA3147	endopeptidase Clp: ATP-binding chain A	magenta
orthoMCL3857	RPA3386	possible amidohydrolase	magenta
orthoMCL0898	RPA3408	Nitroreductase family	magenta
orthoMCL3352	RPA3440	probable amidase	magenta
orthoMCL4417	RPA3552	putative short-chain dehydrogenase/reductase	magenta
orthoMCL0782	RPA3622	hypothetical protein	magenta
orthoMCL0755	RPA3660	urease alpha subunit	magenta
orthoMCL0725	RPA3693	putative cytochrome c	magenta
orthoMCL3832	RPA3839	putative transcriptional regulator PchR, AraC family	magenta
orthoMCL3803	RPA4165	conserved hypothetical protein	magenta
orthoMCL0459	RPA4241	CBS domain	magenta
orthoMCL0426	RPA4293	hypothetical protein	magenta
orthoMCL0248	RPA4628	Protein of unknown function, HesB/YadR/YfhF	magenta
orthoMCL0169	RPA4758	conserved hypothetical protein	magenta
orthoMCL3744	NA	NA	pink
orthoMCL4670	NA	NA	pink
orthoMCL4116	RPA0105	transcriptional regulator, GntR family with aminotransferase domain	pink
orthoMCL2548	RPA0212	ErfK/YbiS/YcfS/YnhG	pink
orthoMCL3587	RPA0766	transcriptional regulator, Crp/Fnr family	pink
orthoMCL2147	RPA0831	cytochrome c oxidase subunit II	pink

orthoMCL4064	RPA0988	putative branched-chain amino acid ABC transporter, ATP-binding protein	pink
orthoMCL2051	RPA0999	conserved hypothetical protein	pink
orthoMCL1914	RPA1194	putative carboxymethylenebutenolidase	pink
orthoMCL4010	RPA1473	possible peptide transport system substrate-binding protein	pink
orthoMCL1737	RPA1637	hypothetical protein	pink
orthoMCL3020	RPA1673	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	pink
orthoMCL1694	RPA1717	hypothetical protein	pink
orthoMCL1685	RPA1747	conserved hypothetical protein	pink
orthoMCL1497	RPA2121	conserved unknown protein	pink
orthoMCL4532	RPA2335	unknown protein	pink
orthoMCL4531	RPA2336	unknown protein	pink
orthoMCL3443	RPA2369	pseudogene-two-component transcriptional regulator, winged helix family	pink
orthoMCL3917	RPA2469	putative lactoylglutathione lyase	pink
orthoMCL4489	RPA2745	possible ferric siderophore receptor protein	pink
orthoMCL4485	RPA2794	putative pH adaptation potassium efflux system component phaG	pink
orthoMCL1114	RPA2966	nitrogen regulatory protein P-II	pink
orthoMCL2893	RPA3033	possible acetylornitine deacetylase	pink
orthoMCL1093	RPA3035	hypothetical protein	pink
orthoMCL1085	RPA3044	Protein of unknown function UPF0061	pink
orthoMCL4456	RPA3323	unknown protein	pink
orthoMCL4448	RPA3333	putative curli production assembly/transport component csgg precursor	pink
orthoMCL4443	RPA3341	putative acetyltransferase	pink
orthoMCL4439	RPA3346	possible mannosyltransferase	pink
orthoMCL4437	RPA3351	possible glycosyl transferase	pink
orthoMCL4430	RPA3362	unknown protein	pink
orthoMCL3860	RPA3383	possible taurine transport system permease protein	pink
orthoMCL0879	RPA3433	possible salicylate hydroxylase	pink
orthoMCL0814	RPA3548	possible serine protease/outer membrane autotransporter	pink
orthoMCL0789	RPA3610	conserved hypothetical protein	pink
orthoMCL0516	RPA4153	putative iron transport system permease protein	pink
orthoMCL4384	RPA4166	putative nitrilase	pink
orthoMCL4374	RPA4326	sulfate transporter family protein	pink
orthoMCL0323	RPA4442	ErfK/YbiS/YcfS/YnhG	pink
orthoMCL0279	RPA4569	DUF344	pink
orthoMCL3657	RPA0095	putative multidrug efflux membrane fusion protein	red
orthoMCL3173	RPA0276	PAP/25A core domain:DNA polymerase, beta-like region	red
orthoMCL3172	RPA0277	conserved hypothetical protein	red
orthoMCL2402	RPA0404	6-pyruvoyl-tetrahydropterin synthase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1)	red
orthoMCL4626	RPA0990	Bacteriophytochrome (light-regulated signal transduction histidine kinase), PhyB5	red
orthoMCL4046	RPA1206	aldehyde dehydrogenase	red
orthoMCL4621	RPA1207	PAS domain:sigma-54-dependent transcriptional regulator, Fis family	red
orthoMCL1905	RPA1208	putative dioxygenase	red
orthoMCL1904	RPA1209	conserved hypothetical protein	red

orthoMCL3048	RPA1275	conserved hypothetical protein	red
orthoMCL4602	RPA1393	conserved hypothetical protein	red
orthoMCL3530	RPA1475	hypothetical protein	red
orthoMCL1683	RPA1749	putative branched-chain amino acid transport system permease protein	red
orthoMCL1681	RPA1751	putative branched-chain amino acid transport system ATP-binding protein	red
orthoMCL5719	RPA1925	putative transmembrane protein	red
orthoMCL3488	RPA1958	transcriptional regulator, TetR family	red
orthoMCL3487	RPA1959	putative RND efflux membrane protein	red
orthoMCL1558	RPA1985	probable diacylglycerol kinase	red
orthoMCL3470	RPA2112	putative nitrate transporter component, nrtA	red
orthoMCL3947	RPA2306	Nickel-dependent hydrogenase b-type cytochrome subunit	red
orthoMCL3941	RPA2375	conserved hypothetical protein	red
orthoMCL3924	RPA2409	possible AmiR antitermination protein	red
orthoMCL3438	RPA2410	putative urea/short-chain amide transport system substrate-binding protein	red
orthoMCL5648	RPA2544	conserved hypothetical protein	red
orthoMCL3409	RPA2705	hypothetical protein	red
orthoMCL3406	RPA2732	conserved hypothetical protein	red
orthoMCL1230	RPA2803	hypothetical protein	red
orthoMCL5611	RPA3009	light harvesting protein B-800-850, beta chain C (antenna pigment protein, beta chain C) (LH II-C beta)	red
orthoMCL5610	RPA3010	pseudo light-harvesting protein	red
orthoMCL4474	RPA3011	unknown protein	red
orthoMCL5609	RPA3013	light harvesting protein B-800-850, beta chain D (antenna pigment protein, beta chain D) (LH II-D beta)	red
orthoMCL2897	RPA3017	response regulator receiver (CheY-like protein)	red
orthoMCL5608	RPA3036	hypothetical protein	red
orthoMCL3372	RPA3196	hypothetical protein	red
orthoMCL3368	RPA3258	conserved hypothetical protein	red
orthoMCL5595	RPA3280	possible heme receptor	red
orthoMCL5594	RPA3282	RNA polymerase ECF-type sigma factor, possible FecI	red
orthoMCL2873	RPA3308	ycfI, putative structural proteins	red
orthoMCL0917	RPA3312	Transglutaminase-like domain	red
orthoMCL3869	RPA3320	hypothetical protein	red
orthoMCL0894	RPA3416	protein with 2 CBS domains	red
orthoMCL5575	RPA3486	putative branched-chain amino acid transport system substrate-binding protein	red
orthoMCL5571	RPA3597	conserved hypothetical protein	red
orthoMCL0712	RPA3726	conserved unknown protein	red
orthoMCL0587	RPA3943	conserved hypothetical protein	red
orthoMCL5535	RPA4097	transcriptional regulator, TetR family	red
orthoMCL4389	RPA4114	transcriptional regulator, LysR family	red
orthoMCL3804	RPA4163	putative taurine transport system ATP-binding protein	red
orthoMCL2791	RPA4217	conserved unknown protein	red
orthoMCL2769	RPA4467	putative sulfur oxidation protein soxY	red
orthoMCL0313	RPA4474	transcriptional regulator, ArsR family	red
orthoMCL4367	RPA4484	putative sensor (PAS) domain for methyl-accepting chemotaxis sensory transducer	red

orthoMCL0272	RPA4602	ferredoxin like protein, fixX	red
orthoMCL0265	RPA4609	putative nifU protein	red
orthoMCL0258	RPA4616	nitrogenase reductase-associated ferredoxin, nifN	red
orthoMCL0251	RPA4624	hypothetical protein	red
orthoMCL0246	RPA4630	nitrogen fixation protein nifB	red
orthoMCL5492	RPA4690	hypothetical protein	red
orthoMCL3775	RPA4714	hypothetical protein	red
orthoMCL2690	NA	NA	turquoise
orthoMCL2696	NA	NA	turquoise
orthoMCL2705	NA	NA	turquoise
orthoMCL3216	NA	NA	turquoise
orthoMCL3219	NA	NA	turquoise
orthoMCL3687	NA	NA	turquoise
orthoMCL3701	NA	NA	turquoise
orthoMCL3725	NA	NA	turquoise
orthoMCL3737	NA	NA	turquoise
orthoMCL4141	NA	NA	turquoise
orthoMCL4148	NA	NA	turquoise
orthoMCL4150	NA	NA	turquoise
orthoMCL4175	NA	NA	turquoise
orthoMCL4190	NA	NA	turquoise
orthoMCL4233	NA	NA	turquoise
orthoMCL4241	NA	NA	turquoise
orthoMCL4271	NA	NA	turquoise
orthoMCL4783	NA	NA	turquoise
orthoMCL4785	NA	NA	turquoise
orthoMCL4786	NA	NA	turquoise
orthoMCL4787	NA	NA	turquoise
orthoMCL4791	NA	NA	turquoise
orthoMCL4792	NA	NA	turquoise
orthoMCL4793	NA	NA	turquoise
orthoMCL4794	NA	NA	turquoise
orthoMCL4806	NA	NA	turquoise
orthoMCL4814	NA	NA	turquoise
orthoMCL4815	NA	NA	turquoise
orthoMCL4829	NA	NA	turquoise
orthoMCL4895	NA	NA	turquoise
orthoMCL4896	NA	NA	turquoise
orthoMCL4947	NA	NA	turquoise
orthoMCL5035	NA	NA	turquoise
orthoMCL5264	NA	NA	turquoise
orthoMCL5316	NA	NA	turquoise
orthoMCL5324	NA	NA	turquoise
orthoMCL5374	NA	NA	turquoise
orthoMCL5383	NA	NA	turquoise
orthoMCL5421	NA	NA	turquoise
orthoMCL5440	NA	NA	turquoise

orthoMCL5468	NA	NA	turquoise
orthoMCL5469	NA	NA	turquoise
orthoMCL5470	NA	NA	turquoise
orthoMCL5471	NA	NA	turquoise
orthoMCL5472	NA	NA	turquoise
orthoMCL5473	NA	NA	turquoise
orthoMCL5475	NA	NA	turquoise
orthoMCL3656	RPA0097	putative flagellar basal-body rod protein flgC	turquoise
orthoMCL2497	RPA0275	putative ammonium transporter AmtB	turquoise
orthoMCL2488	RPA0286	putative diguanylate cyclase (GGDEF) with HAMP domain	turquoise
orthoMCL2435	RPA0345	putative protoporphyrin IX magnesium chelatase bchO	turquoise
orthoMCL3638	RPA0347	putative isocitrate dehydrogenase kinase/phosphatase	turquoise
orthoMCL2412	RPA0390	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL) with PAS domains	turquoise
orthoMCL2383	RPA0429	catalase/peroxidase	turquoise
orthoMCL2310	RPA0532	beta-ketothiolase, acetoacetyl-CoA reductase	turquoise
orthoMCL4085	RPA0540	hypothetical protein	turquoise
orthoMCL2273	RPA0597	possible competence protein F (COMF)	turquoise
orthoMCL3608	RPA0684	hypothetical protein	turquoise
orthoMCL2217	RPA0703	conserved hypothetical protein	turquoise
orthoMCL3585	RPA0776	possible gamma-glutamyltranspeptidase precursor	turquoise
orthoMCL3584	RPA0778	conserved hypothetical protein	turquoise
orthoMCL2184	RPA0779	conserved hypothetical protein	turquoise
orthoMCL2183	RPA0780	possible acyl-CoA dehydrogenase	turquoise
orthoMCL3582	RPA0799	putative carbonic anhydrase	turquoise
orthoMCL2127	RPA0852	two-component transcriptional regulator, LuxR family	turquoise
orthoMCL2080	RPA0919	ABC transporter, ATP-binding protein	turquoise
orthoMCL3093	RPA0968	putative hydrogenase expression/formation protein hupG	turquoise
orthoMCL2052	RPA0998	conserved hypothetical protein	turquoise
orthoMCL2046	RPA1014	conserved hypothetical protein	turquoise
orthoMCL0098	RPA1015	transcriptional regulator, PadR family	turquoise
orthoMCL2042	RPA1023	hypothetical protein	turquoise
orthoMCL2026	RPA1042	conserved unknown protein	turquoise
orthoMCL1968	RPA1134	conserved hypothetical protein	turquoise
orthoMCL1951	RPA1154	conserved hypothetical protein	turquoise
orthoMCL3064	RPA1213	conserved unknown protein	turquoise
orthoMCL3061	RPA1216	putative ABC transporter, permease protein	turquoise
orthoMCL3555	RPA1240	biopolymer transport protein ExbD	turquoise
orthoMCL1888	RPA1241	possible tonB transport protein	turquoise
orthoMCL1880	RPA1263	putative II.1 protein	turquoise
orthoMCL3548	RPA1286	hypothetical protein	turquoise
orthoMCL1857	RPA1406	conserved hypothetical protein	turquoise
orthoMCL3040	RPA1407	conserved hypothetical protein	turquoise
orthoMCL3039	RPA1408	putative ABC transporter protein	turquoise
orthoMCL1856	RPA1409	possible taurine transport system permease protein	turquoise
orthoMCL3038	RPA1410	possible taurine transport system protein	turquoise
orthoMCL3036	RPA1424	possible selenocysteine lyase	turquoise

orthoMCL3035	RPA1425	serine acetyltransferase	turquoise
orthoMCL3031	RPA1429	putative coenzyme F390 synthetase	turquoise
orthoMCL1851	RPA1466	putative glutamyl-tRNA(Gln) amidotransferase subunit A	turquoise
orthoMCL1850	RPA1474	hypothetical protein	turquoise
orthoMCL0044	RPA1493	PucC, possible chlorophyll Major Facilitator Superfamily (MFS) exporter	turquoise
orthoMCL0091	RPA1495	unknown protein	turquoise
orthoMCL1735	RPA1639	possible oxalate/formate Major Facilitator Family (MFS) antiporter	turquoise
orthoMCL1732	RPA1647	putative esterase	turquoise
orthoMCL1725	RPA1659	conserved unknown protein	turquoise
orthoMCL1680	RPA1752	branched-chain amino acid transport system ATP-binding protein	turquoise
orthoMCL1671	RPA1772	putative phosphoenolpyruvate carboxylase	turquoise
orthoMCL3512	RPA1780	Phenylacetic acid degradation-related protein	turquoise
orthoMCL2984	RPA1825	conserved hypothetical protein	turquoise
orthoMCL1633	RPA1857	conserved hypothetical protein	turquoise
orthoMCL1623	RPA1878	putative 6-aminohexanoate-dimer hydrolase	turquoise
orthoMCL3492	RPA1908	hypothetical protein	turquoise
orthoMCL1591	RPA1930	two-component transcriptional regulator, winged helix family	turquoise
orthoMCL3978	RPA1951	possible FusE-MFP/HlyD family membrane fusion protein	turquoise
orthoMCL4566	RPA1987	hypothetical protein	turquoise
orthoMCL1515	RPA2063	putative NosF protein (an ABC transporter)	turquoise
orthoMCL3468	RPA2114	putative nitrate transport system ATP-binding protein	turquoise
orthoMCL3467	RPA2115	putative cyanate lyase	turquoise
orthoMCL1469	RPA2166	conserved hypothetical protein	turquoise
orthoMCL1438	RPA2275	putative branched-chain amino acid transport system ATP-binding protein	turquoise
orthoMCL4525	RPA2345	unknown protein	turquoise
orthoMCL2934	RPA2439	conserved hypothetical protein	turquoise
orthoMCL1402	RPA2463	putative cysteine desulfurase, nifS homolog	turquoise
orthoMCL0008	RPA2493	possible P-methylase	turquoise
orthoMCL1335	RPA2591	putative nitrogen regulation protein nifR3	turquoise
orthoMCL1334	RPA2592	nitrogen regulatory signal transduction histidine kinase NtrB	turquoise
orthoMCL1333	RPA2593	nitrogen assimilation regulatory protein ntrC Response regulator	turquoise
orthoMCL2925	RPA2651	Endoribonuclease, protein synthesis inhibitor	turquoise
orthoMCL1244	RPA2769	putative diguanylate cyclase (GGDEF)	turquoise
orthoMCL2915	RPA2805	nitrile hydratase alpha subunit	turquoise
orthoMCL2914	RPA2806	putative nitrile hydratase beta subunit	turquoise
orthoMCL2913	RPA2807	conserved hypothetical protein	turquoise
orthoMCL1196	RPA2856	Protein of unknown function, HesB/YadR/YfhF	turquoise
orthoMCL0020	RPA3015	Bacteriophytochrome (light-regulated signal transduction histidine kinase), PhyB1	turquoise
orthoMCL1077	RPA3053	cold shock DNA binding protein	turquoise
orthoMCL1023	RPA3133	response regulator receiver (CheY-like protein)	turquoise
orthoMCL2884	RPA3153	conserved hypothetical protein	turquoise
orthoMCL4462	RPA3212	unknown protein	turquoise
orthoMCL4461	RPA3214	hypothetical protein	turquoise
orthoMCL0973	RPA3218	Protein of unknown function UPF0047	turquoise
orthoMCL0902	RPA3393	conserved hypothetical protein	turquoise

orthoMCL0854	RPA3501	conserved unknown protein	turquoise
orthoMCL0842	RPA3514	conserved hypothetical protein	turquoise
orthoMCL0840	RPA3516	possible bcr efflux pump, Major Facilitator Superfamily (MFS)	turquoise
orthoMCL0776	RPA3629	conserved hypothetical protein	turquoise
orthoMCL0753	RPA3663	urease gamma subunit	turquoise
orthoMCL0752	RPA3664	possible urease accessory protein D	turquoise
orthoMCL0751	RPA3665	possible ATP-binding component of ABC transporter	turquoise
orthoMCL0750	RPA3666	possible ATP-binding component of ABC transporter	turquoise
orthoMCL0748	RPA3668	putative ABC transporter system permease component, possible fused protein	turquoise
orthoMCL0747	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor	turquoise
orthoMCL0741	RPA3676	putative type IV prepilin peptidase, cpaA	turquoise
orthoMCL0698	RPA3747	putative ethanolamine ammonia-lyase light chain	turquoise
orthoMCL0696	RPA3749	ethanolamine ammonia-lyase large subunit	turquoise
orthoMCL0694	RPA3751	unknown protein	turquoise
orthoMCL0691	RPA3754	Transglutaminase-like domain	turquoise
orthoMCL0059	RPA3789	putative ATP-binding component of a transport system	turquoise
orthoMCL2832	RPA3790	putative efflux protein	turquoise
orthoMCL0668	RPA3791	probable transcriptional regulator, TetR family	turquoise
orthoMCL0611	RPA3898	Flagellar basal body-associated protein FliL	turquoise
orthoMCL0600	RPA3912	hypothetical protein	turquoise
orthoMCL0581	RPA3957	Hpt domain	turquoise
orthoMCL0485	RPA4203	putative sensor (PAS) domain for methyl-accepting chemotaxis sensory transducer	turquoise
orthoMCL3279	RPA4210	hypothetical protein	turquoise
orthoMCL0481	RPA4212	conserved hypothetical protein	turquoise
orthoMCL2790	RPA4232	hypothetical protein	turquoise
orthoMCL0466	RPA4234	anaerobic aromatic degradation regulator aadR, Crp/Fnr family	turquoise
orthoMCL0465	RPA4235	putative cytochrome c, class I	turquoise
orthoMCL0458	RPA4242	conserved hypothetical protein	turquoise
orthoMCL0451	RPA4249	response regulator receiver (CheY-like protein)	turquoise
orthoMCL0450	RPA4250	nitrogen fixation transcriptional regulator fixK2, Crp/Fnr family	turquoise
orthoMCL0449	RPA4251	O-acetylhomoserine sulfhydrylase	turquoise
orthoMCL0411	RPA4328	elongation factor G, EF-G	turquoise
orthoMCL0345	RPA4411	glycerol-3-phosphate regulon repressor glpR, DeoR family	turquoise
orthoMCL3780	RPA4597	possible seryl-tRNA synthetase	turquoise
orthoMCL3779	RPA4598	possible acyl-CoA dehydrogenase	turquoise
orthoMCL3246	RPA4600	conserved unknown protein	turquoise
orthoMCL0244	RPA4632	Mo/Fe nitrogenase specific transcriptional regulator, NifA	turquoise
orthoMCL0238	RPA4647	probable transcriptional regulator KdgR, IclR family	turquoise
orthoMCL0214	RPA4691	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL0212	RPA4695	transcriptional regulator ligR, LysR family	turquoise
orthoMCL3239	RPA4712	TPR repeat	turquoise
orthoMCL0204	RPA4713	hypothetical protein	turquoise
orthoMCL0203	RPA4715	molybdate transport system ATP-binding protein	turquoise
orthoMCL0202	RPA4716	molybdate transport system permease protein	turquoise

orthoMCL0201	RPA4717	molybdate transport system substrate-binding protein	turquoise
orthoMCL0200	RPA4718	putative molybdate transport system transcriptional regulator, ModE	turquoise
orthoMCL0179	RPA4745	hypothetical protein	turquoise
orthoMCL0170	RPA4756	Helix-turn-helix protein, CopG family	turquoise
orthoMCL3237	RPA4757	possible outer membrane receptor for iron transport	turquoise
orthoMCL0145	RPA4785	Polysaccharide deacetylase	turquoise
orthoMCL0141	RPA4789	sensor histidine kinase	turquoise
orthoMCL0139	RPA4791	two-component transcriptional regulator, winged helix family	turquoise
orthoMCL0137	RPA4793	cytochrome bd-quinol oxidase subunit I	turquoise
orthoMCL3233	RPA4819	hypothetical protein	turquoise
orthoMCL0121	RPA4830	Metal dependent phosphohydrolase with a response regulator receiver domain	turquoise
orthoMCL2677	RPA0010	possible glutamate uptake transcriptional regulator, AsnC family	yellow
orthoMCL2674	RPA0016	cytochrome-c oxidase fixP chain	yellow
orthoMCL3178	RPA0133	ABC sulfate transport system, periplasmic binding protein	yellow
orthoMCL4107	RPA0134	hydrogenase gamma-fused hydrogenase large and small subunit	yellow
orthoMCL5801	RPA0239	hypothetical protein	yellow
orthoMCL2378	RPA0435	putative ribosome-binding factor A	yellow
orthoMCL3152	RPA0550	RNA polymerase ECF-type sigma factor	yellow
orthoMCL3597	RPA0744	putative high potential iron sulfur protein (HiPIP)	yellow
orthoMCL2181	RPA0782	conserved unknown protein	yellow
orthoMCL2128	RPA0851	possible MFS transporter	yellow
orthoMCL2049	RPA1001	conserved hypothetical protein	yellow
orthoMCL4062	RPA1006	possible protocatechuate 4,5-dioxygenase small subunit (AB035121)	yellow
orthoMCL4061	RPA1007	possible 2,3-dihydroxyphenylpropionate 1,2-dioxygenase	yellow
orthoMCL0049	RPA1017	Nitrogen fixation-related protein	yellow
orthoMCL2036	RPA1032	Glycosyl transferase, family 2	yellow
orthoMCL1999	RPA1091	hypothetical protein	yellow
orthoMCL3037	RPA1423	putative membrane protein	yellow
orthoMCL4028	RPA1441	possible uridylate kinase	yellow
orthoMCL4027	RPA1442	possible uridine monophosphate kinase	yellow
orthoMCL1707	RPA1692	putative transmembrane transport protein	yellow
orthoMCL1704	RPA1702	putative acyl-CoA ligase	yellow
orthoMCL1682	RPA1750	putative branched-chain amino acid transport system permease protein	yellow
orthoMCL1483	RPA2141	conserved hypothetical protein	yellow
orthoMCL1437	RPA2276	putative branched-chain amino acid transport system ATP-binding protein	yellow
orthoMCL1436	RPA2277	possible ABC transporter, permease protein	yellow
orthoMCL3933	RPA2384	putative ironIII transport permease protein	yellow
orthoMCL4522	RPA2386	conserved hypothetical protein	yellow
orthoMCL1101	RPA3012	light harvesting protein B-800-850, alpha chain D (antenna pigment protein, alpha chain D) (LH II-D alpha)	yellow
orthoMCL0959	RPA3235	50S ribosomal protein L6	yellow
orthoMCL0958	RPA3236	30S ribosomal protein S8	yellow
orthoMCL0957	RPA3237	30S ribosomal protein S14	yellow
orthoMCL0956	RPA3238	50S ribosomal protein L5	yellow
orthoMCL0955	RPA3239	50S ribosomal protein L24	yellow

orthoMCL0954	RPA3240	50S ribosomal protein L14	yellow
orthoMCL0953	RPA3241	30S ribosomal protein S17	yellow
orthoMCL0952	RPA3242	50S ribosomal protein L29	yellow
orthoMCL0951	RPA3243	50S ribosomal protein L16	yellow
orthoMCL0950	RPA3244	30S ribosomal protein S3	yellow
orthoMCL5598	RPA3276	possible tonB protein	yellow
orthoMCL0921	RPA3293	putative branched-chain amino acid transport system ATP-binding protein	yellow
orthoMCL0920	RPA3294	putative branched-chain amino acid transport system ATP-binding protein	yellow
orthoMCL0069	RPA3295	possible branched-chain amino acid ABC transporter, permease protein	yellow
orthoMCL0068	RPA3296	possible ABC transporter subunit (U75364)	yellow
orthoMCL0067	RPA3297	possible branched-chain amino acid transport system substrate-binding protein	yellow
orthoMCL0916	RPA3313	hypothetical protein	yellow
orthoMCL2857	RPA3470	putative sugar uptake ABC transporter periplasmic solute-binding protein precursor	yellow
orthoMCL3849	RPA3485	putative racemase	yellow
orthoMCL0836	RPA3521	putative UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase	yellow
orthoMCL4416	RPA3560	DUF81	yellow
orthoMCL2841	RPA3724	possible high-affinity leucine-isoleucine-valine transport system	yellow
orthoMCL0682	RPA3770	conserved unknown protein	yellow
orthoMCL2822	RPA3874	hypothetical protein	yellow
orthoMCL3297	RPA3994	putative diguanylate cyclase (GGDEF)	yellow
orthoMCL3813	RPA4007	hypothetical protein	yellow
orthoMCL2816	RPA4020	possible branched-chain amino acid transport system permease protein	yellow
orthoMCL2815	RPA4021	putative branched-chain amino acid ABC transport system permease protein livm (liv-I protein m)	yellow
orthoMCL2812	RPA4026	possible ABC transporter,ATP binding protein	yellow
orthoMCL0537	RPA4042	putative long-chain-fatty-acid--CoA ligase	yellow
orthoMCL0536	RPA4043	possible ABC transporter, permease protein	yellow
orthoMCL0535	RPA4044	putative branched-chain amino acid transport system permease protein	yellow
orthoMCL0534	RPA4045	possible branched-chain amino acid ABC transport system substrate-binding protein	yellow
orthoMCL4377	RPA4292	light harvesting protein B-800-850, alpha chain B (antenna pigment protein, alpha chain B) (LH II-B alpha)	yellow
orthoMCL2768	RPA4468	conserved unknown protein	yellow
orthoMCL2767	RPA4469	conserved unknown protein	yellow
orthoMCL2763	RPA4473	conserved hypothetical protein	yellow
orthoMCL0298	RPA4508	conserved hypothetical protein	yellow
orthoMCL0136	RPA4794	putative cytochrome bd-quinol oxidase subunit II	yellow

Module members are sorted by module colors and strain CGA009's gene numbering (RPA number).

Table 4.5. List of module members in the down-regulated NF-high/PM-high co-expression networks.

orthoMCL ID	RPA Number	Product Description	Module Color
orthoMCL3667	NA	NA	black
orthoMCL3691	NA	NA	black
orthoMCL2658	RPA0035	putative phenylalanine-tRNA ligase beta chain	black
orthoMCL2559	RPA0201	conserved hypothetical protein	black
orthoMCL2454	RPA0323	Protein of unknown function UPF0102	black
orthoMCL2387	RPA0424	transcriptional regulator, FUR family	black
orthoMCL2327	RPA0505	conserved hypothetical protein	black
orthoMCL2253	RPA0618	unknown protein	black
orthoMCL2181	RPA0782	conserved unknown protein	black
orthoMCL2098	RPA0892	glutamate synthase (NADPH) small chain	black
orthoMCL3576	RPA0895	putative 3-oxoacyl-[ACP] reductase	black
orthoMCL2012	RPA1065	Metal dependent phosphohydrolase	black
orthoMCL3071	RPA1110	putative 3-oxoacyl-acyl carrier protein reductase	black
orthoMCL5770	RPA1235	possible Leucine-Binding Protein (LBP)	black
orthoMCL3558	RPA1236	putative acyl-CoA dehydrogenase	black
orthoMCL4030	RPA1416	putative branched-chain amino acid transport system ATP-binding protein	black
orthoMCL4596	RPA1417	putative branched-chain amino acid transport system ATP-binding protein	black
orthoMCL0093	RPA1445	putative oligopeptide transport ATP-binding protein	black
orthoMCL1541	RPA2017	putative lipid A biosynthesis lauroyl acyltransferase	black
orthoMCL1536	RPA2022	specialized acyl carrier protein	black
orthoMCL1512	RPA2083	putative cobyrinic acid a,c-diamide synthase	black
orthoMCL1510	RPA2085	cobalamin biosynthesis protein G	black
orthoMCL1507	RPA2088	precorrin 3 methylase	black
orthoMCL3465	RPA2131	hypothetical protein	black
orthoMCL1472	RPA2158	hypothetical protein	black
orthoMCL4549	RPA2262	putative glutamate permease	black
orthoMCL2928	RPA2490	conserved hypothetical protein	black
orthoMCL1390	RPA2492	Conserved hypothetical protein	black
orthoMCL3914	RPA2545	possible outer membrane protein	black
orthoMCL1216	RPA2828	conserved hypothetical protein	black
orthoMCL1099	RPA3027	potassium uptake protein Kup	black
orthoMCL1025	RPA3129	50S ribosomal protein L33	black
orthoMCL3374	RPA3193	DUF35	black
orthoMCL3370	RPA3204	conserved hypothetical protein	black
orthoMCL0036	RPA3216	sensor histidine kinase with multiple PAS and a response regulator receiver domain	black
orthoMCL0932	RPA3272	50S ribosomal protein L1	black
orthoMCL5592	RPA3292	hypothetical protein	black
orthoMCL4415	RPA3587	hypothetical protein	black
orthoMCL0709	RPA3733	conserved unknown protein	black
orthoMCL2830	RPA3829	putative transcription elongation factor greA homologue	black

orthoMCL0561	RPA3980	nucleotide sugar epimerase	black
orthoMCL0539	RPA4038	possible ABC transport system ATP-binding protein	black
orthoMCL0502	RPA4176	ribosomal protein S21	black
orthoMCL0498	RPA4183	hypothetical protein	black
orthoMCL0340	RPA4418	conserved unknown protein	black
orthoMCL3263	RPA4476	conserved unknown protein	black
orthoMCL0304	RPA4497	putative lemA protein	black
orthoMCL4348	RPA4649	probable ABC transporter permease protein	black
orthoMCL4347	RPA4650	putative spermidine/putrescine transport system ATP-binding protein	black
orthoMCL0164	RPA4766	O-succinylhomoserine sulfhydrylase	black
orthoMCL3771	RPA4804	conserved hypothetical protein	black
orthoMCL2702	NA	NA	blue
orthoMCL3224	NA	NA	blue
orthoMCL4680	NA	NA	blue
orthoMCL4856	NA	NA	blue
orthoMCL2681	RPA0001	chromosomal replication initiator protein DnaA	blue
orthoMCL2656	RPA0038	ribosomal protein L20	blue
orthoMCL2650	RPA0044	bacitracin resistance protein	blue
orthoMCL2634	RPA0060	conserved unknown protein	blue
orthoMCL2632	RPA0062	conserved unknown protein	blue
orthoMCL2587	RPA0172	unknown protein	blue
orthoMCL2580	RPA0179	putative H+-transporting ATP synthase delta chain.	blue
orthoMCL2555	RPA0205	heme exporter protein C (ABC transporter permease component)	blue
orthoMCL2551	RPA0209	putative cell division protein FtsY	blue
orthoMCL2531	RPA0232	putative carbonic anhydrase	blue
orthoMCL2530	RPA0233	putative Citrate lyase beta chain (acyl lyase subunit) (citE)	blue
orthoMCL2527	RPA0240	3-isopropylmalate dehydratase	blue
orthoMCL2526	RPA0241	50s ribosomal protein L19	blue
orthoMCL2522	RPA0245	signal recognition particle protein	blue
orthoMCL2518	RPA0250	diaminopimelate epimerase	blue
orthoMCL2478	RPA0297	conserved hypothetical protein	blue
orthoMCL2467	RPA0309	imidazoleglycerol-phosphate dehydratase	blue
orthoMCL2449	RPA0329	ribonuclease PH	blue
orthoMCL4090	RPA0376	putative L-isoaspartyl protein carboxyl methyltransferase	blue
orthoMCL2407	RPA0396	Cfr family protein	blue
orthoMCL2406	RPA0397	conserved unknown protein	blue
orthoMCL2360	RPA0455	trytophanyl-tRNA synthetase	blue
orthoMCL2337	RPA0493	50S ribosomal protein L28	blue
orthoMCL2325	RPA0508	acetyl-CoA carboxylase carboxyltransferase alpha subunit	blue
orthoMCL3159	RPA0513	putative acetyl-CoA acetyltransferase	blue
orthoMCL2280	RPA0584	transcriptional accessory protein	blue
orthoMCL2267	RPA0604	putative aspartokinase, alpha and beta subunits	blue
orthoMCL2249	RPA0622	putative methionyl-tRNA formyltransferase	blue
orthoMCL2239	RPA0633	probable ribonuclease p protein component (protein c5)	blue
orthoMCL3126	RPA0663	putative transcriptional regulator, badM	blue
orthoMCL2211	RPA0715	putative cobalamin synthesis protein cobW	blue

orthoMCL2209	RPA0717	putative cob(I)alamin adenosyltransferase	blue
orthoMCL2148	RPA0830	conserved unknown protein	blue
orthoMCL2117	RPA0865	homospermidine synthase	blue
orthoMCL2116	RPA0867	Endoribonuclease L-PSP	blue
orthoMCL2084	RPA0915	putative NADPH quinone oxidoreductase	blue
orthoMCL2081	RPA0918	possible 50S ribosomal protein L31	blue
orthoMCL3571	RPA0996	Alpha/beta hydrolase fold	blue
orthoMCL2024	RPA1045	glycyl-tRNA synthetase alpha chain	blue
orthoMCL1993	RPA1097	DUF28	blue
orthoMCL1966	RPA1137	unknown protein	blue
orthoMCL1955	RPA1150	putative histidyl-tRNA synthetase	blue
orthoMCL3067	RPA1176	conserved hypothetical protein	blue
orthoMCL5773	RPA1232	putative branched-chain amino acid transport system ATP-binding protein	blue
orthoMCL1869	RPA1278	GatB/Yqey	blue
orthoMCL1838	RPA1505	putative porphobilinogen deaminase	blue
orthoMCL1774	RPA1589	30S ribosomal protein S4	blue
orthoMCL4585	RPA1651	possible leucine/isoleucine/valine-binding protein precursor	blue
orthoMCL1718	RPA1669	unknown protein	blue
orthoMCL1698	RPA1712	putative enoyl-CoA hydratase	blue
orthoMCL1622	RPA1879	Choloylglycine hydrolase	blue
orthoMCL1585	RPA1939	possible cation efflux protein	blue
orthoMCL1575	RPA1956	conserved hypothetical protein	blue
orthoMCL1553	RPA1999	GCN5-related N-acetyltransferase:Aminotransferase, class-II	blue
orthoMCL1505	RPA2090	precorrin isomerase CobH	blue
orthoMCL4552	RPA2110	probable transcriptional regulator, AraC family	blue
orthoMCL1418	RPA2436	putative biotin carboxyl carrier protein of acetyl-CoA carboxylase	blue
orthoMCL1359	RPA2557	Glycosyl transferase, family 2	blue
orthoMCL1358	RPA2558	conserved hypothetical protein	blue
orthoMCL1313	RPA2659	UDP-N-acetylglucosamine pyrophosphorylase	blue
orthoMCL1207	RPA2841	survival protein surE	blue
orthoMCL1183	RPA2874	enolase	blue
orthoMCL1176	RPA2888	triose-phosphate isomerase	blue
orthoMCL1149	RPA2922	30S ribosomal protein S2	blue
orthoMCL1117	RPA2962	putative trigger factor	blue
orthoMCL1076	RPA3056	nucleoside-diphosphate-kinase	blue
orthoMCL1075	RPA3057	ABC transporter, duplicated ATPase domains	blue
orthoMCL1058	RPA3075	putative malonyl CoA-acyl carrier protein transacylase	blue
orthoMCL1039	RPA3107	Glu-tRNA(Gln) amidotransferase subunit A	blue
orthoMCL0941	RPA3254	30S ribosomal protein S7	blue
orthoMCL0940	RPA3255	30S ribosomal protein S12	blue
orthoMCL0914	RPA3364	putative rRNA methylase	blue
orthoMCL3348	RPA3458	possible TrapT family, dctP subunit, C4-dicarboxylate periplasmic binding protein	blue
orthoMCL0702	RPA3743	possible phytoene synthase-related protein	blue
orthoMCL0671	RPA3784	putative dolichol-phosphate mannosyltransferase	blue
orthoMCL0642	RPA3833	tRNA/rRNA methyltransferase	blue

orthoMCL0630	RPA3871	Nuclear protein SET	blue
orthoMCL3299	RPA3981	conserved hypothetical protein	blue
orthoMCL0558	RPA3985	ADP-L-glycero-D-mannoheptose-6-epimerase	blue
orthoMCL0492	RPA4190	conserved unknown protein	blue
orthoMCL0420	RPA4308	putative phosphoglycerate dehydrogenase	blue
orthoMCL0411	RPA4328	elongation factor G, EF-G	blue
orthoMCL0395	RPA4354	putative GTP-binding protein	blue
orthoMCL0387	RPA4362	ribose-phosphate pyrophosphokinase	blue
orthoMCL0384	RPA4366	pyrroline-5-carboxylate reductase	blue
orthoMCL0365	RPA4388	GCN5-related N-acetyltransferase	blue
orthoMCL2780	RPA4404	putative periplasmic binding ABC transporter protein, probable sugar binding	blue
orthoMCL0350	RPA4406	sugar ABC transport system, permease component	blue
orthoMCL0347	RPA4409	ATP-binding protein of sugar ABC transporter	blue
orthoMCL2777	RPA4459	putative flavocytochrome C sulfide dehydrogenase, flavoprotein subunit	blue
orthoMCL2762	RPA4475	conserved hypothetical protein	blue
orthoMCL0302	RPA4501	phnA-like protein	blue
orthoMCL2759	RPA4506	putative short-chain alcohol dehydrogenase	blue
orthoMCL0216	RPA4688	possible inner mitochondrial membrane protein Sco1p	blue
orthoMCL0210	RPA4706	DedA family	blue
orthoMCL0190	RPA4730	ATP-binding protein of ABC transporter, duplicated ATPase domains	blue
orthoMCL2657	RPA0037	phenylalanyl-tRNA synthetase, alpha-subunit	brown
orthoMCL5802	RPA0237	conserved hypothetical protein	brown
orthoMCL2410	RPA0392	argininosuccinate synthase	brown
orthoMCL2378	RPA0435	putative ribosome-binding factor A	brown
orthoMCL2372	RPA0443	possible transcriptional regulator, XRE family	brown
orthoMCL4643	RPA0546	hypothetical protein	brown
orthoMCL3152	RPA0550	RNA polymerase ECF-type sigma factor	brown
orthoMCL0108	RPA0668	putative ABC transporter subunit, substrate-binding component	brown
orthoMCL4077	RPA0786	putative Adenylate/Guanylate cyclase	brown
orthoMCL2115	RPA0869	GCN5-related N-acetyltransferase	brown
orthoMCL2071	RPA0937	extragenic suppressor protein SuhB	brown
orthoMCL4062	RPA1006	possible protocatechuate 4,5-dioxygenase small subunit (AB035121)	brown
orthoMCL4060	RPA1009	possible cytochrome P450	brown
orthoMCL2029	RPA1039	possible isopentenyl monophosphate kinase	brown
orthoMCL3561	RPA1115	conserved hypothetical protein	brown
orthoMCL1895	RPA1224	putative indolepyruvate ferredoxin oxidoreductase, alpha subunit	brown
orthoMCL5772	RPA1233	putative branched-chain amino acid transport system ATP-binding protein	brown
orthoMCL1864	RPA1298	putative 3-oxoacyl-acyl carrier protein synthase III	brown
orthoMCL3520	RPA1582	Ferredoxin:Adenylate/Guanylate cyclase	brown
orthoMCL3022	RPA1620	unknown protein	brown
orthoMCL1728	RPA1653	conserved unknown protein	brown
orthoMCL3518	RPA1664	Glyoxalase/Bleomycin resistance protein/dioxygenase domain	brown
orthoMCL1704	RPA1702	putative acyl-CoA ligase	brown
orthoMCL1689	RPA1730	possible hydrolase	brown

orthoMCL0089	RPA1792	putative branched-chain amino acid transport system ATP-binding protein	brown
orthoMCL1642	RPA1839	putative dihydroneopterin aldolase	brown
orthoMCL1579	RPA1948	pyrroloquinoline quinone biosynthesis protein C	brown
orthoMCL1554	RPA1997	cysteine-tRNA ligase	brown
orthoMCL5704	RPA2159	hypothetical protein	brown
orthoMCL2950	RPA2184	putative oxidoreductase	brown
orthoMCL1438	RPA2275	putative branched-chain amino acid transport system ATP-binding protein	brown
orthoMCL1437	RPA2276	putative branched-chain amino acid transport system ATP-binding protein	brown
orthoMCL1436	RPA2277	possible ABC transporter, permease protein	brown
orthoMCL1435	RPA2278	possible ABC transporter, permease protein	brown
orthoMCL5677	RPA2350	putative O-acetylhomoserine sulfhydrylase	brown
orthoMCL5663	RPA2366	hypothetical protein	brown
orthoMCL1417	RPA2437	3-dehydroquinate dehydratase type 2	brown
orthoMCL4505	RPA2542	possible TrapT family, dctQ subunit, C4-dicarboxylate transport	brown
orthoMCL4504	RPA2543	TrapT family, dctP subunit, C4-dicarboxylate periplasmic binding protein	brown
orthoMCL1322	RPA2605	S-adenosylmethionine tRNA ribosyltransferase	brown
orthoMCL1299	RPA2675	Protein of unknown function UPF0004:Elongator protein 3/MiaB/NifB	brown
orthoMCL3905	RPA2676	transcriptional regulator, LysR family	brown
orthoMCL1245	RPA2768	ribosomal protein S9	brown
orthoMCL1177	RPA2887	possible secG: preprotein translocase	brown
orthoMCL1163	RPA2906	glutamyl-tRNA synthetase	brown
orthoMCL4475	RPA2999	putative nitrogen regulatory IIA protein	brown
orthoMCL1079	RPA3051	5'-phosphoribosyl-5-aminoimidazole synthetase	brown
orthoMCL3384	RPA3055	Integral membrane protein TerC family	brown
orthoMCL1056	RPA3078	30S ribosomal protein S18	brown
orthoMCL1055	RPA3079	hypothetical protein	brown
orthoMCL1053	RPA3081	unknown protein	brown
orthoMCL0949	RPA3245	50S ribosomal protein L22	brown
orthoMCL0926	RPA3286	possible phosphoglycerate mutase	brown
orthoMCL0069	RPA3295	possible branched-chain amino acid ABC transporter, permease protein	brown
orthoMCL0068	RPA3296	possible ABC transporter subunit (U75364)	brown
orthoMCL3363	RPA3302	conserved hypothetical protein	brown
orthoMCL0806	RPA3564	putative tyrosine phenol-lyase	brown
orthoMCL0703	RPA3742	putative poly-isoprenyl transferase	brown
orthoMCL2836	RPA3759	putative 5-carboxymethyl-2-hydroxymuconate isomerase	brown
orthoMCL0687	RPA3765	putative phenylacetic acid degradation protein PaaD	brown
orthoMCL2816	RPA4020	possible branched-chain amino acid transport system permease protein	brown
orthoMCL2815	RPA4021	putative branched-chain amino acid ABC transport system permease protein livm (liv-I protein m)	brown
orthoMCL2814	RPA4022	putative branched-chain amino acid ABC transport system, ATP-binding protein	brown
orthoMCL0543	RPA4034	ABC transporter, periplasmic branched chain amino acid binding protein	brown
orthoMCL0537	RPA4042	putative long-chain-fatty-acid--CoA ligase	brown

orthoMCL0536	RPA4043	possible ABC transporter, permease protein	brown
orthoMCL0535	RPA4044	putative branched-chain amino acid transport system permease protein	brown
orthoMCL0534	RPA4045	possible branched-chain amino acid ABC transport system substrate-binding protein	brown
orthoMCL0521	RPA4137	conserved unknown protein	brown
orthoMCL2794	RPA4198	Amidohydrolase 2	brown
orthoMCL0445	RPA4255	NADH-ubiquinone dehydrogenase chain K	brown
orthoMCL0431	RPA4269	ribonuclease H	brown
orthoMCL0430	RPA4270	homoserine kinase	brown
orthoMCL4376	RPA4296	unknown protein	brown
orthoMCL0419	RPA4309	phosphoserine aminotransferase	brown
orthoMCL3270	RPA4321	putative enoyl-CoA hydratase paaG	brown
orthoMCL2771	RPA4465	sulfur/thiosulfate oxidation protein SoxB	brown
orthoMCL2763	RPA4473	conserved hypothetical protein	brown
orthoMCL0300	RPA4504	acetyl-CoA synthetase	brown
orthoMCL4358	RPA4554	TrapT family, dctM subunit, C4-dicarboxylate transport	brown
orthoMCL0230	RPA4667	putative carbon-monoxide dehydrogenase large subunit	brown
orthoMCL0229	RPA4668	carbon monoxide dehydrogenase chain C	brown
orthoMCL0029	RPA4698	possible acyl transferase	brown
orthoMCL0167	RPA4760	unknown protein	brown
orthoMCL5491	RPA4795	conserved hypothetical protein	brown
orthoMCL0124	RPA4821	putative 5'-methylthioadenosine phosphorylase	brown
orthoMCL2664	RPA0028	bifunctional purine biosynthesis protein	green
orthoMCL2581	RPA0178	putative H+-transporting ATP synthase alpha chain.	green
orthoMCL2434	RPA0346	conserved hypothetical protein	green
orthoMCL2384	RPA0427	enoyl-acyl carrier protein reductase	green
orthoMCL2314	RPA0527	conserved hypothetical protein	green
orthoMCL2299	RPA0557	cysteine synthase, cytosolic O-acetylserine(thiol)lyase	green
orthoMCL2069	RPA0939	possible thiamine-phosphate pyrophosphorylase	green
orthoMCL2042	RPA1023	hypothetical protein	green
orthoMCL1997	RPA1093	possible GTP cyclohydrolase II, riboflavin biosynthesis	green
orthoMCL1956	RPA1149	ATP phosphoribosyltransferase	green
orthoMCL3068	RPA1163	possible epoxide hydrolase	green
orthoMCL3037	RPA1423	putative membrane protein	green
orthoMCL3013	RPA1707	putative feruloyl-CoA synthetase	green
orthoMCL1582	RPA1945	putative acyl-CoA transferase	green
orthoMCL1532	RPA2029	putative phosphoserine phosphatase	green
orthoMCL1514	RPA2081	possible Fe ABC Transporter	green
orthoMCL1511	RPA2084	precorrin 3 or 4 methylase	green
orthoMCL2951	RPA2173	GCN5-related N-acetyltransferase	green
orthoMCL1445	RPA2201	quinone oxidoreductase	green
orthoMCL1415	RPA2445	conserved hypothetical protein	green
orthoMCL1325	RPA2602	peptidyl prolyl cis-trans isomerase	green
orthoMCL1305	RPA2667	conserved unknown protein	green
orthoMCL1246	RPA2767	ribosomal protein L13	green
orthoMCL1206	RPA2845	seryl-tRNA synthetase	green

orthoMCL1181	RPA2879	2-dehydro-3-deoxyphosphooctonate aldolase	green
orthoMCL1151	RPA2920	uridylate kinase	green
orthoMCL1057	RPA3077	possible 30S ribosomal protein S6	green
orthoMCL0990	RPA3175	propionyl-CoA carboxylase precursor, biotin carrier protein	green
orthoMCL0931	RPA3273	50S ribosomal protein L11	green
orthoMCL0930	RPA3274	transcription antitermination protein, NusG	green
orthoMCL0925	RPA3287	putative 3-oxoacyl-acyl carrier protein reductase	green
orthoMCL3364	RPA3301	putative lipid transfer protein	green
orthoMCL0874	RPA3450	putative malonic semialdehyde oxidative decarboxylase	green
orthoMCL0847	RPA3509	permease, ABC-2-type transport system	green
orthoMCL2846	RPA3719	putative high-affinity branched-chain amino acid transport system ATP-binding protein	green
orthoMCL2845	RPA3720	putative branched-chain amino acid transport system ATP-binding protein	green
orthoMCL2844	RPA3721	possible ABC transporter, permease protein	green
orthoMCL0647	RPA3817	adenylosuccinate lyase	green
orthoMCL0550	RPA4015	S-adenosyl L-homocysteine hydrolase	green
orthoMCL0540	RPA4037	putative ABC transport system ATP-binding protein	green
orthoMCL0500	RPA4181	nicotinamide nucleotide transhydrogenase, subunit alpha2	green
orthoMCL0499	RPA4182	nicotinamide nucleotide transhydrogenase, subunit alpha1	green
orthoMCL0450	RPA4250	nitrogen fixation transcriptional regulator fixK2, Crp/Fnr family	green
orthoMCL0394	RPA4355	putative peptidyl-tRNA hydrolase	green
orthoMCL0393	RPA4356	putative 50S ribosomal protein L25	green
orthoMCL0377	RPA4373	possible protease	green
orthoMCL0349	RPA4407	permease protein of sugar ABC transporter	green
orthoMCL0338	RPA4420	conserved hypothetical protein	green
orthoMCL0326	RPA4439	putative histidinol-phosphate aminotransferase	green
orthoMCL2776	RPA4460	putative flavocytochrome C sulfide dehydrogenase, flavoprotein subunit	green
orthoMCL2772	RPA4464	sulfite dehydrogenase	green
orthoMCL2770	RPA4466	putative sulfur oxidation protein soxZ	green
orthoMCL2769	RPA4467	putative sulfur oxidation protein soxY	green
orthoMCL2768	RPA4468	conserved unknown protein	green
orthoMCL0198	RPA4721	possible+E2677 pyruvate-flavodoxin oxidoreductase	green
orthoMCL0177	RPA4749	electron transfer flavoprotein alpha-subunit, (ETFLS)	green
orthoMCL0176	RPA4750	electron transfer flavoprotein beta chain, (ETFSS)	green
orthoMCL2714	NA	NA	greenyellow
orthoMCL4206	NA	NA	greenyellow
orthoMCL4314	NA	NA	greenyellow
orthoMCL4751	NA	NA	greenyellow
orthoMCL4834	NA	NA	greenyellow
orthoMCL5116	NA	NA	greenyellow
orthoMCL5190	NA	NA	greenyellow
orthoMCL2213	RPA0713	conserved unknown protein	greenyellow
orthoMCL2212	RPA0714	bifunctional cobinamide kinase, cobinamide phosphate guanylyltransferase protein	greenyellow
orthoMCL2203	RPA0723	possible hemin ABC transporter, permease component	greenyellow
orthoMCL2064	RPA0947	conserved hypothetical protein	greenyellow

orthoMCL0097	RPA1016	ubiquinol-cytochrome-c reductase, Rieske iron-sulfur protein	greenyellow
orthoMCL1874	RPA1270	conserved hypothetical protein	greenyellow
orthoMCL0095	RPA1420	putative inner membrane component for iron transport	greenyellow
orthoMCL1855	RPA1421	possible efflux protein	greenyellow
orthoMCL1854	RPA1422	unknown protein	greenyellow
orthoMCL1819	RPA1527	photosynthetic reaction center L subunit	greenyellow
orthoMCL1818	RPA1528	photosynthetic reaction center M protein	greenyellow
orthoMCL1722	RPA1665	hypothetical protein	greenyellow
orthoMCL1688	RPA1735	pseudogene of fused ABC transporter ATPase and permease domains	greenyellow
orthoMCL1552	RPA2000	putative isopropyl malate synthase	greenyellow
orthoMCL1549	RPA2006	putative phosphatidylserine decarboxylase	greenyellow
orthoMCL3969	RPA2068	conserved unknown protein	greenyellow
orthoMCL1508	RPA2087	putative precorrin 6x reductase	greenyellow
orthoMCL1506	RPA2089	precorrin 2 methylase	greenyellow
orthoMCL1434	RPA2279	unknown protein	greenyellow
orthoMCL0038	RPA2443	probable antioxidant protein	greenyellow
orthoMCL1261	RPA2739	conserved unknown protein	greenyellow
orthoMCL0685	RPA3767	phenylacetic acid degradation protein paaB	greenyellow
orthoMCL0569	RPA3971	phosphomethylpyrimidine kinase (hmp-phosphate kinase)	greenyellow
orthoMCL0524	RPA4078	conserved hypothetical protein	greenyellow
orthoMCL3260	RPA4491	conserved hypothetical protein	greenyellow
orthoMCL0303	RPA4498	anthranilate synthase	greenyellow
orthoMCL5214	NA	NA	grey
orthoMCL2691	NA	NA	magenta
orthoMCL2706	NA	NA	magenta
orthoMCL2679	RPA0003	putative RecF protein	magenta
orthoMCL4124	RPA0027	2-dehydro-3-deoxyphosphoheptonate aldolase	magenta
orthoMCL2649	RPA0045	putative NADH dehydrogenase (ubiquinone) 1 alpha subcomplex	magenta
orthoMCL2595	RPA0164	gamma-glutamyl phosphate reductase	magenta
orthoMCL2558	RPA0202	aconitate hydratase	magenta
orthoMCL2535	RPA0227	beta-isopropylmalate dehydrogenase	magenta
orthoMCL3175	RPA0234	methyl-accepting chemotaxis receptor/sensory transducer	magenta
orthoMCL2441	RPA0337	orotidine 5`-phosphate decarboxylase	magenta
orthoMCL3639	RPA0340	phosphoglycerate mutase	magenta
orthoMCL3578	RPA0863	possible MgtC-magnesium transport	magenta
orthoMCL3577	RPA0866	putative nucleoside diphosphate kinase regulator	magenta
orthoMCL2067	RPA0943	phosphoglycerate kinase	magenta
orthoMCL0048	RPA1018	conserved hypothetical protein	magenta
orthoMCL2045	RPA1020	possible membrane fusion protein precursor	magenta
orthoMCL1926	RPA1181	Haloacid dehalogenase-like hydrolase	magenta
orthoMCL3065	RPA1200	possible phosphoglycerate mutase	magenta
orthoMCL4597	RPA1415	possible branched-chain amino acid transport system substrate-binding protein	magenta
orthoMCL4595	RPA1418	possible transport system permease protein	magenta
orthoMCL4594	RPA1419	possible transport system permease protein	magenta
orthoMCL0001	RPA1491	light harvesting protein B-800-850, beta chain E (antenna pigment protein, beta chain E) (LH II-E beta)	magenta

orthoMCL1768	RPA1597	phosphoribosylformylglycinamide synthetase	magenta
orthoMCL1700	RPA1708	putative acyl-CoA dehydrogenase	magenta
orthoMCL0018	RPA1709	putative acyl-CoA dehydrogenase	magenta
orthoMCL2990	RPA1789	putative branched-chain amino acid transport system substrate-binding protein	magenta
orthoMCL1446	RPA2200	inosine monophosphate dehydrogenase	magenta
orthoMCL1443	RPA2203	GMP synthetase	magenta
orthoMCL1432	RPA2281	putative low-affinity phosphate transport protein	magenta
orthoMCL1279	RPA2703	DNA topoisomerase IV subunitA	magenta
orthoMCL1150	RPA2921	elongation factor Ts	magenta
orthoMCL0980	RPA3192	MaoC-like dehydratase:Asparaginase/glutaminase	magenta
orthoMCL0942	RPA3253	elongation factor G	magenta
orthoMCL0716	RPA3715	acetyl-CoA acetyltransferase	magenta
orthoMCL0628	RPA3876	fumarate hydratase, class I	magenta
orthoMCL0556	RPA3988	putative phosphatase	magenta
orthoMCL2801	RPA4142	PilT protein, N-terminal	magenta
orthoMCL0428	RPA4272	conserved unknown protein	magenta
orthoMCL0422	RPA4303	conserved hypothetical protein	magenta
orthoMCL0414	RPA4320	L-lactate dehydrogenase	magenta
orthoMCL2727	RPA4798	putative acyl-CoA dehydrogenase	magenta
orthoMCL2726	RPA4799	possible acyl-CoA dehydrogenase	magenta
orthoMCL4158	NA	NA	pink
orthoMCL4204	NA	NA	pink
orthoMCL5010	NA	NA	pink
orthoMCL5050	NA	NA	pink
orthoMCL5226	NA	NA	pink
orthoMCL5378	NA	NA	pink
orthoMCL2640	RPA0054	putative small heat shock protein	pink
orthoMCL2599	RPA0160	possible acetyltransferases.	pink
orthoMCL2538	RPA0223	conserved hypothetical protein	pink
orthoMCL2504	RPA0267	possible thioredoxin	pink
orthoMCL2499	RPA0272	GlnK, nitrogen regulatory protein P-II	pink
orthoMCL4097	RPA0273	ammonium transporter AmtB	pink
orthoMCL2445	RPA0333	heat shock protein DnaK (70)	pink
orthoMCL2403	RPA0403	conserved hypothetical protein	pink
orthoMCL2068	RPA0940	fructose-bisphosphate aldolase	pink
orthoMCL2046	RPA1014	conserved hypothetical protein	pink
orthoMCL1964	RPA1139	conserved unknown protein	pink
orthoMCL0009	RPA1140	chaperonin GroEL1, cpn60	pink
orthoMCL3064	RPA1213	conserved unknown protein	pink
orthoMCL3555	RPA1240	biopolymer transport protein ExbD	pink
orthoMCL3036	RPA1424	possible selenocysteine lyase	pink
orthoMCL3030	RPA1430	putative outer membrane protein	pink
orthoMCL1748	RPA1624	conserved hypothetical protein	pink
orthoMCL2986	RPA1798	putative periplasmic binding protein for ABC transporter for branched chain amino acids	pink
orthoMCL1470	RPA2165	chaperonin GroES2, cpn10	pink

orthoMCL4525	RPA2345	unknown protein	pink
orthoMCL2930	RPA2488	conserved unknown protein	pink
orthoMCL1387	RPA2502	unknown protein	pink
orthoMCL3383	RPA3068	conserved hypothetical protein	pink
orthoMCL0973	RPA3218	Protein of unknown function UPF0047	pink
orthoMCL0848	RPA3508	ATP-binding component of ABC transporter	pink
orthoMCL3279	RPA4210	hypothetical protein	pink
orthoMCL0481	RPA4212	conserved hypothetical protein	pink
orthoMCL0466	RPA4234	anaerobic aromatic degradation regulator aadR, Crp/Fnr family	pink
orthoMCL0465	RPA4235	putative cytochrome c, class I	pink
orthoMCL0451	RPA4249	response regulator receiver (CheY-like protein)	pink
orthoMCL0449	RPA4251	O-acetylhomoserine sulfhydrylase	pink
orthoMCL2782	RPA4348	conserved hypothetical protein	pink
orthoMCL0336	RPA4423	conserved unknown protein	pink
orthoMCL0335	RPA4425	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	pink
orthoMCL0238	RPA4647	probable transcriptional regulator KdgR, IclR family	pink
orthoMCL0201	RPA4717	molybdate transport system substrate-binding protein	pink
orthoMCL0200	RPA4718	putative molybdate transport system transcriptional regulator, ModE	pink
orthoMCL0132	RPA4809	putative ABC transporter, ATP-binding protein	pink
orthoMCL3715	NA	NA	purple
orthoMCL5061	NA	NA	purple
orthoMCL5062	NA	NA	purple
orthoMCL5063	NA	NA	purple
orthoMCL5064	NA	NA	purple
orthoMCL3629	RPA0407	possible TonB-dependent receptor (outer membrane siderophore receptor)	purple
orthoMCL2341	RPA0489	ferredoxin II	purple
orthoMCL2288	RPA0571	two-component transcriptional regulator RegR, Fis family	purple
orthoMCL3132	RPA0657	benzoyl-CoA reductase subunit	purple
orthoMCL3131	RPA0658	benzoyl-CoA reductase subunit	purple
orthoMCL3130	RPA0659	benzoyl-CoA reductase subunit	purple
orthoMCL2215	RPA0711	Nitroreductase family	purple
orthoMCL3602	RPA0724	putative high-affinity nickel-transport protein	purple
orthoMCL2177	RPA0787	putative heat shock protein (htpX)	purple
orthoMCL2074	RPA0932	conserved unknown protein	purple
orthoMCL2066	RPA0944	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	purple
orthoMCL2054	RPA0958	putative acyl-CoA dehydrogenase	purple
orthoMCL1881	RPA1259	putative cation-transporting P-type ATPase	purple
orthoMCL1842	RPA1501	possible coenzyme F420 hydrogenase beta subunit	purple
orthoMCL3006	RPA1736	putative beta-glucosidase	purple
orthoMCL1669	RPA1776	possible lipid transfer protein	purple
orthoMCL3475	RPA2091	hypothetical protein	purple
orthoMCL2960	RPA2094	putative nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	purple
orthoMCL2959	RPA2095	possible cobalamin (5`-phosphate) synthase	purple
orthoMCL1501	RPA2097	putative cobF protein	purple
orthoMCL1454	RPA2185	nodN-like protein	purple

orthoMCL1453	RPA2186	possible 3-oxo-(acyl) acyl carrier protein reductase	purple
orthoMCL2929	RPA2489	hypothetical protein	purple
orthoMCL1274	RPA2716	conserved hypothetical protein	purple
orthoMCL0927	RPA3285	putative dehydrogenase	purple
orthoMCL2824	RPA3870	hypothetical protein	purple
orthoMCL0487	RPA4200	YbaK / prolyl-tRNA synthetases associated domain	purple
orthoMCL3278	RPA4288	conserved hypothetical protein	purple
orthoMCL2783	RPA4346	putative acetyl-CoA acyltransferase	purple
orthoMCL0209	RPA4707	conserved hypothetical protein	purple
orthoMCL0197	RPA4722	possible glutamate synthase, small subunit	purple
orthoMCL3236	RPA4803	putative outer membrane hemin/siderophore receptor protein	purple
orthoMCL0128	RPA4815	heat shock protein HtpG	purple
orthoMCL2701	NA	NA	red
orthoMCL2677	RPA0010	possible glutamate uptake transcriptional regulator, AsnC family	red
orthoMCL3633	RPA0373	thioredoxin	red
orthoMCL2302	RPA0545	conserved hypothetical protein	red
orthoMCL2190	RPA0751	putative phosphoadenosine phosphosulfate reductase	red
orthoMCL2103	RPA0887	UTP-glucose-1-phosphate uridylyltransferase	red
orthoMCL1893	RPA1226	putative 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit	red
orthoMCL1892	RPA1227	putative 2-oxoglutarate ferredoxin oxidoreductase, beta subunit	red
orthoMCL1891	RPA1228	putative 2-oxoglutarate ferredoxin oxidoreductase, gamma subunit	red
orthoMCL3557	RPA1237	possible acyl-CoA dehydrogenase	red
orthoMCL3553	RPA1260	Universal stress protein (Usp)	red
orthoMCL1733	RPA1646	hypothetical protein	red
orthoMCL2989	RPA1791	branched-chain amino acid transport system ATP-binding protein	red
orthoMCL1643	RPA1838	putative 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	red
orthoMCL1624	RPA1868	conserved hypothetical protein	red
orthoMCL1540	RPA2018	alcohol dehydrogenase	red
orthoMCL1537	RPA2021	3-hydroxymyristoyl-acyl carrier protein dehydratase	red
orthoMCL1530	RPA2031	acetolactate synthase (large subunit)	red
orthoMCL2948	RPA2193	putative ABC transporter, perplasmic binding protein, branched chain amino acids	red
orthoMCL2937	RPA2404	probable transcriptional regulator, AraC family	red
orthoMCL1263	RPA2737	unknown protein	red
orthoMCL3391	RPA2955	possible RND efflux membrane fusion protein precursor	red
orthoMCL2889	RPA3100	Uncharacterized protein family UPF0065	red
orthoMCL0948	RPA3246	30S ribosomal protein S19	red
orthoMCL0947	RPA3247	50S ribosomal protein L2	red
orthoMCL0946	RPA3248	50S ribosomal protein L23	red
orthoMCL0945	RPA3249	50S ribosomal protein L4	red
orthoMCL0944	RPA3250	50S ribosomal protein L3	red
orthoMCL0943	RPA3251	30S ribosomal protein S10	red
orthoMCL3362	RPA3303	MaoC-like dehydratase	red
orthoMCL0813	RPA3549	possible hydrolase	red
orthoMCL3836	RPA3728	conserved unknown protein	red
orthoMCL0686	RPA3766	phenylacetic acid degradation protein paaC	red

orthoMCL0542	RPA4035	possible ABC transport system permease protein	red
orthoMCL0541	RPA4036	possible branched-chain amino acid ABC transporter, permease protein	red
orthoMCL0533	RPA4046	putative branched-chain amino acid ABC transport system ATP-binding protein	red
orthoMCL0506	RPA4171	conserved unknown protein	red
orthoMCL0478	RPA4215	putative siroheme synthase	red
orthoMCL3274	RPA4302	methyl-accepting chemotaxis receptor/sensory transducer	red
orthoMCL0356	RPA4398	putative branched-chain amino acid ABC transporter, ATP-binding protein	red
orthoMCL0355	RPA4399	possible branched-chain amino acid transport system ATP-binding protein	red
orthoMCL0032	RPA4422	transcriptional regulator, Crp/Fnr family	red
orthoMCL2775	RPA4461	possible cytochrome subunit of sulfide dehydrogenase	red
orthoMCL2773	RPA4463	possible cytochrome	red
orthoMCL2767	RPA4469	conserved unknown protein	red
orthoMCL2766	RPA4470	DUF336	red
orthoMCL2765	RPA4471	conserved hypothetical protein	red
orthoMCL2764	RPA4472	putative c-type cytochrome biogenesis protein	red
orthoMCL0239	RPA4645	fructose-1,6-bisphosphatase	red
orthoMCL3773	RPA4761	conserved hypothetical protein	red
orthoMCL0056	RPA4818	conserved hypothetical protein	red
orthoMCL2684	NA	NA	turquoise
orthoMCL2697	NA	NA	turquoise
orthoMCL3215	NA	NA	turquoise
orthoMCL4157	NA	NA	turquoise
orthoMCL4188	NA	NA	turquoise
orthoMCL4848	NA	NA	turquoise
orthoMCL0052	RPA0139	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL3644	RPA0141	chemotaxis signal transduction/oligomerization protein CheW1-1	turquoise
orthoMCL4105	RPA0145	Metal dependent phosphohydrolase with a response regulator receiver domain	turquoise
orthoMCL2579	RPA0180	Rare lipoprotein A	turquoise
orthoMCL2524	RPA0243	putative 16S rRNA processing protein.	turquoise
orthoMCL2523	RPA0244	ribosomal protein S16	turquoise
orthoMCL2519	RPA0249	hypothetical protein	turquoise
orthoMCL2298	RPA0558	putative phosphoribosylaminoimidazole-succinocarboxamide synthase	turquoise
orthoMCL2275	RPA0595	conserved hypothetical protein	turquoise
orthoMCL2250	RPA0621	putative N-formylmethionylaminoacyl-tRNA deformylase	turquoise
orthoMCL3608	RPA0684	hypothetical protein	turquoise
orthoMCL3604	RPA0707	putative periplasmic divalent cation resistance protein CutA	turquoise
orthoMCL5790	RPA0710	hypothetical protein	turquoise
orthoMCL2194	RPA0747	putative sulfate ABC transporter, ATP-binding component	turquoise
orthoMCL2193	RPA0748	possible sulfate ABC transporter, permease component	turquoise
orthoMCL2192	RPA0749	putative sulfate ABC transporter, permease component	turquoise
orthoMCL2191	RPA0750	sulfate ABC transporter, periplasmic binding protein component	turquoise
orthoMCL2114	RPA0870	putative ornithine decarboxylase	turquoise
orthoMCL2101	RPA0889	small heat shock protein	turquoise

orthoMCL3093	RPA0968	putative hydrogenase expression/formation protein hupG	turquoise
orthoMCL2053	RPA0994	unknown protein	turquoise
orthoMCL2048	RPA1002	putative molybdenum transport system protein	turquoise
orthoMCL3568	RPA1012	possible integral membrane protein	turquoise
orthoMCL1901	RPA1212	Alpha/beta hydrolase fold	turquoise
orthoMCL1896	RPA1223	hypothetical protein	turquoise
orthoMCL1890	RPA1229	probable aerobic phenylacetate-CoA ligase	turquoise
orthoMCL5752	RPA1362	putative sulfate ester transport system substrate-binding protein	turquoise
orthoMCL5751	RPA1363	sulfate ester transport system permease protein	turquoise
orthoMCL5750	RPA1364	sulfate ester transport system ATP-binding protein	turquoise
orthoMCL5749	RPA1365	putative sulfatase	turquoise
orthoMCL5748	RPA1366	putative sulfur oxidation protein	turquoise
orthoMCL5747	RPA1367	putative sulfur oxidation protein	turquoise
orthoMCL3035	RPA1425	serine acetyltransferase	turquoise
orthoMCL3034	RPA1426	ABC transporter, ATP-binding protein	turquoise
orthoMCL3033	RPA1427	putative ABC transporter, permease protein	turquoise
orthoMCL3032	RPA1428	possible lipoprotein	turquoise
orthoMCL3031	RPA1429	putative coenzyme F390 synthetase	turquoise
orthoMCL4017	RPA1461	conserved hypothetical protein	turquoise
orthoMCL1798	RPA1549	possible photosynthetic complex assembly protein	turquoise
orthoMCL4005	RPA1578	ferredoxin--NADP+ reductase	turquoise
orthoMCL1769	RPA1594	putative tryptophan synthase beta chain	turquoise
orthoMCL5735	RPA1601	possible transcriptional regulator, MarR family	turquoise
orthoMCL1749	RPA1623	conserved unknown protein	turquoise
orthoMCL1662	RPA1802	conserved hypothetical protein	turquoise
orthoMCL5723	RPA1830	hypothetical protein	turquoise
orthoMCL5722	RPA1831	conserved hypothetical protein	turquoise
orthoMCL2970	RPA2034	Permeases of the drug/metabolite transporter (DMT) superfamily	turquoise
orthoMCL3974	RPA2037	possible periplasmic binding protein	turquoise
orthoMCL1502	RPA2096	cobyric acid synthase	turquoise
orthoMCL3474	RPA2098	Glutamine amidotransferase class-I	turquoise
orthoMCL1490	RPA2132	hypothetical protein	turquoise
orthoMCL3955	RPA2268	Flavin reductase-like	turquoise
orthoMCL5688	RPA2316	enoyl-CoA hydratase	turquoise
orthoMCL5687	RPA2317	putative CoA transferase, small subunit B	turquoise
orthoMCL5686	RPA2318	possible glutaconate CoA-transferase, subunit A	turquoise
orthoMCL4535	RPA2319	Uncharacterized protein family UPF0065	turquoise
orthoMCL5685	RPA2320	possible TctA subunit of the Tripartite Tricarboxylate Transport(TTT) Family	turquoise
orthoMCL5684	RPA2321	possible TctB subunit of the Tripartite Tricarboxylate Transporter(TTT) Family	turquoise
orthoMCL4526	RPA2343	transcriptional regulator, GntR family	turquoise
orthoMCL5681	RPA2344	conserved unknown protein	turquoise
orthoMCL5680	RPA2346	hypothetical protein	turquoise
orthoMCL5679	RPA2347	possible vanadium nitrogenase associated protein N (U51863)	turquoise
orthoMCL3943	RPA2348	possible nitrogenase molybdenum-iron protein alpha chain (nitrogenase component I) (dinitrogenase)	turquoise
orthoMCL5678	RPA2349	putative homocysteine synthase	turquoise

orthoMCL5676	RPA2351	transcriptional regulator, AsnC family	turquoise
orthoMCL5675	RPA2352	conserved hypothetical protein	turquoise
orthoMCL5674	RPA2353	putative nitrogenase NifH subunit	turquoise
orthoMCL5673	RPA2354	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	turquoise
orthoMCL5672	RPA2355	possible nitrogenase NifB	turquoise
orthoMCL4524	RPA2356	putative cystathionine beta-synthase	turquoise
orthoMCL5671	RPA2357	cystathionine gamma-lyase	turquoise
orthoMCL5670	RPA2358	2OG-Fe(II) oxygenase superfamily	turquoise
orthoMCL5669	RPA2359	putative periplasmic protein	turquoise
orthoMCL5668	RPA2360	ABC transporter, ATP-binding protein	turquoise
orthoMCL5667	RPA2361	putative ABC transporter, permease protein	turquoise
orthoMCL5666	RPA2362	putative O-acetylhomoserine sulfhydrylase	turquoise
orthoMCL5665	RPA2363	possible nitrogenase molybdenum-iron protein alpha chain (nitrogenase component I) (dinitrogenase)	turquoise
orthoMCL5664	RPA2364	possible nitrogenase iron-molybdenum cofactor biosynthesis protein NifE homolog	turquoise
orthoMCL4523	RPA2365	putative L-allo-threonine aldolase	turquoise
orthoMCL5645	RPA2607	possible transcriptional regulator, XRE family, CUPIN domain	turquoise
orthoMCL4502	RPA2608	possible sulfonate binding protein	turquoise
orthoMCL3416	RPA2609	possible monooxygenase	turquoise
orthoMCL4501	RPA2610	aliphatic sulfonate transport ATP-binding protein, Subunit of ABC transporter	turquoise
orthoMCL3912	RPA2611	putative aliphatic sulfonate transport membrane component. Permease subunit of an ABC transporter	turquoise
orthoMCL4500	RPA2612	methanesulfonate sulfonatase MsuD (monooxygenase)	turquoise
orthoMCL5644	RPA2613	putative aliphatic sulfonate binding protein, subunit of ABC transporter	turquoise
orthoMCL4499	RPA2614	conserved hypothetical protein	turquoise
orthoMCL4498	RPA2615	putative nitrogenase iron protein (nitrogenase component II) (nitrogenase reductase)	turquoise
orthoMCL4496	RPA2617	possible vanadium nitrogenase associated protein vnfN (U51863)	turquoise
orthoMCL5643	RPA2618	putative sulfonate transport system substrate-binding protein	turquoise
orthoMCL5642	RPA2619	possible GMC-type oxidoreductase	turquoise
orthoMCL5641	RPA2620	unknown protein	turquoise
orthoMCL3415	RPA2621	conserved unknown protein	turquoise
orthoMCL4495	RPA2622	putative sulfonate transport system, ATP-binding protein. Subunit of ABC transporter	turquoise
orthoMCL4494	RPA2623	possible sulfonate transport system permease protein. Subunit of an ABC transporter	turquoise
orthoMCL3911	RPA2624	putative sulfonate transport system substrate-binding protein	turquoise
orthoMCL5640	RPA2625	putative cystathionine or methionine gamma-lyase	turquoise
orthoMCL4493	RPA2626	possible hydrolase	turquoise
orthoMCL5639	RPA2627	putative carboxylesterase	turquoise
orthoMCL5638	RPA2628	polar amino acid ABC transport substrate-binding protein, aapJ-2	turquoise
orthoMCL5637	RPA2629	polar amino acid ABC transpor permease protein, aapQ-2	turquoise
orthoMCL5636	RPA2630	polar amino acid ABC transport system protein, aapM-2	turquoise
orthoMCL5635	RPA2631	possible ABC transport substrate-binding periplasmic protein	turquoise
orthoMCL5634	RPA2632	ABC transport permease protein, possibly for nitrate	turquoise
orthoMCL3910	RPA2633	putative ATP-binding protein of ABC transporter, possibly for nitrate	turquoise

orthoMCL5633	RPA2634	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	turquoise
orthoMCL5632	RPA2635	putative nitrogenase reductase NifH subunit	turquoise
orthoMCL5631	RPA2637	possible nitrogenase iron-molybdenum protein subunit nifK homolog	turquoise
orthoMCL5630	RPA2639	probable L-2-amino-thiazoline-4-carboxylic acid hydrolase	turquoise
orthoMCL5627	RPA2735	putative RND multidrug efflux membrane fusion protein MexC precursor	turquoise
orthoMCL1220	RPA2823	conserved hypothetical protein	turquoise
orthoMCL1219	RPA2825	conserved unknown protein	turquoise
orthoMCL1101	RPA3012	light harvesting protein B-800-850, alpha chain D (antenna pigment protein, alpha chain D) (LH II-D alpha)	turquoise
orthoMCL1097	RPA3030	hypothetical protein	turquoise
orthoMCL1005	RPA3152	hypothetical protein	turquoise
orthoMCL2884	RPA3153	conserved hypothetical protein	turquoise
orthoMCL2873	RPA3308	ycfI, putative structural proteins	turquoise
orthoMCL0916	RPA3313	hypothetical protein	turquoise
orthoMCL4428	RPA3369	hypothetical protein	turquoise
orthoMCL0062	RPA3546	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL3331	RPA3585	hypothetical protein	turquoise
orthoMCL0594	RPA3935	hypothetical protein	turquoise
orthoMCL5550	RPA4004	conserved hypothetical protein	turquoise
orthoMCL0501	RPA4180	nicotinamide nucleotide transhydrogenase, subunit beta	turquoise
orthoMCL0480	RPA4213	sulfite reductase hemoprotein subunit	turquoise
orthoMCL0472	RPA4223	response regulator receiver (CheY-like protein) with unknown domain	turquoise
orthoMCL0471	RPA4224	unknown protein	turquoise
orthoMCL0448	RPA4252	NADH-ubiquinone dehydrogenase chain N	turquoise
orthoMCL3801	RPA4284	YceI like family	turquoise
orthoMCL4378	RPA4285	Nitroreductase family	turquoise
orthoMCL0427	RPA4287	conserved hypothetical protein	turquoise
orthoMCL3267	RPA4402	hypothetical protein	turquoise
orthoMCL2774	RPA4462	Lipocalin-related protein and Bos/Can/Equ allergen	turquoise
orthoMCL3795	RPA4477	Possible membrane protein with ATP/GTP-binding site motif A (P-loop)	turquoise
orthoMCL0237	RPA4653	biotin sulfoxide reductase	turquoise
orthoMCL5497	RPA4656	possible sugar kinase	turquoise
orthoMCL5495	RPA4658	zinc-binding dehydrogenases (related to alcohol dehydrogenase, NADPH quinone oxidoreductase)	turquoise
orthoMCL0235	RPA4660	putative alpha,alpha-trehalose-phosphate synthase (UDP-forming) (trehalose-6-phosphate synthase)	turquoise
orthoMCL2736	RPA4719	molybdo-pterin binding protein	turquoise
orthoMCL0199	RPA4720	conserved hypothetical protein	turquoise
orthoMCL0136	RPA4794	putative cytochrome bd-quinol oxidase subunit II	turquoise
orthoMCL0115	NA	NA	yellow
orthoMCL3194	NA	NA	yellow
orthoMCL3207	NA	NA	yellow
orthoMCL3221	NA	NA	yellow
orthoMCL3227	NA	NA	yellow
orthoMCL3762	NA	NA	yellow

orthoMCL4181	NA	NA	yellow
orthoMCL4663	NA	NA	yellow
orthoMCL4669	NA	NA	yellow
orthoMCL4672	NA	NA	yellow
orthoMCL4679	NA	NA	yellow
orthoMCL4703	NA	NA	yellow
orthoMCL4778	NA	NA	yellow
orthoMCL5230	NA	NA	yellow
orthoMCL5267	NA	NA	yellow
orthoMCL3637	RPA0348	possible hydrolase	yellow
orthoMCL3169	RPA0372	transcriptional regulators, AsnC family	yellow
orthoMCL2365	RPA0450	transcriptional regulator, FUR family	yellow
orthoMCL4088	RPA0465	possible transmembrane protein	yellow
orthoMCL2348	RPA0477	conserved hypothetical protein	yellow
orthoMCL3123	RPA0673	Hydroxybenzoate anaerobic degradation regulatory protein HbaR, Crp/Fnr family	yellow
orthoMCL3122	RPA0676	possible enoyl-CoA hydratase/isomerase family member	yellow
orthoMCL2188	RPA0753	putative CysN/CysC bifunctional enzyme, ATP-sulfurylase large subunit and adenylyl sulfate kinase	yellow
orthoMCL2182	RPA0781	putative cytochrome c552 precursor	yellow
orthoMCL2172	RPA0794	a-type carbonic anhydrase	yellow
orthoMCL2149	RPA0828	transcriptional regulator, MarR family	yellow
orthoMCL3111	RPA0829	organic hydroperoxide resistance protein	yellow
orthoMCL4628	RPA0925	hypothetical protein	yellow
orthoMCL3573	RPA0926	hypothetical protein	yellow
orthoMCL0010	RPA0945	transketolase	yellow
orthoMCL3092	RPA0969	hydrogenase expression/formation protein hupH	yellow
orthoMCL3091	RPA0970	putative rubredoxin hupI	yellow
orthoMCL3090	RPA0971	putative hydrogenase expression/formation protein hupJ	yellow
orthoMCL3089	RPA0972	putative hydrogenase expression/formation protein hupK	yellow
orthoMCL3088	RPA0973	hydrogenase formation/expression protein hypA	yellow
orthoMCL3087	RPA0974	hydrogenase expression/formation protein hypB	yellow
orthoMCL3079	RPA0991	possible transcriptional regulator, Crp/Fnr family	yellow
orthoMCL3076	RPA1004	hypothetical protein	yellow
orthoMCL0045	RPA1492	light harvesting protein B-800-850, alpha chain E (antenna pigment protein, alpha chain E) (LH II-E alpha)	yellow
orthoMCL1791	RPA1559	ribulose-bisphosphate carboxylase large chain	yellow
orthoMCL1790	RPA1560	ribulose-bisphosphate carboxylase small chain	yellow
orthoMCL1789	RPA1561	cbbX protein homolog	yellow
orthoMCL1767	RPA1598	Peptidylprolyl isomerase, FKBP-type:Acyltransferase 3 family	yellow
orthoMCL1703	RPA1703	putative acetyl-CoA acyltransferase	yellow
orthoMCL1702	RPA1704	probable transcriptional regulator, TetR family	yellow
orthoMCL1661	RPA1803	putative DNA polymerase III alpha chain	yellow
orthoMCL1660	RPA1809	hypothetical protein	yellow
orthoMCL1648	RPA1829	LAO/AO transport system kinase	yellow
orthoMCL3482	RPA2003	probable cytochrome c precursor	yellow
orthoMCL1529	RPA2032	acetolactate synthase (small subunit)	yellow
orthoMCL1521	RPA2046	2-isopropylmalate synthase	yellow

orthoMCL1519	RPA2048	possible TrapT family, dctQ subunit, C4-dicarboxylate transport	yellow
orthoMCL2968	RPA2060	regulatory protein NosR	yellow
orthoMCL2967	RPA2061	nitrous-oxide reductase precursor NosZ	yellow
orthoMCL3472	RPA2106	conserved unknown protein	yellow
orthoMCL3463	RPA2162	possible serine protease/outer membrane autotransporter	yellow
orthoMCL4542	RPA2294	probable transcriptional regulator, TetR family	yellow
orthoMCL3414	RPA2640	Isochorismatase hydrolase family	yellow
orthoMCL1320	RPA2644	putative ABC transporter permease protein	yellow
orthoMCL1273	RPA2717	conserved hypothetical protein	yellow
orthoMCL2905	RPA2895	possible small heat shock protein	yellow
orthoMCL1015	RPA3141	possible VirR protein	yellow
orthoMCL4468	RPA3172	transcriptional regulator, LysR family	yellow
orthoMCL0014	RPA3252	elongation factor Tu	yellow
orthoMCL4426	RPA3400	hypothetical protein	yellow
orthoMCL2867	RPA3401	hypothetical protein	yellow
orthoMCL3351	RPA3453	putative enoyl-CoA hydratase	yellow
orthoMCL0846	RPA3510	conserved unknown protein	yellow
orthoMCL0805	RPA3565	ErfK/YbiS/YcfS/YnhG	yellow
orthoMCL0758	RPA3653	Protein of unknown function UPF0033	yellow
orthoMCL0549	RPA4016	methionine S-adenosyltransferase	yellow
orthoMCL3285	RPA4144	conserved hypothetical protein	yellow
orthoMCL0461	RPA4239	conserved unknown protein	yellow
orthoMCL0456	RPA4244	conserved unknown protein	yellow
orthoMCL0455	RPA4245	conserved unknown protein	yellow
orthoMCL0396	RPA4353	conserved hypothetical protein	yellow
orthoMCL0352	RPA4403	morphinone reductase	yellow
orthoMCL0277	RPA4571	hypothetical protein	yellow
orthoMCL0242	RPA4641	ribulose-bisphosphate carboxylase form II	yellow
orthoMCL0241	RPA4642	fructose-bisphosphate aldolase	yellow
orthoMCL0240	RPA4644	phosphoribulokinase (phosphopentokinase) (PRK)	yellow
orthoMCL0138	RPA4792	RNA polymerase ECF-type sigma factor	yellow

Module members are sorted by module colors and strain CGA009's gene numbering (RPA number).

Table 4.6. List of module members in the up-regulated NF-low/NF-high co-expression networks.

orthoMCL ID	RPA number	Product Description	Module Color
orthoMCL3687	NA	NA	black
orthoMCL3710	NA	NA	black
orthoMCL3711	NA	NA	black
orthoMCL4150	NA	NA	black
orthoMCL4157	NA	NA	black
orthoMCL4158	NA	NA	black
orthoMCL4211	NA	NA	black
orthoMCL4233	NA	NA	black
orthoMCL4256	NA	NA	black
orthoMCL4279	NA	NA	black
orthoMCL4282	NA	NA	black
orthoMCL4771	NA	NA	black
orthoMCL4778	NA	NA	black
orthoMCL4787	NA	NA	black
orthoMCL4835	NA	NA	black
orthoMCL5050	NA	NA	black
orthoMCL5446	NA	NA	black
orthoMCL5451	NA	NA	black
orthoMCL5482	NA	NA	black
orthoMCL3629	RPA0407	possible TonB-dependent receptor (outer membrane siderophore receptor)	black
orthoMCL4088	RPA0465	possible transmembrane protein	black
orthoMCL2347	RPA0478	conserved hypothetical protein	black
orthoMCL3606	RPA0705	hypothetical protein	black
orthoMCL2211	RPA0715	putative cobalamin synthesis protein cobW	black
orthoMCL2206	RPA0720	putative ABC-type cobalamin/Fe3+-siderophores transport systems, periplasmic components	black
orthoMCL3602	RPA0724	putative high-affinity nickel-transport protein	black
orthoMCL2053	RPA0994	unknown protein	black
orthoMCL3555	RPA1240	biopolymer transport protein ExbD	black
orthoMCL1811	RPA1535	cytochrome c2	black
orthoMCL1794	RPA1553	conserved hypothetical protein	black
orthoMCL3022	RPA1620	unknown protein	black
orthoMCL1723	RPA1662	hypothetical protein	black
orthoMCL3005	RPA1742	putative tartrate dehydrogenase	black
orthoMCL1685	RPA1747	conserved hypothetical protein	black
orthoMCL3978	RPA1951	possible FusE-MFP/HlyD family membrane fusion protein	black
orthoMCL1558	RPA1985	probable diacylglycerol kinase	black
orthoMCL4566	RPA1987	hypothetical protein	black
orthoMCL1514	RPA2081	possible Fe ABC Transporter	black
orthoMCL1513	RPA2082	putative uroporphyrin III methylase	black
orthoMCL1512	RPA2083	putative cobyrinic acid a,c-diamide synthase	black
orthoMCL1511	RPA2084	precorrin 3 or 4 methylase	black

orthoMCL1510	RPA2085	cobalamin biosynthesis protein G	black
orthoMCL1508	RPA2087	putative precorrin 6x reductase	black
orthoMCL1507	RPA2088	precorrin 3 methylase	black
orthoMCL1506	RPA2089	precorrin 2 methylase	black
orthoMCL1505	RPA2090	precorrin isomerase CobH	black
orthoMCL3475	RPA2091	hypothetical protein	black
orthoMCL1502	RPA2096	cobyric acid synthase	black
orthoMCL1449	RPA2195	possible exopolyphosphatase	black
orthoMCL1406	RPA2458	putative cytochrome b561	black
orthoMCL1224	RPA2815	possible outer membrane protein	black
orthoMCL1181	RPA2879	2-dehydro-3-deoxyphosphooctonate aldolase	black
orthoMCL2897	RPA3017	response regulator receiver (CheY-like protein)	black
orthoMCL4463	RPA3210	possible prolyl oligopeptidase family Dienelactone hydrolase family	black
orthoMCL0941	RPA3254	30S ribosomal protein S7	black
orthoMCL0695	RPA3750	methyl-accepting chemotaxis receptor/sensory transducer	black
orthoMCL3829	RPA3860	hypothetical protein	black
orthoMCL0600	RPA3912	hypothetical protein	black
orthoMCL0500	RPA4181	nicotinamide nucleotide transhydrogenase, subunit alpha2	black
orthoMCL0303	RPA4498	anthranilate synthase	black
orthoMCL0238	RPA4647	probable transcriptional regulator KdgR, IclR family	black
orthoMCL3242	RPA4683	hypothetical protein	black
orthoMCL0169	RPA4758	conserved hypothetical protein	black
orthoMCL3236	RPA4803	putative outer membrane hemin/siderophore receptor protein	black
orthoMCL0121	RPA4830	Metal dependent phosphohydrolase with a response regulator receiver domain	black
orthoMCL2684	NA	NA	blue
orthoMCL2719	NA	NA	blue
orthoMCL2720	NA	NA	blue
orthoMCL3207	NA	NA	blue
orthoMCL3219	NA	NA	blue
orthoMCL3691	NA	NA	blue
orthoMCL4137	NA	NA	blue
orthoMCL4152	NA	NA	blue
orthoMCL4156	NA	NA	blue
orthoMCL4200	NA	NA	blue
orthoMCL4252	NA	NA	blue
orthoMCL4336	NA	NA	blue
orthoMCL4664	NA	NA	blue
orthoMCL4666	NA	NA	blue
orthoMCL4670	NA	NA	blue
orthoMCL4676	NA	NA	blue
orthoMCL4703	NA	NA	blue
orthoMCL4704	NA	NA	blue
orthoMCL4776	NA	NA	blue
orthoMCL4777	NA	NA	blue
orthoMCL5278	NA	NA	blue
orthoMCL5279	NA	NA	blue

orthoMCL5436	NA	NA	blue
orthoMCL4125	RPA0009	circadian clock protein	blue
orthoMCL4124	RPA0027	2-dehydro-3-deoxyphosphoheptonate aldolase	blue
orthoMCL2640	RPA0054	putative small heat shock protein	blue
orthoMCL3656	RPA0097	putative flagellar basal-body rod protein flgC	blue
orthoMCL2602	RPA0157	conserved unknown protein	blue
orthoMCL2599	RPA0160	possible acetyltransferases.	blue
orthoMCL2546	RPA0215	possible general stress protein 26	blue
orthoMCL3176	RPA0229	hypothetical protein	blue
orthoMCL2459	RPA0317	sensor histidine kinase with a response regulator receiver domain	blue
orthoMCL2435	RPA0345	putative protoporphyrin IX magnesium chelatase bchO	blue
orthoMCL3166	RPA0401	NAD binding site:Amine oxidase	blue
orthoMCL2304	RPA0543	unknown protein	blue
orthoMCL3608	RPA0684	hypothetical protein	blue
orthoMCL2182	RPA0781	putative cytochrome c552 precursor	blue
orthoMCL2181	RPA0782	conserved unknown protein	blue
orthoMCL3110	RPA0864	hypothetical protein	blue
orthoMCL3079	RPA0991	possible transcriptional regulator, Crp/Fnr family	blue
orthoMCL4060	RPA1009	possible cytochrome P450	blue
orthoMCL2046	RPA1014	conserved hypothetical protein	blue
orthoMCL2001	RPA1089	hypothetical protein	blue
orthoMCL1933	RPA1173	cold shock DNA binding protein	blue
orthoMCL1903	RPA1210	conserved hypothetical protein	blue
orthoMCL1902	RPA1211	hypothetical protein	blue
orthoMCL1901	RPA1212	Alpha/beta hydrolase fold	blue
orthoMCL1892	RPA1227	putative 2-oxoglutarate ferredoxin oxidoreductase, beta subunit	blue
orthoMCL1883	RPA1246	conserved unknown protein	blue
orthoMCL1871	RPA1274	possible Dps protein family starvation-inducible DNA-binding protein	blue
orthoMCL3048	RPA1275	conserved hypothetical protein	blue
orthoMCL4010	RPA1473	possible peptide transport system substrate-binding protein	blue
orthoMCL1849	RPA1481	response regulator receiver (CheY-like protein)	blue
orthoMCL0043	RPA1494	unknown protein	blue
orthoMCL1844	RPA1499	hypothetical protein	blue
orthoMCL1834	RPA1510	conserved unknown protein	blue
orthoMCL1831	RPA1513	CrtB phytoene synthase	blue
orthoMCL1821	RPA1525	light-harvesting complex 1 beta chain	blue
orthoMCL1800	RPA1547	photosynthetic complex (LH1) assembly protein LhaA, probable Major Facilitator Superfamily (MFS) transporter	blue
orthoMCL1767	RPA1598	Peptidylprolyl isomerase, FKBP-type:Acyltransferase 3 family	blue
orthoMCL1763	RPA1605	conserved hypothetical protein	blue
orthoMCL1737	RPA1637	hypothetical protein	blue
orthoMCL3517	RPA1675	methyl-accepting chemotaxis receptor/sensory transducer with PAS domain	blue
orthoMCL4580	RPA1727	hypothetical protein	blue
orthoMCL3006	RPA1736	putative beta-glucosidase	blue
orthoMCL2974	RPA1992	possible NtrR protein	blue
orthoMCL3967	RPA2070	hypothetical protein	blue

orthoMCL1420	RPA2434	sensor histidine kinase	blue
orthoMCL3422	RPA2519	hypothetical protein	blue
orthoMCL1274	RPA2716	conserved hypothetical protein	blue
orthoMCL1273	RPA2717	conserved hypothetical protein	blue
orthoMCL1244	RPA2769	putative diguanylate cyclase (GGDEF)	blue
orthoMCL1232	RPA2801	Collagen triple helix repeat	blue
orthoMCL2910	RPA2820	DUF433	blue
orthoMCL1219	RPA2825	conserved unknown protein	blue
orthoMCL3888	RPA2860	hypothetical protein	blue
orthoMCL2907	RPA2869	possible flavin-dependent oxidoreductase	blue
orthoMCL3398	RPA2876	periplasmic glucans biosynthesis protein OpgG	blue
orthoMCL2905	RPA2895	possible small heat shock protein	blue
orthoMCL2904	RPA2896	hypothetical protein	blue
orthoMCL3881	RPA2988	GCN5-related N-acetyltransferase	blue
orthoMCL1092	RPA3037	conserved unknown protein	blue
orthoMCL3873	RPA3170	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL) with PAS and GAF domains	blue
orthoMCL2873	RPA3308	ycfI, putative structural proteins	blue
orthoMCL0918	RPA3309	conserved unknown protein	blue
orthoMCL0007	RPA3311	glycosyl hydrolase	blue
orthoMCL0917	RPA3312	Transglutaminase-like domain	blue
orthoMCL0916	RPA3313	hypothetical protein	blue
orthoMCL4455	RPA3324	hypothetical protein	blue
orthoMCL4444	RPA3338	possible Condensation domain, peptide synthetase	blue
orthoMCL3358	RPA3339	possible McyE polykeitde synthase and peptide synthetase (AF183408)	blue
orthoMCL4442	RPA3342	possible glycosyl hydrolase	blue
orthoMCL3356	RPA3378	hypothetical protein	blue
orthoMCL3859	RPA3384	possible ABC related periplasmic binding protein	blue
orthoMCL2868	RPA3399	cold shock DNA binding protein	blue
orthoMCL0065	RPA3419	possible transcriptional regulator, Crp/Fnr family	blue
orthoMCL0888	RPA3424	hypothetical protein	blue
orthoMCL3851	RPA3460	hypothetical protein	blue
orthoMCL3850	RPA3484	PilT protein, N-terminal	blue
orthoMCL0846	RPA3510	conserved unknown protein	blue
orthoMCL3330	RPA3599	hypothetical protein	blue
orthoMCL0794	RPA3600	bacterioferritin	blue
orthoMCL0759	RPA3651	Ku domain	blue
orthoMCL2850	RPA3652	conserved hypothetical protein	blue
orthoMCL0720	RPA3702	methionine synthase	blue
orthoMCL2840	RPA3725	possible leucine/isoleucine/valine-binding protein precursor	blue
orthoMCL2836	RPA3759	putative 5-carboxymethyl-2-hydroxymuconate isomerase	blue
orthoMCL0628	RPA3876	fumarate hydratase, class I	blue
orthoMCL0567	RPA3973	cytochrome c556	blue
orthoMCL3297	RPA3994	putative diguanylate cyclase (GGDEF)	blue
orthoMCL0534	RPA4045	possible branched-chain amino acid ABC transport system substrate-binding protein	blue
orthoMCL0521	RPA4137	conserved unknown protein	blue

orthoMCL3285	RPA4144	conserved hypothetical protein	blue
orthoMCL0506	RPA4171	conserved unknown protein	blue
orthoMCL2798	RPA4178	conserved unknown protein	blue
orthoMCL3279	RPA4210	hypothetical protein	blue
orthoMCL2791	RPA4217	conserved unknown protein	blue
orthoMCL0472	RPA4223	response regulator receiver (CheY-like protein) with unknown domain	blue
orthoMCL0471	RPA4224	unknown protein	blue
orthoMCL0470	RPA4225	RNA polymerase ECF-type sigma factor	blue
orthoMCL0427	RPA4287	conserved hypothetical protein	blue
orthoMCL4377	RPA4292	light harvesting protein B-800-850, alpha chain B (antenna pigment protein, alpha chain B) (LH II-B alpha)	blue
orthoMCL4376	RPA4296	unknown protein	blue
orthoMCL0340	RPA4418	conserved unknown protein	blue
orthoMCL0032	RPA4422	transcriptional regulator, Crp/Fnr family	blue
orthoMCL0314	RPA4458	hypothetical protein	blue
orthoMCL2764	RPA4472	putative c-type cytochrome biogenesis protein	blue
orthoMCL3261	RPA4480	possible RND divalent metal cation efflux membrane fusion protein CzcB precursor	blue
orthoMCL4365	RPA4500	hypothetical protein	blue
orthoMCL3777	RPA4640	hypothetical protein	blue
orthoMCL0235	RPA4660	putative alpha,alpha-trehalose-phosphate synthase (UDP-forming) (trehalose-6-phosphate synthase)	blue
orthoMCL0229	RPA4668	carbon monoxide dehydrogenase chain C	blue
orthoMCL2705	NA	NA	brown
orthoMCL3204	NA	NA	brown
orthoMCL3664	NA	NA	brown
orthoMCL3672	NA	NA	brown
orthoMCL3688	NA	NA	brown
orthoMCL3690	NA	NA	brown
orthoMCL4243	NA	NA	brown
orthoMCL4314	NA	NA	brown
orthoMCL4866	NA	NA	brown
orthoMCL4873	NA	NA	brown
orthoMCL4895	NA	NA	brown
orthoMCL4935	NA	NA	brown
orthoMCL4936	NA	NA	brown
orthoMCL4968	NA	NA	brown
orthoMCL4999	NA	NA	brown
orthoMCL5000	NA	NA	brown
orthoMCL5033	NA	NA	brown
orthoMCL5047	NA	NA	brown
orthoMCL5049	NA	NA	brown
orthoMCL5081	NA	NA	brown
orthoMCL5090	NA	NA	brown
orthoMCL5091	NA	NA	brown
orthoMCL5104	NA	NA	brown
orthoMCL5116	NA	NA	brown

orthoMCL5134	NA	NA	brown
orthoMCL5137	NA	NA	brown
orthoMCL5138	NA	NA	brown
orthoMCL5155	NA	NA	brown
orthoMCL5156	NA	NA	brown
orthoMCL5164	NA	NA	brown
orthoMCL5169	NA	NA	brown
orthoMCL5211	NA	NA	brown
orthoMCL2585	RPA0174	putative dinucleoside polyphosphate hydrolase (AP4A pyrophosphatase) (invasion protein A, NUDIX family hydrolase, NUDH subfamily.	brown
orthoMCL2560	RPA0200	conserved hypothetical protein	brown
orthoMCL3635	RPA0362	putative class A beta-lactamase precursor (penicillinase)	brown
orthoMCL2416	RPA0386	Ku domain	brown
orthoMCL2279	RPA0586	putative PAN2 protein	brown
orthoMCL3147	RPA0639	RNA polymerase ECF-type sigma factor	brown
orthoMCL3123	RPA0673	Hydroxybenzoate anaerobic degradation regulatory protein HbaR, Crp/Fnr family	brown
orthoMCL3610	RPA0678	possible pyridine nucleotide-linked oxidoreductase, possible glutamate synthase	brown
orthoMCL2199	RPA0732	putative NAD-dependent formate dehydrogenase gamma subunit	brown
orthoMCL3085	RPA0976	putative hypC	brown
orthoMCL2000	RPA1090	transcriptional regulatorn, Crp/Fnr family	brown
orthoMCL1907	RPA1203	conserved hypothetical protein	brown
orthoMCL1900	RPA1219	conserved unknown protein	brown
orthoMCL1896	RPA1223	hypothetical protein	brown
orthoMCL0045	RPA1492	light harvesting protein B-800-850, alpha chain E (antenna pigment protein, alpha chain E) (LH II-E alpha)	brown
orthoMCL0044	RPA1493	PucC, possible chlorophyll Major Facilitator Superfamily (MFS) exporter	brown
orthoMCL1832	RPA1512	phytoene dehydrogenase CrtI	brown
orthoMCL1828	RPA1518	methoxyneurosporene dehydrogenase	brown
orthoMCL1827	RPA1519	geranylgeranyl pyrophosphate synthase	brown
orthoMCL1825	RPA1521	2-desacetyl-2-hydroxyethyl bacteriochlorophyllide a dehydrogenase	brown
orthoMCL1824	RPA1522	bacteriochlorophyllide reductase subunit BchX	brown
orthoMCL1822	RPA1524	bacteriochlorophyllide reductase subunit	brown
orthoMCL1820	RPA1526	light-harvesting complex 1 alpha chain	brown
orthoMCL1732	RPA1647	putative esterase	brown
orthoMCL1721	RPA1666	putative coproporphyrinogen oxidase III	brown
orthoMCL1720	RPA1667	putative 4-vinyl protochlorophyllide reductase	brown
orthoMCL1719	RPA1668	Mg-protoporphyrin IX monomethyl ester oxidative cyclase 66kD subunit	brown
orthoMCL1718	RPA1669	unknown protein	brown
orthoMCL1715	RPA1672	conserved hypothetical protein	brown
orthoMCL1698	RPA1712	putative enoyl-CoA hydratase	brown
orthoMCL1677	RPA1763	putative long-chain-fatty-acid CoA ligase	brown
orthoMCL1672	RPA1771	hypothetical protein	brown
orthoMCL1660	RPA1809	hypothetical protein	brown
orthoMCL1622	RPA1879	Choloylglycine hydrolase	brown
orthoMCL1498	RPA2120	putative hemin binding protein	brown

orthoMCL1497	RPA2121	conserved unknown protein	brown
orthoMCL1492	RPA2126	conserved unknown protein	brown
orthoMCL1483	RPA2141	conserved hypothetical protein	brown
orthoMCL1479	RPA2145	putative enoyl-CoA hydratase/isomerase	brown
orthoMCL1456	RPA2182	putative glutathione S-transferase	brown
orthoMCL2944	RPA2248	conserved hypothetical protein	brown
orthoMCL3453	RPA2323	4-aminobutyrate aminotransferase	brown
orthoMCL3449	RPA2327	putative ABC transporter oligopeptide-binding protein	brown
orthoMCL3432	RPA2417	putative 3-ketoacyl-CoA reductase	brown
orthoMCL2931	RPA2483	hypothetical protein	brown
orthoMCL5647	RPA2566	putative EA59 gene protein, phage lambda	brown
orthoMCL3906	RPA2668	unknown protein	brown
orthoMCL2915	RPA2805	nitrile hydratase alpha subunit	brown
orthoMCL2914	RPA2806	putative nitrile hydratase beta subunit	brown
orthoMCL1030	RPA3120	DNA processing chain A	brown
orthoMCL3869	RPA3320	hypothetical protein	brown
orthoMCL3868	RPA3336	possible lipopeptide antibiotics iturin a biosynthesis protein	brown
orthoMCL0854	RPA3501	conserved unknown protein	brown
orthoMCL0034	RPA3518	Excinuclease ABC, C subunit, N-terminal	brown
orthoMCL0782	RPA3622	hypothetical protein	brown
orthoMCL0684	RPA3768	phenylacetic acid degradation protein paaA	brown
orthoMCL2835	RPA3771	unknown protein	brown
orthoMCL3833	RPA3826	conserved hypothetical protein	brown
orthoMCL3831	RPA3840	putative hydroxamate-type ferrisiderophore receptor	brown
orthoMCL0627	RPA3878	conserved unknown protein	brown
orthoMCL0606	RPA3906	unknown protein	brown
orthoMCL0590	RPA3940	Universal stress protein (Usp)	brown
orthoMCL0522	RPA4082	possible phage-like integrase	brown
orthoMCL4389	RPA4114	transcriptional regulator, LysR family	brown
orthoMCL3809	RPA4122	Conjugal transfer protein TrbD	brown
orthoMCL2800	RPA4147	putative nucleoside phosphorylase	brown
orthoMCL0465	RPA4235	putative cytochrome c, class I	brown
orthoMCL0461	RPA4239	conserved unknown protein	brown
orthoMCL0458	RPA4242	conserved hypothetical protein	brown
orthoMCL0456	RPA4244	conserved unknown protein	brown
orthoMCL0454	RPA4246	CBS domain:Transport-associated domain	brown
orthoMCL0431	RPA4269	ribonuclease H	brown
orthoMCL0422	RPA4303	conserved hypothetical protein	brown
orthoMCL0285	RPA4534	hypothetical protein	brown
orthoMCL0273	RPA4601	conserved hypothetical protein	brown
orthoMCL0214	RPA4691	methyl-accepting chemotaxis receptor/sensory transducer	brown
orthoMCL0056	RPA4818	conserved hypothetical protein	brown
orthoMCL2675	RPA0014	conserved hypothetical protein	green
orthoMCL2606	RPA0090	hypothetical protein	green
orthoMCL0053	RPA0091	hypothetical protein	green
orthoMCL3658	RPA0093	transcriptional regulator, TetR family	green
orthoMCL4122	RPA0099	putative oligopeptide ABC transporter (ATP-binding protein)	green

orthoMCL4120	RPA0101	putative dipeptide transport system permease protein 1	green
orthoMCL4119	RPA0102	putative ABC transporter oligopeptide-binding protein	green
orthoMCL4118	RPA0103	Cobalamin synthesis protein/P47K	green
orthoMCL3178	RPA0133	ABC sulfate transport system, periplasmic binding protein	green
orthoMCL2579	RPA0180	Rare lipoprotein A	green
orthoMCL2563	RPA0196	putative ABC transporter, ATP-binding protein	green
orthoMCL2558	RPA0202	aconitate hydratase	green
orthoMCL4079	RPA0685	transcriptional regulator, IclR family	green
orthoMCL5789	RPA0728	conserved unknown protein	green
orthoMCL2196	RPA0735	putative FdsC protein, formate dehydrogenase chain D	green
orthoMCL2195	RPA0736	possible NAD-dependent formate dehydrogenase delta subunit	green
orthoMCL3591	RPA0759	putative oligopeptide ABC transporter, permease component	green
orthoMCL3589	RPA0761	possible oligopeptide ABC transporter, periplasmic binding protein component	green
orthoMCL2172	RPA0794	a-type carbonic anhydrase	green
orthoMCL3094	RPA0967	hydrogenase expression/formation protein hupF	green
orthoMCL4056	RPA1082	conserved hypothetical protein	green
orthoMCL4055	RPA1083	conserved hypothetical protein	green
orthoMCL4054	RPA1084	possible minor curlin subunit precursor (fimbrin sef17 minor subunit).	green
orthoMCL4053	RPA1085	hypothetical protein	green
orthoMCL4052	RPA1086	possible curli production assembly/transport component csgg precursor	green
orthoMCL4051	RPA1087	possible transglycosylase SLT domain	green
orthoMCL1915	RPA1193	cytochrome b/c1 precursor	green
orthoMCL5768	RPA1309	possible transposase	green
orthoMCL1799	RPA1548	H subunit of photosynthetic reaction center complex	green
orthoMCL1681	RPA1751	putative branched-chain amino acid transport system ATP-binding protein	green
orthoMCL1671	RPA1772	putative phosphoenolpyruvate carboxylase	green
orthoMCL3513	RPA1773	putative DMT superfamily multidrug-efflux transporter	green
orthoMCL1470	RPA2165	chaperonin GroES2, cpn10	green
orthoMCL3959	RPA2237	possible lytic transglycosylase	green
orthoMCL1434	RPA2279	unknown protein	green
orthoMCL4543	RPA2288	Streptomyces cyclase/dehydrase	green
orthoMCL3922	RPA2418	psuedogene of ABC transporter, ATPase subunit	green
orthoMCL3423	RPA2516	unknown protein	green
orthoMCL4507	RPA2530	hypothetical protein	green
orthoMCL5648	RPA2544	conserved hypothetical protein	green
orthoMCL5643	RPA2618	putative sulfonate transport system substrate-binding protein	green
orthoMCL3911	RPA2624	putative sulfonate transport system substrate-binding protein	green
orthoMCL3903	RPA2678	putative permease protein, subunit of ABC transporter	green
orthoMCL3406	RPA2732	conserved hypothetical protein	green
orthoMCL3884	RPA2883	hypothetical protein	green
orthoMCL1170	RPA2894	conserved hypothetical protein	green
orthoMCL1113	RPA2967	glutamine synthetase I	green
orthoMCL1107	RPA2977	ribonucleotide reductase	green
orthoMCL1096	RPA3031	possible Acetyltransferase (GNAT) family	green

orthoMCL2891	RPA3088	conserved hypothetical protein	green
orthoMCL3376	RPA3123	hypothetical protein	green
orthoMCL0070	RPA3195	related to Pyruvate ferredoxin/flavodoxin oxidoreductase	green
orthoMCL5600	RPA3213	hypothetical protein	green
orthoMCL5591	RPA3306	nitrite reductase, major outer membrane copper-containing protein	green
orthoMCL4452	RPA3329	conserved hypothetical protein	green
orthoMCL4450	RPA3331	hypothetical protein	green
orthoMCL4449	RPA3332	hypothetical protein	green
orthoMCL4443	RPA3341	putative acetyltransferase	green
orthoMCL5586	RPA3343	possible mannosyltransferase B	green
orthoMCL4439	RPA3346	possible mannosyltransferase	green
orthoMCL3867	RPA3348	possible polysaccharide export protein (AF040104)	green
orthoMCL3866	RPA3349	putative exopolysaccharide biosynthesis protein	green
orthoMCL3865	RPA3350	putative aminotransferase	green
orthoMCL4437	RPA3351	possible glycosyl transferase	green
orthoMCL3864	RPA3352	glycosyltransferase	green
orthoMCL4436	RPA3354	hypothetical protein	green
orthoMCL3862	RPA3357	putative glycosyltransferase	green
orthoMCL4417	RPA3552	putative short-chain dehydrogenase/reductase	green
orthoMCL0682	RPA3770	conserved unknown protein	green
orthoMCL2806	RPA4068	hypothetical protein	green
orthoMCL0464	RPA4236	Mce4/Rv3499c/MTV023.06c protein	green
orthoMCL0450	RPA4250	nitrogen fixation transcriptional regulator fixK2, Crp/Fnr family	green
orthoMCL0411	RPA4328	elongation factor G, EF-G	green
orthoMCL0410	RPA4329	conserved unknown protein	green
orthoMCL3252	RPA4578	Basic membrane lipoprotein	green
orthoMCL0250	RPA4625	NifZ domain	green
orthoMCL2707	NA	NA	magenta
orthoMCL2417	RPA0385	possible DNA repair protein RadC	magenta
orthoMCL2184	RPA0779	conserved hypothetical protein	magenta
orthoMCL3091	RPA0970	putative rubredoxin hupI	magenta
orthoMCL3076	RPA1004	hypothetical protein	magenta
orthoMCL1895	RPA1224	putative indolepyruvate ferredoxin oxidoreductase, alpha subunit	magenta
orthoMCL1894	RPA1225	possible pyruvate ferredoxin/flavodoxin oxidoreductas 4Fe-4S binding domain	magenta
orthoMCL1893	RPA1226	putative 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit	magenta
orthoMCL1891	RPA1228	putative 2-oxoglutarate ferredoxin oxidoreductase, gamma subunit	magenta
orthoMCL1890	RPA1229	probable aerobic phenylacetate-CoA ligase	magenta
orthoMCL1864	RPA1298	putative 3-oxoacyl-acyl carrier protein synthase III	magenta
orthoMCL1791	RPA1559	ribulose-bisphosphate carboxylase large chain	magenta
orthoMCL1790	RPA1560	ribulose-bisphosphate carboxylase small chain	magenta
orthoMCL1789	RPA1561	cbbX protein homolog	magenta
orthoMCL3460	RPA2218	possible ATP-dependent RNA helicase	magenta
orthoMCL3459	RPA2238	conserved hypothetical protein	magenta
orthoMCL1222	RPA2818	conserved hypothetical protein	magenta
orthoMCL0873	RPA3451	possible transcriptional regulator, TetR family	magenta
orthoMCL0806	RPA3564	putative tyrosine phenol-lyase	magenta

orthoMCL2814	RPA4022	putative branched-chain amino acid ABC transport system, ATP-binding protein	magenta
orthoMCL0455	RPA4245	conserved unknown protein	magenta
orthoMCL0419	RPA4309	phosphoserine aminotransferase	magenta
orthoMCL0300	RPA4504	acetyl-CoA synthetase	magenta
orthoMCL3253	RPA4574	hypothetical protein	magenta
orthoMCL0263	RPA4611	putative nitrogen fixation protein nifQ	magenta
orthoMCL0262	RPA4612	ferredoxin 2[4Fe-4S] III, fdxB	magenta
orthoMCL0261	RPA4613	DUF683	magenta
orthoMCL0257	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE	magenta
orthoMCL0256	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK	magenta
orthoMCL0255	RPA4619	nitrogenase molybdenum-iron protein alpha chain, nifD	magenta
orthoMCL0254	RPA4620	nitrogenase iron protein, nifH	magenta
orthoMCL0252	RPA4623	conserved hypothetical protein	magenta
orthoMCL0251	RPA4624	hypothetical protein	magenta
orthoMCL0198	RPA4721	possible+E2677 pyruvate-flavodoxin oxidoreductase	magenta
orthoMCL0197	RPA4722	possible glutamate synthase, small subunit	magenta
orthoMCL4107	RPA0134	hydrogenase gamma-fused hydrogenase large and small subunit	pink
orthoMCL4106	RPA0135	possible oxidoreductase	pink
orthoMCL2508	RPA0260	possible photosynthesis gene regulator, AppA/PpaA family	pink
orthoMCL2507	RPA0261	unknown protein	pink
orthoMCL5799	RPA0326	DUF24, predicted transcriptional regulator, related to MarR family	pink
orthoMCL2303	RPA0544	LrgA family holin protein	pink
orthoMCL2299	RPA0557	cysteine synthase, cytosolic O-acetylserine(thiol)lyase	pink
orthoMCL5788	RPA0729	conserved hypothetical protein	pink
orthoMCL5787	RPA0730	conserved hypothetical protein	pink
orthoMCL3584	RPA0778	conserved hypothetical protein	pink
orthoMCL3582	RPA0799	putative carbonic anhydrase	pink
orthoMCL2125	RPA0854	5-aminolevulinic acid synthase (ALAS)	pink
orthoMCL2104	RPA0886	putative diguanylate cyclase (GGDEF)	pink
orthoMCL0048	RPA1018	conserved hypothetical protein	pink
orthoMCL4057	RPA1081	hypothetical protein	pink
orthoMCL1951	RPA1154	conserved hypothetical protein	pink
orthoMCL3046	RPA1285	conserved hypothetical protein	pink
orthoMCL3548	RPA1286	hypothetical protein	pink
orthoMCL3530	RPA1475	hypothetical protein	pink
orthoMCL1686	RPA1746	transcriptional regulator, LysR family	pink
orthoMCL4571	RPA1877	conserved hypothetical protein	pink
orthoMCL1563	RPA1979	unknown protein	pink
orthoMCL5677	RPA2350	putative O-acetylhomoserine sulfhydrylase	pink
orthoMCL3421	RPA2547	hypothetical protein	pink
orthoMCL1275	RPA2715	possible transcriptional regulator, MarR family	pink
orthoMCL4488	RPA2747	hypothetical protein	pink
orthoMCL3389	RPA2979	conserved hypothetical protein	pink
orthoMCL5612	RPA3007	hypothetical protein	pink
orthoMCL0020	RPA3015	Bacteriophytochrome (light-regulated signal transduction histidine kinase), PhyB1	pink

orthoMCL1011	RPA3146	conserved hypothetical protein	pink
orthoMCL1008	RPA3149	conserved hypothetical protein	pink
orthoMCL0984	RPA3185	methyl-accepting chemotaxis receptor/sensory transducer	pink
orthoMCL4456	RPA3323	unknown protein	pink
orthoMCL5580	RPA3409	possible metal-dependent hydrolases	pink
orthoMCL0890	RPA3421	conserved hypothetical protein	pink
orthoMCL4415	RPA3587	hypothetical protein	pink
orthoMCL4412	RPA3598	methyl-accepting chemotaxis receptor/sensory transducer	pink
orthoMCL0694	RPA3751	unknown protein	pink
orthoMCL2828	RPA3857	possible esterase/lipase/outer membrane autotransporter	pink
orthoMCL5555	RPA3959	putative amidase	pink
orthoMCL5551	RPA4003	conserved hypothetical protein	pink
orthoMCL3806	RPA4145	dissimilatory nitrite reductase	pink
orthoMCL0499	RPA4182	nicotinamide nucleotide transhydrogenase, subunit alpha1	pink
orthoMCL3274	RPA4302	methyl-accepting chemotaxis receptor/sensory transducer	pink
orthoMCL0231	RPA4666	carbon-monoxide dehydrogenase small subunit	pink
orthoMCL3773	RPA4761	conserved hypothetical protein	pink
orthoMCL2730	RPA4764	Metallo-phosphoesterase:Tat pathway signal	pink
orthoMCL4117	RPA0104	Amidohydrolase	red
orthoMCL4090	RPA0376	putative L-isoaspartyl protein carboxyl methyltransferase	red
orthoMCL4089	RPA0378	putative alpha-D-galactoside galactohydrolase	red
orthoMCL2383	RPA0429	catalase/peroxidase	red
orthoMCL3613	RPA0579	hypothetical protein	red
orthoMCL2197	RPA0734	NAD-dependent formate dehydrogenase alpha subunit	red
orthoMCL3595	RPA0746	possible deca-heme c-type cytochrome.	red
orthoMCL3093	RPA0968	putative hydrogenase expression/formation protein hupG	red
orthoMCL3092	RPA0969	hydrogenase expression/formation protein hupH	red
orthoMCL3089	RPA0972	putative hydrogenase expression/formation protein hupK	red
orthoMCL3088	RPA0973	hydrogenase formation/expression protein hypA	red
orthoMCL3087	RPA0974	hydrogenase expression/formation protein hypB	red
orthoMCL4069	RPA0983	possible phthalate dioxygenase	red
orthoMCL2048	RPA1002	putative molybdenum transport system protein	red
orthoMCL2042	RPA1023	hypothetical protein	red
orthoMCL2027	RPA1041	conserved hypothetical protein	red
orthoMCL2026	RPA1042	conserved unknown protein	red
orthoMCL1897	RPA1222	conserved hypothetical protein	red
orthoMCL1889	RPA1239	putative biopolymer transport protein ExbB	red
orthoMCL1874	RPA1270	conserved hypothetical protein	red
orthoMCL1863	RPA1299	conserved hypothetical protein	red
orthoMCL1860	RPA1305	possible flagellar hook length determination protein	red
orthoMCL3546	RPA1344	hypothetical protein	red
orthoMCL4603	RPA1383	putative transcriptional regulator, ModE family	red
orthoMCL1854	RPA1422	unknown protein	red
orthoMCL1810	RPA1536	transcriptional regulator PpsR2	red
orthoMCL5733	RPA1652	possible flagellar basal-body rod protein FlgG	red
orthoMCL1670	RPA1774	OmpA/MotB domain, possible porin	red
orthoMCL4574	RPA1854	possible transcriptional regulator, TetR family	red

orthoMCL2979	RPA1889	hypothetical protein	red
orthoMCL1545	RPA2011	puative potassium uptake protein Kup	red
orthoMCL3463	RPA2162	possible serine protease/outer membrane autotransporter	red
orthoMCL1469	RPA2166	conserved hypothetical protein	red
orthoMCL5694	RPA2231	conjugal transfer protein trbD	red
orthoMCL3425	RPA2478	probable transcriptional regulator, AraC family	red
orthoMCL3915	RPA2484	conserved hypothetical protein	red
orthoMCL3418	RPA2578	possible alginate o-acetyltransferase AlgJ	red
orthoMCL4493	RPA2626	possible hydrolase	red
orthoMCL2917	RPA2786	hypothetical protein	red
orthoMCL4483	RPA2859	hypothetical protein	red
orthoMCL2896	RPA3018	response regulator receiver:histidine kinase	red
orthoMCL3371	RPA3201	formate/nitrate transporter	red
orthoMCL3871	RPA3219	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL)	red
orthoMCL2878	RPA3222	DUF24, predicted transcriptional regulator, related to MarR family	red
orthoMCL5589	RPA3317	hypothetical protein	red
orthoMCL4451	RPA3330	conserved hypothetical protein	red
orthoMCL4448	RPA3333	putative curli production assembly/transport component csgg precursor	red
orthoMCL0915	RPA3340	peptide synthetase (fragment)	red
orthoMCL4441	RPA3344	possible oxidoreductase	red
orthoMCL3863	RPA3353	putative serine/threonine protein phosphatase	red
orthoMCL4435	RPA3355	putative exopolysaccharide polymerization protein	red
orthoMCL4434	RPA3356	unknown protein	red
orthoMCL3338	RPA3526	conserved hypothetical protein	red
orthoMCL3841	RPA3621	putative vanillate O-demethylase oxidoreductase	red
orthoMCL0721	RPA3701	putative 5,10-methylenetetrahydrofolate reductase	red
orthoMCL3318	RPA3762	putative 2-oxo-3-ene-1,7-dioic acid hydratase	red
orthoMCL3312	RPA3863	transcriptional regulator, Crp/Fnr family	red
orthoMCL5552	RPA3998	hypothetical protein	red
orthoMCL3815	RPA3999	possible coenzyme PQQ synthesis protein E	red
orthoMCL3813	RPA4007	hypothetical protein	red
orthoMCL2809	RPA4029	possible branched-chain amino acid ABC transport system substrate-binding protein	red
orthoMCL0538	RPA4041	putative branched-chain amino acid ABC transport system ATP-binding protein	red
orthoMCL0304	RPA4497	putative lemA protein	red
orthoMCL0284	RPA4537	conserved hypothetical protein	red
orthoMCL5492	RPA4690	hypothetical protein	red
orthoMCL0199	RPA4720	conserved hypothetical protein	red
orthoMCL0115	NA	NA	turquoise
orthoMCL2688	NA	NA	turquoise
orthoMCL2701	NA	NA	turquoise
orthoMCL2714	NA	NA	turquoise
orthoMCL3194	NA	NA	turquoise
orthoMCL3212	NA	NA	turquoise
orthoMCL3215	NA	NA	turquoise
orthoMCL3218	NA	NA	turquoise

orthoMCL3224	NA	NA	turquoise
orthoMCL3701	NA	NA	turquoise
orthoMCL3713	NA	NA	turquoise
orthoMCL4205	NA	NA	turquoise
orthoMCL4206	NA	NA	turquoise
orthoMCL4210	NA	NA	turquoise
orthoMCL4231	NA	NA	turquoise
orthoMCL4248	NA	NA	turquoise
orthoMCL4270	NA	NA	turquoise
orthoMCL4751	NA	NA	turquoise
orthoMCL4847	NA	NA	turquoise
orthoMCL4848	NA	NA	turquoise
orthoMCL4904	NA	NA	turquoise
orthoMCL4930	NA	NA	turquoise
orthoMCL4949	NA	NA	turquoise
orthoMCL4972	NA	NA	turquoise
orthoMCL5023	NA	NA	turquoise
orthoMCL5034	NA	NA	turquoise
orthoMCL5053	NA	NA	turquoise
orthoMCL5095	NA	NA	turquoise
orthoMCL5125	NA	NA	turquoise
orthoMCL5126	NA	NA	turquoise
orthoMCL5139	NA	NA	turquoise
orthoMCL5146	NA	NA	turquoise
orthoMCL5174	NA	NA	turquoise
orthoMCL5415	NA	NA	turquoise
orthoMCL2677	RPA0010	possible glutamate uptake transcriptional regulator, AsnC family	turquoise
orthoMCL2671	RPA0019	cytochrome-c oxidase fixN chain, heme and copper binding subunit	turquoise
orthoMCL2660	RPA0033	conserved hypothetical protein	turquoise
orthoMCL0052	RPA0139	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL3645	RPA0140	chemotaxis signal transduction/oligomerization protein CheW1-2	turquoise
orthoMCL3643	RPA0142	multidomain chemotaxis histidine kinase CheA1 (Hpt, CheA, & CheW domains)	turquoise
orthoMCL3642	RPA0143	response regulator receiver, CheY1	turquoise
orthoMCL3641	RPA0144	Sulfate transporter/antisigma-factor antagonist domain	turquoise
orthoMCL2509	RPA0259	HTransmembrane sensor and HAMP domains	turquoise
orthoMCL2498	RPA0274	GlnK, nitrogen regulatory protein P-II	turquoise
orthoMCL2497	RPA0275	putative ammonium transporter AmtB	turquoise
orthoMCL2488	RPA0286	putative diguanylate cyclase (GGDEF) with HAMP domain	turquoise
orthoMCL2476	RPA0299	putative maf protein	turquoise
orthoMCL2456	RPA0321	two-component transcriptional regulator, LuxR family	turquoise
orthoMCL3637	RPA0348	possible hydrolase	turquoise
orthoMCL3633	RPA0373	thioredoxin	turquoise
orthoMCL2413	RPA0389	putative penicillin-insensitive murein endopeptidase A	turquoise
orthoMCL0006	RPA0538	putative Omp2b porin	turquoise
orthoMCL2204	RPA0722	putative iron(III) dicitrate ABC transporter, ATP-binding component FecE	turquoise

orthoMCL3597	RPA0744	putative high potential iron sulfur protein (HiPIP)	turquoise
orthoMCL3596	RPA0745	possible outer membrane protein precursor	turquoise
orthoMCL2193	RPA0748	possible sulfate ABC transporter, permease component	turquoise
orthoMCL2192	RPA0749	putative sulfate ABC transporter, permease component	turquoise
orthoMCL2191	RPA0750	sulfate ABC transporter, periplasmic binding protein component	turquoise
orthoMCL2190	RPA0751	putative phosphoadenosine phosphosulfate reductase	turquoise
orthoMCL2189	RPA0752	putative ATP sulfurylase small subunit	turquoise
orthoMCL2188	RPA0753	putative CysN/CysC bifunctional enzyme, ATP-sulfurylase large subunit and adenylyl sulfate kinase	turquoise
orthoMCL2149	RPA0828	transcriptional regulator, MarR family	turquoise
orthoMCL2127	RPA0852	two-component transcriptional regulator, LuxR family	turquoise
orthoMCL2101	RPA0889	small heat shock protein	turquoise
orthoMCL2056	RPA0956	hypothetical protein	turquoise
orthoMCL0099	RPA1010	Beta-lactamase-like:ATP/GTP-binding site motif A (P-loop)	turquoise
orthoMCL2041	RPA1024	putative oxidoreductase	turquoise
orthoMCL2040	RPA1025	possible Ectothiorhodospira Vacuolata Cytochrome	turquoise
orthoMCL2008	RPA1071	conserved hypothetical protein	turquoise
orthoMCL2006	RPA1073	possible ADP-RIBOSE PHOSPHOHYDROLASE	turquoise
orthoMCL1968	RPA1134	conserved hypothetical protein	turquoise
orthoMCL1885	RPA1244	conserved unknown protein	turquoise
orthoMCL1884	RPA1245	conserved hypothetical protein	turquoise
orthoMCL3050	RPA1256	putative formamidase regulatory protein FmdB	turquoise
orthoMCL3049	RPA1271	conserved hypothetical protein	turquoise
orthoMCL1861	RPA1304	possible flagellar basal-body rod modification protein FlgD	turquoise
orthoMCL3038	RPA1410	possible taurine transport system protein	turquoise
orthoMCL3037	RPA1423	putative membrane protein	turquoise
orthoMCL3036	RPA1424	possible selenocysteine lyase	turquoise
orthoMCL3035	RPA1425	serine acetyltransferase	turquoise
orthoMCL3034	RPA1426	ABC transporter, ATP-binding protein	turquoise
orthoMCL3033	RPA1427	putative ABC transporter, permease protein	turquoise
orthoMCL3032	RPA1428	possible lipoprotein	turquoise
orthoMCL3031	RPA1429	putative coenzyme F390 synthetase	turquoise
orthoMCL3030	RPA1430	putative outer membrane protein	turquoise
orthoMCL0023	RPA1489	two-component transcriptional regulator, LuxR family	turquoise
orthoMCL0001	RPA1491	light harvesting protein B-800-850, beta chain E (antenna pigment protein, beta chain E) (LH II-E beta)	turquoise
orthoMCL0091	RPA1495	unknown protein	turquoise
orthoMCL1819	RPA1527	photosynthetic reaction center L subunit	turquoise
orthoMCL1818	RPA1528	photosynthetic reaction center M protein	turquoise
orthoMCL1798	RPA1549	possible photosynthetic complex assembly protein	turquoise
orthoMCL1797	RPA1550	possible photosynthetic complex assembly protein	turquoise
orthoMCL1796	RPA1551	hypothetical protein	turquoise
orthoMCL1793	RPA1554	5-aminolevulinic acid synthase (ALAS)	turquoise
orthoMCL1785	RPA1573	LemA family	turquoise
orthoMCL0021	RPA1628	chemotaxis signal transduction/oligomerization protein CheW2	turquoise
orthoMCL1744	RPA1630	chemotaxis methylesterase, CheB2	turquoise
orthoMCL1728	RPA1653	conserved unknown protein	turquoise

orthoMCL3019	RPA1678	chemotaxis methyltransferase, CheR3	turquoise
orthoMCL3017	RPA1680	putative response regulator and cylclic diguanylate phosphodiesterase (EAL)	turquoise
orthoMCL1706	RPA1693	superoxide dismutase	turquoise
orthoMCL3014	RPA1697	Competence-damaged protein	turquoise
orthoMCL1694	RPA1717	hypothetical protein	turquoise
orthoMCL3514	RPA1743	hypothetical protein	turquoise
orthoMCL1684	RPA1748	putative branched-chain amino acid transport system substrate-binding protein	turquoise
orthoMCL1651	RPA1822	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL0087	RPA1964	hypothetical protein	turquoise
orthoMCL1549	RPA2006	putative phosphatidylserine decarboxylase	turquoise
orthoMCL1547	RPA2009	possible chemotaxis protein motA	turquoise
orthoMCL1530	RPA2031	acetolactate synthase (large subunit)	turquoise
orthoMCL3970	RPA2067	hypothetical protein	turquoise
orthoMCL3472	RPA2106	conserved unknown protein	turquoise
orthoMCL3461	RPA2171	unknown protein	turquoise
orthoMCL1461	RPA2177	DUF404	turquoise
orthoMCL1441	RPA2251	putative plasmid stabilization protein	turquoise
orthoMCL0084	RPA2256	transcriptional regulator, ArsR family, ArsR1	turquoise
orthoMCL3956	RPA2267	hypothetical protein	turquoise
orthoMCL3954	RPA2293	hypothetical protein	turquoise
orthoMCL4542	RPA2294	probable transcriptional regulator, TetR family	turquoise
orthoMCL0004	RPA2297	conserved unknown protein	turquoise
orthoMCL3457	RPA2298	hypothetical protein	turquoise
orthoMCL1424	RPA2424	conserved unknown protein	turquoise
orthoMCL1401	RPA2464	sufB, needed for fhuF Fe-S center stability	turquoise
orthoMCL1375	RPA2517	conserved hypothetical protein	turquoise
orthoMCL1341	RPA2585	hypothetical protein	turquoise
orthoMCL4502	RPA2608	possible sulfonate binding protein	turquoise
orthoMCL3416	RPA2609	possible monooxygenase	turquoise
orthoMCL4501	RPA2610	aliphatic sulfonate transport ATP-binding protein, Subunit of ABC transporter	turquoise
orthoMCL3912	RPA2611	putative aliphatic sulfonate transport membrane component. Permease subunit of an ABC transporter	turquoise
orthoMCL4500	RPA2612	methanesulfonate sulfonatase MsuD (monooxygenase)	turquoise
orthoMCL4499	RPA2614	conserved hypothetical protein	turquoise
orthoMCL4498	RPA2615	putative nitrogenase iron protein (nitrogenase component II) (nitrogenase reductase)	turquoise
orthoMCL4496	RPA2617	possible vanadium nitrogenase associated protein vnfN (U51863)	turquoise
orthoMCL3402	RPA2755	possible DNA-binding stress protein	turquoise
orthoMCL2916	RPA2787	putative signal transduction histidine kinase with PAS/PAC domain	turquoise
orthoMCL1185	RPA2871	probable RhtB family transporter, amino acid efflux	turquoise
orthoMCL1022	RPA3134	conserved unknown protein	turquoise
orthoMCL0979	RPA3198	conserved hypothetical protein	turquoise
orthoMCL0977	RPA3205	Type III secretion proteins, related to flagellar biosynthesis protein FlhB	turquoise
orthoMCL0973	RPA3218	Protein of unknown function UPF0047	turquoise
orthoMCL2876	RPA3256	hypothetical protein	turquoise

orthoMCL2867	RPA3401	hypothetical protein	turquoise
orthoMCL0892	RPA3418	conserved hypothetical protein	turquoise
orthoMCL3852	RPA3457	Biotin/lipoyl attachment:Biotin-requiring enzyme, attachment site	turquoise
orthoMCL0805	RPA3565	ErfK/YbiS/YcfS/YnhG	turquoise
orthoMCL0784	RPA3616	putative diguanylate cyclase (GGDEF) with PAS/PAC domain	turquoise
orthoMCL3838	RPA3661	Metal dependent phosphohydrolase	turquoise
orthoMCL0753	RPA3663	urease gamma subunit	turquoise
orthoMCL0747	RPA3669	putative urea short-chain amide or branched-chain amino acid uptake ABC transporter periplasmic solute-binding protein precursor	turquoise
orthoMCL0746	RPA3670	putative ATP-dependent RNA helicase	turquoise
orthoMCL0741	RPA3676	putative type IV prepilin peptidase, cpaA	turquoise
orthoMCL0698	RPA3747	putative ethanolamine ammonia-lyase light chain	turquoise
orthoMCL0697	RPA3748	Elongator protein 3/MiaB/NifB	turquoise
orthoMCL0691	RPA3754	Transglutaminase-like domain	turquoise
orthoMCL2826	RPA3867	conserved hypothetical protein	turquoise
orthoMCL0610	RPA3899	putative flagellar basal-body rod protein flgF	turquoise
orthoMCL0607	RPA3902	putative flagellar L-ring protein FlgH	turquoise
orthoMCL0603	RPA3909	flagellar P-ring protein FlgI	turquoise
orthoMCL0602	RPA3910	conserved hypothetical protein	turquoise
orthoMCL0595	RPA3932	probable flagellar hook protein flgE	turquoise
orthoMCL0581	RPA3957	Hpt domain	turquoise
orthoMCL3817	RPA3960	conserved unknown protein	turquoise
orthoMCL3284	RPA4148	putative ribose-phosphate pyrophosphokinase	turquoise
orthoMCL2799	RPA4149	Beta-lactamase-like	turquoise
orthoMCL0481	RPA4212	conserved hypothetical protein	turquoise
orthoMCL0480	RPA4213	sulfite reductase hemoprotein subunit	turquoise
orthoMCL0479	RPA4214	conserved hypothetical protein	turquoise
orthoMCL0478	RPA4215	putative siroheme synthase	turquoise
orthoMCL0449	RPA4251	O-acetylhomoserine sulfhydrylase	turquoise
orthoMCL0345	RPA4411	glycerol-3-phosphate regulon repressor glpR, DeoR family	turquoise
orthoMCL0336	RPA4423	conserved unknown protein	turquoise
orthoMCL2775	RPA4461	possible cytochrome subunit of sulfide dehydrogenase	turquoise
orthoMCL0312	RPA4485	adenine glycosylase mutY	turquoise
orthoMCL0213	RPA4692	unknown protein	turquoise
orthoMCL0138	RPA4792	RNA polymerase ECF-type sigma factor	turquoise
orthoMCL2696	NA	NA	yellow
orthoMCL2697	NA	NA	yellow
orthoMCL2706	NA	NA	yellow
orthoMCL3227	NA	NA	yellow
orthoMCL3667	NA	NA	yellow
orthoMCL3668	NA	NA	yellow
orthoMCL3725	NA	NA	yellow
orthoMCL4180	NA	NA	yellow
orthoMCL4181	NA	NA	yellow
orthoMCL4663	NA	NA	yellow
orthoMCL4669	NA	NA	yellow

orthoMCL4672	NA	NA	yellow
orthoMCL4679	NA	NA	yellow
orthoMCL4680	NA	NA	yellow
orthoMCL4681	NA	NA	yellow
orthoMCL4755	NA	NA	yellow
orthoMCL4757	NA	NA	yellow
orthoMCL4758	NA	NA	yellow
orthoMCL4792	NA	NA	yellow
orthoMCL4793	NA	NA	yellow
orthoMCL4795	NA	NA	yellow
orthoMCL4856	NA	NA	yellow
orthoMCL4963	NA	NA	yellow
orthoMCL5214	NA	NA	yellow
orthoMCL3169	RPA0372	transcriptional regulators, AsnC family	yellow
orthoMCL2342	RPA0488	possible CarD-like transcriptional regulator	yellow
orthoMCL2213	RPA0713	conserved unknown protein	yellow
orthoMCL2209	RPA0717	putative cob(I)alamin adenosyltransferase	yellow
orthoMCL2203	RPA0723	possible hemin ABC transporter, permease component	yellow
orthoMCL4629	RPA0893	conserved hypothetical protein	yellow
orthoMCL4029	RPA1431	putative NAD+ ADP-ribosyltransferase	yellow
orthoMCL4017	RPA1461	conserved hypothetical protein	yellow
orthoMCL1850	RPA1474	hypothetical protein	yellow
orthoMCL1714	RPA1674	response regulator receiver, CheY3	yellow
orthoMCL1526	RPA2038	putative oxidoreductase	yellow
orthoMCL1509	RPA2086	putative precorrin 6y methylase	yellow
orthoMCL2960	RPA2094	putative nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	yellow
orthoMCL2959	RPA2095	possible cobalamin (5`-phosphate) synthase	yellow
orthoMCL1501	RPA2097	putative cobF protein	yellow
orthoMCL4546	RPA2270	hypothetical protein	yellow
orthoMCL4525	RPA2345	unknown protein	yellow
orthoMCL3943	RPA2348	possible nitrogenase molybdenum-iron protein alpha chain (nitrogenase component I) (dinitrogenase)	yellow
orthoMCL5676	RPA2351	transcriptional regulator, AsnC family	yellow
orthoMCL5675	RPA2352	conserved hypothetical protein	yellow
orthoMCL5674	RPA2353	putative nitrogenase NifH subunit	yellow
orthoMCL5673	RPA2354	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	yellow
orthoMCL5672	RPA2355	possible nitrogenase NifB	yellow
orthoMCL4524	RPA2356	putative cystathionine beta-synthase	yellow
orthoMCL5671	RPA2357	cystathionine gamma-lyase	yellow
orthoMCL5670	RPA2358	2OG-Fe(II) oxygenase superfamily	yellow
orthoMCL5669	RPA2359	putative periplasmic protein	yellow
orthoMCL5668	RPA2360	ABC transporter, ATP-binding protein	yellow
orthoMCL5667	RPA2361	putative ABC transporter, permease protein	yellow
orthoMCL5666	RPA2362	putative O-acetylhomoserine sulfhydrylase	yellow
orthoMCL4523	RPA2365	putative L-allo-threonine aldolase	yellow
orthoMCL0008	RPA2493	possible P-methylase	yellow

orthoMCL5645	RPA2607	possible transcriptional regulator, XRE family, CUPIN domain	yellow
orthoMCL5644	RPA2613	putative aliphatic sulfonate binding protein, subunit of ABC transporter	yellow
orthoMCL5635	RPA2631	possible ABC transport substrate-binding periplasmic protein	yellow
orthoMCL5633	RPA2634	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	yellow
orthoMCL5632	RPA2635	putative nitrogenase reductase NifH subunit	yellow
orthoMCL3905	RPA2676	transcriptional regulator, LysR family	yellow
orthoMCL2912	RPA2808	conserved hypothetical protein	yellow
orthoMCL1179	RPA2882	conserved hypothetical protein	yellow
orthoMCL5611	RPA3009	light harvesting protein B-800-850, beta chain C (antenna pigment protein, beta chain C) (LH II-C beta)	yellow
orthoMCL5610	RPA3010	pseudo light-harvesting protein	yellow
orthoMCL4474	RPA3011	unknown protein	yellow
orthoMCL1101	RPA3012	light harvesting protein B-800-850, alpha chain D (antenna pigment protein, alpha chain D) (LH II-D alpha)	yellow
orthoMCL5609	RPA3013	light harvesting protein B-800-850, beta chain D (antenna pigment protein, beta chain D) (LH II-D beta)	yellow
orthoMCL3876	RPA3023	aldo/keto reductase	yellow
orthoMCL1093	RPA3035	hypothetical protein	yellow
orthoMCL5608	RPA3036	hypothetical protein	yellow
orthoMCL2885	RPA3144	conserved hypothetical protein	yellow
orthoMCL0947	RPA3247	50S ribosomal protein L2	yellow
orthoMCL0946	RPA3248	50S ribosomal protein L23	yellow
orthoMCL0945	RPA3249	50S ribosomal protein L4	yellow
orthoMCL0944	RPA3250	50S ribosomal protein L3	yellow
orthoMCL0943	RPA3251	30S ribosomal protein S10	yellow
orthoMCL0014	RPA3252	elongation factor Tu	yellow
orthoMCL0942	RPA3253	elongation factor G	yellow
orthoMCL4446	RPA3335	hypothetical protein	yellow
orthoMCL2869	RPA3389	hypothetical protein	yellow
orthoMCL2852	RPA3581	putative 8.8 KD protein Y4HR	yellow
orthoMCL3331	RPA3585	hypothetical protein	yellow
orthoMCL3317	RPA3792	putative diguanylate cyclase (GGDEF)	yellow
orthoMCL0594	RPA3935	hypothetical protein	yellow
orthoMCL0569	RPA3971	phosphomethylpyrimidine kinase (hmp-phosphate kinase)	yellow
orthoMCL0555	RPA3989	putative lipopolysaccharide-heptosyl-transferase	yellow
orthoMCL5547	RPA4011	possible serine protease/outer membrane autotransporter	yellow

Module members are sorted by module colors and strain CGA009's gene numbering (RPA number).

Table 4.7. List of module members in the down-regulated NF-low/NF-high co-expression networks.

orthoMCL ID	RPA Number	Product Description	Module Color
orthoMCL5470	NA	NA	black
orthoMCL5472	NA	NA	black
orthoMCL2679	RPA0003	putative RecF protein	black
orthoMCL2493	RPA0281	possible exodeoxyribonuclease III	black
orthoMCL2489	RPA0285	Protein of unknown function UPF0001	black
orthoMCL2474	RPA0301	putative DNA polymerase III epsilon chain	black
orthoMCL2276	RPA0594	putative mutator protein mutT	black
orthoMCL2253	RPA0618	unknown protein	black
orthoMCL2249	RPA0622	putative methionyl-tRNA formyltransferase	black
orthoMCL2248	RPA0623	putative tRNA-pseudouridine synthase	black
orthoMCL3131	RPA0658	benzoyl-CoA reductase subunit	black
orthoMCL2214	RPA0712	putative nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	black
orthoMCL2115	RPA0869	GCN5-related N-acetyltransferase	black
orthoMCL2114	RPA0870	putative ornithine decarboxylase	black
orthoMCL2102	RPA0888	putative lysophospholipase L2	black
orthoMCL2081	RPA0918	possible 50S ribosomal protein L31	black
orthoMCL2080	RPA0919	ABC transporter, ATP-binding protein	black
orthoMCL3080	RPA0982	putativetranscriptional regulator PcaR, IclR family	black
orthoMCL1973	RPA1128	conserved hypothetical protein	black
orthoMCL1966	RPA1137	unknown protein	black
orthoMCL1954	RPA1151	conserved hypothetical protein	black
orthoMCL1940	RPA1166	N-6 Adenine-specific DNA methylase:Conserved hypothetical protein 95	black
orthoMCL1908	RPA1201	chorismate synthase	black
orthoMCL1830	RPA1514	putative coproporphyrinogen III oxidase precursor	black
orthoMCL1781	RPA1579	L-carnitine dehydratase/bile acid-inducible protein F	black
orthoMCL1751	RPA1619	hypothetical protein	black
orthoMCL1704	RPA1702	putative acyl-CoA ligase	black
orthoMCL2990	RPA1789	putative branched-chain amino acid transport system substrate-binding protein	black
orthoMCL1595	RPA1922	CycL cytochrome C-type biogenesis protein	black
orthoMCL1575	RPA1956	conserved hypothetical protein	black
orthoMCL1558	RPA1985	probable diacylglycerol kinase	black
orthoMCL1510	RPA2085	cobalamin biosynthesis protein G	black
orthoMCL1505	RPA2090	precorrin isomerase CobH	black
orthoMCL3953	RPA2295	unknown protein	black
orthoMCL4494	RPA2623	possible sulfonate transport system permease protein. Subunit of an ABC transporter	black
orthoMCL1286	RPA2694	pyridoxal phosphate biosynthetic protein pdxJ	black
orthoMCL1207	RPA2841	survival protein surE	black
orthoMCL1143	RPA2932	hypothetical protein	black
orthoMCL0072	RPA3092	putative oxidoreductase	black
orthoMCL1035	RPA3113	conserved hypothetical protein	black

orthoMCL0999	RPA3160	DUF218	black
orthoMCL0929	RPA3275	preprotein translocase, SecE subunit	black
orthoMCL0921	RPA3293	putative branched-chain amino acid transport system ATP-binding protein	black
orthoMCL0920	RPA3294	putative branched-chain amino acid transport system ATP-binding protein	black
orthoMCL0069	RPA3295	possible branched-chain amino acid ABC transporter, permease protein	black
orthoMCL0068	RPA3296	possible ABC transporter subunit (U75364)	black
orthoMCL0852	RPA3503	putative D-lactate dehydrogenase, oxidoreductase	black
orthoMCL0800	RPA3575	thiamin biosynthesis ThiG	black
orthoMCL0704	RPA3741	putative oxidoreductase	black
orthoMCL0673	RPA3782	putative rubredoxin reductase	black
orthoMCL0668	RPA3791	probable transcriptional regulator, TetR family	black
orthoMCL0473	RPA4222	hypothetical protein	black
orthoMCL2787	RPA4298	ATP/GTP-binding site motif A (P-loop)	black
orthoMCL0378	RPA4372	Class I peptide chain release factor domain	black
orthoMCL0289	RPA4529	putative arsenate reductase	black
orthoMCL0238	RPA4647	probable transcriptional regulator KdgR, IclR family	black
orthoMCL0115	NA	NA	blue
orthoMCL2714	NA	NA	blue
orthoMCL3216	NA	NA	blue
orthoMCL4205	NA	NA	blue
orthoMCL4219	NA	NA	blue
orthoMCL4286	NA	NA	blue
orthoMCL4874	NA	NA	blue
orthoMCL4890	NA	NA	blue
orthoMCL4912	NA	NA	blue
orthoMCL4951	NA	NA	blue
orthoMCL4955	NA	NA	blue
orthoMCL5036	NA	NA	blue
orthoMCL5211	NA	NA	blue
orthoMCL5411	NA	NA	blue
orthoMCL2669	RPA0022	DSBA oxidoreductase	blue
orthoMCL2635	RPA0059	L-carnitine dehydratase/bile acid-inducible protein F	blue
orthoMCL2629	RPA0065	putative protease IV	blue
orthoMCL0000	RPA0096	putative multidrug-efflux transport protein	blue
orthoMCL3642	RPA0143	response regulator receiver, CheY1	blue
orthoMCL2600	RPA0159	ribosomal protein L27	blue
orthoMCL2476	RPA0299	putative maf protein	blue
orthoMCL3160	RPA0499	DedA family	blue
orthoMCL3619	RPA0516	transcriptional regulator, MarR family	blue
orthoMCL2301	RPA0548	Protein of unknown function UPF0118	blue
orthoMCL2281	RPA0578	unknown protein	blue
orthoMCL2263	RPA0608	conserved unknown protein	blue
orthoMCL2244	RPA0628	conserved unknown protein	blue
orthoMCL2194	RPA0747	putative sulfate ABC transporter, ATP-binding component	blue
orthoMCL2193	RPA0748	possible sulfate ABC transporter, permease component	blue

orthoMCL2191	RPA0750	sulfate ABC transporter, periplasmic binding protein component	blue
orthoMCL2190	RPA0751	putative phosphoadenosine phosphosulfate reductase	blue
orthoMCL2189	RPA0752	putative ATP sulfurylase small subunit	blue
orthoMCL2188	RPA0753	putative CysN/CysC bifunctional enzyme, ATP-sulfurylase large subunit and adenylyl sulfate kinase	blue
orthoMCL2149	RPA0828	transcriptional regulator, MarR family	blue
orthoMCL2145	RPA0833	heme O synthase	blue
orthoMCL2134	RPA0845	probable ATP synthase subunit C TRANSMEMBRANE protein	blue
orthoMCL2100	RPA0890	hypothetical protein	blue
orthoMCL3108	RPA0923	conserved hypothetical protein	blue
orthoMCL1988	RPA1103	Thioesterase superfamily:4-hydroxybenzoyl-CoA thioesterase	blue
orthoMCL1896	RPA1223	hypothetical protein	blue
orthoMCL0095	RPA1420	putative inner membrane component for iron transport	blue
orthoMCL1855	RPA1421	possible efflux protein	blue
orthoMCL3037	RPA1423	putative membrane protein	blue
orthoMCL3035	RPA1425	serine acetyltransferase	blue
orthoMCL3034	RPA1426	ABC transporter, ATP-binding protein	blue
orthoMCL3033	RPA1427	putative ABC transporter, permease protein	blue
orthoMCL3032	RPA1428	possible lipoprotein	blue
orthoMCL3030	RPA1430	putative outer membrane protein	blue
orthoMCL1668	RPA1777	DUF35	blue
orthoMCL1660	RPA1809	hypothetical protein	blue
orthoMCL1628	RPA1864	hypothetical protein	blue
orthoMCL1627	RPA1865	conserved hypothetical protein	blue
orthoMCL1619	RPA1894	hypothetical protein	blue
orthoMCL1540	RPA2018	alcohol dehydrogenase	blue
orthoMCL1537	RPA2021	3-hydroxymyristoyl-acyl carrier protein dehydratase	blue
orthoMCL1536	RPA2022	specialized acyl carrier protein	blue
orthoMCL3476	RPA2071	hypothetical protein	blue
orthoMCL1490	RPA2132	hypothetical protein	blue
orthoMCL1489	RPA2133	conserved hypothetical protein	blue
orthoMCL1472	RPA2158	hypothetical protein	blue
orthoMCL2952	RPA2163	SEC-C motif	blue
orthoMCL1450	RPA2191	DUF188	blue
orthoMCL1364	RPA2549	conserved hypothetical protein	blue
orthoMCL1321	RPA2606	tRNA guanine transglycosylase	blue
orthoMCL1320	RPA2644	putative ABC transporter permease protein	blue
orthoMCL1235	RPA2779	possible hydrolase	blue
orthoMCL1205	RPA2847	putative sec-independent protein translocase component TatC	blue
orthoMCL1176	RPA2888	triose-phosphate isomerase	blue
orthoMCL3384	RPA3055	Integral membrane protein TerC family	blue
orthoMCL1066	RPA3066	dimethyladenosine transferase	blue
orthoMCL2887	RPA3109	conserved hypothetical protein	blue
orthoMCL0950	RPA3244	30S ribosomal protein S3	blue
orthoMCL0933	RPA3271	putative D-isomer specific 2-hydroxyacid dehydrogenase	blue
orthoMCL2874	RPA3278	putative exbB, uptake of enterochelin	blue
orthoMCL0776	RPA3629	conserved hypothetical protein	blue

orthoMCL0769	RPA3639	Cof-subfamily: Cof-like hydrolase	blue
orthoMCL0701	RPA3744	conserved unknown protein	blue
orthoMCL0636	RPA3849	glycine cleavage system protein H	blue
orthoMCL2823	RPA3873	hypothetical protein	blue
orthoMCL2822	RPA3874	hypothetical protein	blue
orthoMCL0618	RPA3888	possible flagellar basal-body rod protein FlgB	blue
orthoMCL0583	RPA3954	thioredoxin reductase	blue
orthoMCL0545	RPA4030	hypothetical protein	blue
orthoMCL2802	RPA4141	conserved hypothetical protein	blue
orthoMCL3284	RPA4148	putative ribose-phosphate pyrophosphokinase	blue
orthoMCL0481	RPA4212	conserved hypothetical protein	blue
orthoMCL0480	RPA4213	sulfite reductase hemoprotein subunit	blue
orthoMCL0479	RPA4214	conserved hypothetical protein	blue
orthoMCL0057	RPA4231	putative oxidoreductase	blue
orthoMCL0459	RPA4241	CBS domain	blue
orthoMCL0449	RPA4251	O-acetylhomoserine sulfhydrylase	blue
orthoMCL0426	RPA4293	hypothetical protein	blue
orthoMCL0416	RPA4317	putative fosmidomycin resistance protein	blue
orthoMCL0385	RPA4364	conserved hypothetical protein	blue
orthoMCL0374	RPA4376	Signal peptidase II, family A8	blue
orthoMCL0356	RPA4398	putative branched-chain amino acid ABC transporter, ATP-binding protein	blue
orthoMCL0312	RPA4485	adenine glycosylase mutY	blue
orthoMCL3260	RPA4491	conserved hypothetical protein	blue
orthoMCL2760	RPA4502	putative outer membrane protein	blue
orthoMCL0298	RPA4508	conserved hypothetical protein	blue
orthoMCL0189	RPA4732	uncharacterized cation transport protein chaC	blue
orthoMCL0182	RPA4741	diaminopimelate decarboxylase	blue
orthoMCL3198	NA	NA	brown
orthoMCL3657	RPA0095	putative multidrug efflux membrane fusion protein	brown
orthoMCL2580	RPA0179	putative H+-transporting ATP synthase delta chain.	brown
orthoMCL2563	RPA0196	putative ABC transporter, ATP-binding protein	brown
orthoMCL2553	RPA0207	unknown protein	brown
orthoMCL2530	RPA0233	putative Citrate lyase beta chain (acyl lyase subunit) (citE)	brown
orthoMCL5802	RPA0237	conserved hypothetical protein	brown
orthoMCL2468	RPA0308	heat shock protein HslV, proteasome-related peptidase subunit	brown
orthoMCL2427	RPA0358	two-component transcriptional regulator ChvI, winged helix family	brown
orthoMCL2378	RPA0435	putative ribosome-binding factor A	brown
orthoMCL2358	RPA0457	possible cabohydrate kinases	brown
orthoMCL4644	RPA0472	OmpA-like transmembrane domain	brown
orthoMCL4085	RPA0540	hypothetical protein	brown
orthoMCL2173	RPA0793	conserved hypothetical protein	brown
orthoMCL2132	RPA0847	FoF1 ATP synthase, subunit I	brown
orthoMCL2130	RPA0849	conserved hypothetical protein	brown
orthoMCL3090	RPA0971	putative hydrogenase expression/formation protein hupJ	brown
orthoMCL4626	RPA0990	Bacteriophytochrome (light-regulated signal transduction histidine kinase), PhyB5	brown

orthoMCL2052	RPA0998	conserved hypothetical protein	brown
orthoMCL5779	RPA1008	probable transcriptional regulator, AraC family	brown
orthoMCL2024	RPA1045	glycyl-tRNA synthetase alpha chain	brown
orthoMCL1999	RPA1091	hypothetical protein	brown
orthoMCL1955	RPA1150	putative histidyl-tRNA synthetase	brown
orthoMCL3065	RPA1200	possible phosphoglycerate mutase	brown
orthoMCL4621	RPA1207	PAS domain:sigma-54-dependent transcriptional regulator, Fis family	brown
orthoMCL1888	RPA1241	possible tonB transport protein	brown
orthoMCL4595	RPA1418	possible transport system permease protein	brown
orthoMCL1854	RPA1422	unknown protein	brown
orthoMCL5740	RPA1564	possible urea/short-chain amide transport system substrate-binding protein	brown
orthoMCL5735	RPA1601	possible transcriptional regulator, MarR family	brown
orthoMCL1722	RPA1665	hypothetical protein	brown
orthoMCL5732	RPA1685	possible serine protease/outer membrane autotransporter	brown
orthoMCL3011	RPA1719	Protein of unknown function UPF0153	brown
orthoMCL4575	RPA1815	hypothetical protein	brown
orthoMCL5723	RPA1830	hypothetical protein	brown
orthoMCL5722	RPA1831	conserved hypothetical protein	brown
orthoMCL3986	RPA1850	methyl-accepting chemotaxis receptor/sensory transducer	brown
orthoMCL1556	RPA1995	conserved hypothetical protein	brown
orthoMCL5709	RPA2026	ferric siderophore receptor	brown
orthoMCL1528	RPA2033	possible AtsE	brown
orthoMCL3464	RPA2152	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL) with PAS domain	brown
orthoMCL5704	RPA2159	hypothetical protein	brown
orthoMCL1470	RPA2165	chaperonin GroES2, cpn10	brown
orthoMCL1464	RPA2174	conserved hypothetical protein	brown
orthoMCL3952	RPA2300	Bacterial regulatory protein, LuxR family	brown
orthoMCL3947	RPA2306	Nickel-dependent hydrogenase b-type cytochrome subunit	brown
orthoMCL5688	RPA2316	enoyl-CoA hydratase	brown
orthoMCL5685	RPA2320	possible TctA subunit of the Tripartite Tricarboxylate Transport(TTT) Family	brown
orthoMCL5683	RPA2340	putative membrane protein	brown
orthoMCL5663	RPA2366	hypothetical protein	brown
orthoMCL3941	RPA2375	conserved hypothetical protein	brown
orthoMCL5659	RPA2400	conserved hypothetical protein	brown
orthoMCL1398	RPA2467	sufS, putative selenosysteine lyase	brown
orthoMCL4511	RPA2497	putative anion ABC transporter, ATP-binding protein	brown
orthoMCL1389	RPA2498	possible ABC transporter, permease protein	brown
orthoMCL4510	RPA2499	possible ABC transporter, periplasmic protein	brown
orthoMCL4509	RPA2500	possible amidase	brown
orthoMCL0016	RPA2535	putative UDP-glucose:(heptosyl) LPS alpha 1,3-glucosyltransferase	brown
orthoMCL1288	RPA2692	RNA polymerase omega subunit	brown
orthoMCL1252	RPA2757	conserved hypothetical protein	brown
orthoMCL1168	RPA2899	conserved hypothetical protein	brown
orthoMCL4482	RPA2908	Cytidine/deoxycytidylate deaminase:Tat pathway signal	brown

orthoMCL5617	RPA2963	hypothetical protein	brown
orthoMCL1097	RPA3030	hypothetical protein	brown
orthoMCL1063	RPA3070	conserved unknown protein	brown
orthoMCL3372	RPA3196	hypothetical protein	brown
orthoMCL5597	RPA3277	putative exbD, uptake of enterochelin	brown
orthoMCL5594	RPA3282	RNA polymerase ECF-type sigma factor, possible FecI	brown
orthoMCL5584	RPA3373	hypothetical protein	brown
orthoMCL2857	RPA3470	putative sugar uptake ABC transporter periplasmic solute-binding protein precursor	brown
orthoMCL0801	RPA3574	putative thiamin biosynthesis ThiS	brown
orthoMCL0724	RPA3694	3-deoxy-manno-octulosonate cytidylyltransferase	brown
orthoMCL0723	RPA3695	chorismate mutase/prephenate dehydratase	brown
orthoMCL2841	RPA3724	possible high-affinity leucine-isoleucine-valine transport system	brown
orthoMCL0585	RPA3945	hypothetical protein	brown
orthoMCL0557	RPA3986	putative ADP-heptose--LPS heptosyltransferase II	brown
orthoMCL4385	RPA4164	possible aliphatic sulfonate binding protein of ABC transporter system	brown
orthoMCL4384	RPA4166	putative nitrilase	brown
orthoMCL5515	RPA4169	Bacterial regulatory protein, GntR family	brown
orthoMCL0492	RPA4190	conserved unknown protein	brown
orthoMCL5510	RPA4280	probable transcriptional regulator, AraC family	brown
orthoMCL0365	RPA4388	GCN5-related N-acetyltransferase	brown
orthoMCL4370	RPA4430	putative TonB-dependent receptor	brown
orthoMCL3793	RPA4482	putative sensor (PAS) domain for methyl-accepting chemotaxis sensory transducer	brown
orthoMCL0308	RPA4490	conserved hypothetical protein	brown
orthoMCL0307	RPA4493	conserved unknown protein	brown
orthoMCL0290	RPA4527	putative septum formation maf protein	brown
orthoMCL0265	RPA4609	putative nifU protein	brown
orthoMCL2749	RPA4622	hypothetical protein	brown
orthoMCL5498	RPA4655	L-fuculose phosphate aldolase	brown
orthoMCL2684	NA	NA	green
orthoMCL2690	NA	NA	green
orthoMCL3208	NA	NA	green
orthoMCL3217	NA	NA	green
orthoMCL3220	NA	NA	green
orthoMCL4257	NA	NA	green
orthoMCL4940	NA	NA	green
orthoMCL5236	NA	NA	green
orthoMCL5331	NA	NA	green
orthoMCL4100	RPA0153	possible outer membrane receptor for Fe transport	green
orthoMCL2599	RPA0160	possible acetyltransferases.	green
orthoMCL2592	RPA0167	DUF163	green
orthoMCL2579	RPA0180	Rare lipoprotein A	green
orthoMCL2566	RPA0193	AFG1-like ATPase	green
orthoMCL2546	RPA0215	possible general stress protein 26	green
orthoMCL2383	RPA0429	catalase/peroxidase	green
orthoMCL2327	RPA0505	conserved hypothetical protein	green

orthoMCL2246	RPA0626	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	green
orthoMCL3608	RPA0684	hypothetical protein	green
orthoMCL2185	RPA0775	hypothetical protein	green
orthoMCL3094	RPA0967	hydrogenase expression/formation protein hupF	green
orthoMCL2050	RPA1000	Nitrogenase-associated protein:Arsenate reductase and related	green
orthoMCL2049	RPA1001	conserved hypothetical protein	green
orthoMCL4060	RPA1009	possible cytochrome P450	green
orthoMCL2035	RPA1033	DUF85:Elongator protein 3/MiaB/NifB	green
orthoMCL2012	RPA1065	Metal dependent phosphohydrolase	green
orthoMCL3048	RPA1275	conserved hypothetical protein	green
orthoMCL1844	RPA1499	hypothetical protein	green
orthoMCL1812	RPA1534	hypothetical protein	green
orthoMCL4005	RPA1578	ferredoxin--NADP+ reductase	green
orthoMCL1764	RPA1604	conserved hypothetical protein	green
orthoMCL3491	RPA1934	hypothetical protein	green
orthoMCL3473	RPA2105	non-heme chloroperoxidase	green
orthoMCL1422	RPA2432	putative signal transduction histidine kinase with response regulator receiver domain	green
orthoMCL1378	RPA2513	elongation factor P	green
orthoMCL1326	RPA2601	phosphopantetheine adenylyltransferase	green
orthoMCL5643	RPA2618	putative sulfonate transport system substrate-binding protein	green
orthoMCL3911	RPA2624	putative sulfonate transport system substrate-binding protein	green
orthoMCL4493	RPA2626	possible hydrolase	green
orthoMCL5639	RPA2627	putative carboxylesterase	green
orthoMCL1273	RPA2717	conserved hypothetical protein	green
orthoMCL1269	RPA2726	putative riboflavin-specific deaminase / reductase	green
orthoMCL1197	RPA2855	dGTP triphosphohydrolase	green
orthoMCL1170	RPA2894	conserved hypothetical protein	green
orthoMCL1160	RPA2910	conserved hypothetical protein	green
orthoMCL1139	RPA2936	putative biotin-protein ligase birA	green
orthoMCL3881	RPA2988	GCN5-related N-acetyltransferase	green
orthoMCL1078	RPA3052	phosphoribosylglycinamide formyltransferase	green
orthoMCL2884	RPA3153	conserved hypothetical protein	green
orthoMCL2873	RPA3308	ycfI, putative structural proteins	green
orthoMCL0918	RPA3309	conserved unknown protein	green
orthoMCL0917	RPA3312	Transglutaminase-like domain	green
orthoMCL0916	RPA3313	hypothetical protein	green
orthoMCL0888	RPA3424	hypothetical protein	green
orthoMCL0803	RPA3568	conserved unknown protein	green
orthoMCL0802	RPA3573	thiamine biosynthesis oxidoreductase thiO	green
orthoMCL0786	RPA3614	conserved unknown protein	green
orthoMCL2850	RPA3652	conserved hypothetical protein	green
orthoMCL0758	RPA3653	Protein of unknown function UPF0033	green
orthoMCL0720	RPA3702	methionine synthase	green
orthoMCL5533	RPA4099	hypothetical protein	green
orthoMCL5532	RPA4100	hypothetical protein	green
orthoMCL0504	RPA4174	phosphoribosylaminoimidazole carboxylase catalytic subunit	green

orthoMCL0502	RPA4176	ribosomal protein S21	green
orthoMCL3273	RPA4305	hypothetical protein	green
orthoMCL2786	RPA4313	putative acetyltransfersase	green
orthoMCL0235	RPA4660	putative alpha,alpha-trehalose-phosphate synthase (UDP-forming) (trehalose-6-phosphate synthase)	green
orthoMCL0118	RPA4835	conserved hypothetical protein	green
orthoMCL3098	RPA0962	hydrogenase small chain	grey
orthoMCL4069	RPA0983	possible phthalate dioxygenase	grey
orthoMCL3061	RPA1216	putative ABC transporter, permease protein	grey
orthoMCL5742	RPA1562	transcriptional regulator, LysR family	grey
orthoMCL0200	RPA4718	putative molybdate transport system transcriptional regulator, ModE	grey
orthoMCL2736	RPA4719	molybdo-pterin binding protein	grey
orthoMCL0199	RPA4720	conserved hypothetical protein	grey
orthoMCL3192	NA	NA	magenta
orthoMCL3757	NA	NA	magenta
orthoMCL3760	NA	NA	magenta
orthoMCL4285	NA	NA	magenta
orthoMCL4292	NA	NA	magenta
orthoMCL4683	NA	NA	magenta
orthoMCL4686	NA	NA	magenta
orthoMCL4687	NA	NA	magenta
orthoMCL4891	NA	NA	magenta
orthoMCL4902	NA	NA	magenta
orthoMCL4975	NA	NA	magenta
orthoMCL2647	RPA0047	conserved hypothetical protein	magenta
orthoMCL3643	RPA0142	multidomain chemotaxis histidine kinase CheA1 (Hpt, CheA, & CheW domains)	magenta
orthoMCL4101	RPA0152	Protein of unknown function, UPF0066	magenta
orthoMCL2447	RPA0331	possible heat shock protein (HSP-70 COFACTOR), grpE	magenta
orthoMCL2387	RPA0424	transcriptional regulator, FUR family	magenta
orthoMCL2366	RPA0449	possible hydrolases/phosphatases	magenta
orthoMCL2280	RPA0584	transcriptional accessory protein	magenta
orthoMCL2264	RPA0607	putative protoporphyrinogen oxidase, hemK protein	magenta
orthoMCL3588	RPA0764	conserved hypothetical protein	magenta
orthoMCL3587	RPA0766	transcriptional regulator, Crp/Fnr family	magenta
orthoMCL3109	RPA0871	conserved hypothetical protein	magenta
orthoMCL2092	RPA0906	putative N-formylglutamate amidohydrolase	magenta
orthoMCL2070	RPA0938	conserved unknown protein	magenta
orthoMCL4068	RPA0984	putative glutamine synthetase-like protein	magenta
orthoMCL3570	RPA0997	possible transcriptional regulator, ArsR family	magenta
orthoMCL2007	RPA1072	conserved hypothetical protein	magenta
orthoMCL1843	RPA1500	unknown protein	magenta
orthoMCL3027	RPA1508	conserved hypothetical protein	magenta
orthoMCL1761	RPA1607	Appr-1"-p processing enzyme family protein homolog	magenta
orthoMCL3486	RPA1960	putative RND efflux membrane protein	magenta
orthoMCL1565	RPA1977	possible TrapT family, dctQ subunit, glutamate transport	magenta
orthoMCL1328	RPA2598	GTP binding protein-like	magenta

orthoMCL2892	RPA3054	possible transcriptional regulator, Crp/Fnr family	magenta
orthoMCL1017	RPA3139	DUF482	magenta
orthoMCL0941	RPA3254	30S ribosomal protein S7	magenta
orthoMCL2867	RPA3401	hypothetical protein	magenta
orthoMCL3351	RPA3453	putative enoyl-CoA hydratase	magenta
orthoMCL3342	RPA3477	exbD, uptake of enterochelin	magenta
orthoMCL3341	RPA3478	possible exbB, uptake of enterochelin	magenta
orthoMCL3340	RPA3481	hypothetical protein	magenta
orthoMCL0756	RPA3658	possible urease accessory protein UreF	magenta
orthoMCL0647	RPA3817	adenylosuccinate lyase	magenta
orthoMCL0587	RPA3943	conserved hypothetical protein	magenta
orthoMCL3281	RPA4179	conserved unknown protein	magenta
orthoMCL2792	RPA4211	glutaminase A	magenta
orthoMCL0400	RPA4349	conserved unknown protein	magenta
orthoMCL0395	RPA4354	putative GTP-binding protein	magenta
orthoMCL2755	RPA4521	Thioesterase superfamily	magenta
orthoMCL0243	RPA4633	short-chain dehydrogenase	magenta
orthoMCL0031	RPA4634	hypothetical protein	magenta
orthoMCL3771	RPA4804	conserved hypothetical protein	magenta
orthoMCL2724	RPA4823	putative long-chain-fatty-acid-CoA ligase	magenta
orthoMCL3191	NA	NA	pink
orthoMCL4925	NA	NA	pink
orthoMCL5022	NA	NA	pink
orthoMCL5193	NA	NA	pink
orthoMCL2634	RPA0060	conserved unknown protein	pink
orthoMCL3183	RPA0114	hypothetical protein	pink
orthoMCL2407	RPA0396	Cfr family protein	pink
orthoMCL2400	RPA0408	conserved GTPase	pink
orthoMCL3614	RPA0561	hypothetical protein	pink
orthoMCL2277	RPA0592	putative glutamate N-acetyltransferase/amino-acid acetyltransferase	pink
orthoMCL2239	RPA0633	probable ribonuclease p protein component (protein c5)	pink
orthoMCL2217	RPA0703	conserved hypothetical protein	pink
orthoMCL2133	RPA0846	Fo ATP synthase subunit A	pink
orthoMCL4046	RPA1206	aldehyde dehydrogenase	pink
orthoMCL2991	RPA1788	possible 4-hydroxybenzoyl-CoA thioesterase	pink
orthoMCL0089	RPA1792	putative branched-chain amino acid transport system ATP-binding protein	pink
orthoMCL1561	RPA1981	ribose 5-phosphate isomerase	pink
orthoMCL1499	RPA2119	putative permease protein of ABC transporter	pink
orthoMCL1498	RPA2120	putative hemin binding protein	pink
orthoMCL1497	RPA2121	conserved unknown protein	pink
orthoMCL1495	RPA2123	conserved unknown protein	pink
orthoMCL1494	RPA2124	tonB dependent iron siderophore receptor	pink
orthoMCL1493	RPA2125	conserved unknown protein	pink
orthoMCL1492	RPA2126	conserved unknown protein	pink
orthoMCL1491	RPA2128	biopolymer transport protein ExbD/TolR	pink

orthoMCL1483	RPA2141	conserved hypothetical protein	pink
orthoMCL1474	RPA2156	hypothetical protein	pink
orthoMCL1396	RPA2470	Protein of unknown function, HesB/YadR/YfhF	pink
orthoMCL1266	RPA2729	antitermination factor, NusB	pink
orthoMCL2921	RPA2754	Polysaccharide deacetylase	pink
orthoMCL0075	RPA2792	putative component of pH adaptation K-efflux system phaE	pink
orthoMCL1140	RPA2935	Beta-lactamase-like	pink
orthoMCL3363	RPA3302	conserved hypothetical protein	pink
orthoMCL2858	RPA3465	putative long-chain-fatty-acid CoA ligase	pink
orthoMCL0864	RPA3474	putative 3-oxoacyl-acyl carrier protein reductase	pink
orthoMCL2856	RPA3479	2OG-Fe(II) oxygenase superfamily:Prolyl 4-hydroxylase, alpha subunit	pink
orthoMCL2855	RPA3480	putative outer membrane receptor for iron transport	pink
orthoMCL0780	RPA3625	Excinuclease ABC, C subunit, N-terminal	pink
orthoMCL0763	RPA3646	putative maltooligosyltrehalose trehalohydrolase	pink
orthoMCL2846	RPA3719	putative high-affinity branched-chain amino acid transport system ATP-binding protein	pink
orthoMCL2845	RPA3720	putative branched-chain amino acid transport system ATP-binding protein	pink
orthoMCL2833	RPA3786	unknown protein	pink
orthoMCL3831	RPA3840	putative hydroxamate-type ferrisiderophore receptor	pink
orthoMCL0517	RPA4152	periplasmic iron binding protein FbpA precursor	pink
orthoMCL0516	RPA4153	putative iron transport system permease protein	pink
orthoMCL0399	RPA4350	hypothetical protein	pink
orthoMCL3267	RPA4402	hypothetical protein	pink
orthoMCL0321	RPA4444	conserved hypothetical protein	pink
orthoMCL0280	RPA4568	putative enoyl-acyl carrier protein reductase	pink
orthoMCL2747	RPA4636	FeoA family	pink
orthoMCL2742	RPA4697	4-oxalomesaconate hydratase	pink
orthoMCL0159	RPA4771	possible heat shock protein HSP33	pink
orthoMCL2723	RPA4828	conserved hypothetical protein	pink
orthoMCL4665	NA	NA	red
orthoMCL4782	NA	NA	red
orthoMCL4124	RPA0027	2-dehydro-3-deoxyphosphoheptonate aldolase	red
orthoMCL2658	RPA0035	putative phenylalanine-tRNA ligase beta chain	red
orthoMCL2524	RPA0243	putative 16S rRNA processing protein.	red
orthoMCL3153	RPA0549	sensor protein ChrR, cytochrome cycA regulator	red
orthoMCL3152	RPA0550	RNA polymerase ECF-type sigma factor	red
orthoMCL2250	RPA0621	putative N-formylmethionylaminoacyl-tRNA deformylase	red
orthoMCL2187	RPA0756	putative amidase	red
orthoMCL2151	RPA0823	Apple domain:N/apple PAN	red
orthoMCL2098	RPA0892	glutamate synthase (NADPH) small chain	red
orthoMCL2066	RPA0944	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	red
orthoMCL3067	RPA1176	conserved hypothetical protein	red
orthoMCL1851	RPA1466	putative glutamyl-tRNA(Gln) amidotransferase subunit A	red
orthoMCL1541	RPA2017	putative lipid A biosynthesis lauroyl acyltransferase	red
orthoMCL1535	RPA2023	conserved hypothetical protein	red
orthoMCL5674	RPA2353	putative nitrogenase NifH subunit	red

orthoMCL5673	RPA2354	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	red
orthoMCL5671	RPA2357	cystathionine gamma-lyase	red
orthoMCL5669	RPA2359	putative periplasmic protein	red
orthoMCL5668	RPA2360	ABC transporter, ATP-binding protein	red
orthoMCL1330	RPA2596	D-alanine aminotransferase	red
orthoMCL1246	RPA2767	ribosomal protein L13	red
orthoMCL1215	RPA2829	conserved hypothetical protein	red
orthoMCL2908	RPA2844	SCP-like extracellular protein	red
orthoMCL1125	RPA2950	NADH-ubiquinone dehydrogenase chain C	red
orthoMCL1055	RPA3079	hypothetical protein	red
orthoMCL1054	RPA3080	putative 50S ribosomal protein L9, cultivar specific nodulation protein Csn1	red
orthoMCL1002	RPA3157	hypothetical protein	red
orthoMCL0965	RPA3229	Adenylate kinase	red
orthoMCL0964	RPA3230	secretion protein SecY	red
orthoMCL0963	RPA3231	50S ribosomal protein L15	red
orthoMCL0962	RPA3232	ribosomal protein L30	red
orthoMCL0961	RPA3233	ribosomal protein S5	red
orthoMCL0960	RPA3234	50S ribosomal protein L18	red
orthoMCL0959	RPA3235	50S ribosomal protein L6	red
orthoMCL0958	RPA3236	30S ribosomal protein S8	red
orthoMCL0957	RPA3237	30S ribosomal protein S14	red
orthoMCL0956	RPA3238	50S ribosomal protein L5	red
orthoMCL0955	RPA3239	50S ribosomal protein L24	red
orthoMCL0954	RPA3240	50S ribosomal protein L14	red
orthoMCL0952	RPA3242	50S ribosomal protein L29	red
orthoMCL0951	RPA3243	50S ribosomal protein L16	red
orthoMCL0949	RPA3245	50S ribosomal protein L22	red
orthoMCL0948	RPA3246	30S ribosomal protein S19	red
orthoMCL0947	RPA3247	50S ribosomal protein L2	red
orthoMCL0946	RPA3248	50S ribosomal protein L23	red
orthoMCL0945	RPA3249	50S ribosomal protein L4	red
orthoMCL0944	RPA3250	50S ribosomal protein L3	red
orthoMCL0943	RPA3251	30S ribosomal protein S10	red
orthoMCL0014	RPA3252	elongation factor Tu	red
orthoMCL3368	RPA3258	conserved hypothetical protein	red
orthoMCL0934	RPA3270	50S ribosomal protein L10	red
orthoMCL0762	RPA3647	putative glycosyl hydrolase	red
orthoMCL0757	RPA3654	putative molybdopterin-guanine dinucleotide biosynthesis protein A	red
orthoMCL0656	RPA3807	putative permease of ABC transporter (high-affinity branched-chain amino acid transport)	red
orthoMCL0478	RPA4215	putative siroheme synthase	red
orthoMCL0450	RPA4250	nitrogen fixation transcriptional regulator fixK2, Crp/Fnr family	red
orthoMCL0448	RPA4252	NADH-ubiquinone dehydrogenase chain N	red
orthoMCL3276	RPA4290	conserved hypothetical protein	red
orthoMCL3795	RPA4477	Possible membrane protein with ATP/GTP-binding site motif A (P-loop)	red

orthoMCL0269	RPA4605	electron trnasfer flavoprotein beta chain fixA	red
orthoMCL0264	RPA4610	Protein of unknown function, HesB/YadR/YfhF	red
orthoMCL0250	RPA4625	NifZ domain	red
orthoMCL3245	RPA4626	Protein of unknown function from Deinococcus and Synechococcus	red
orthoMCL5491	RPA4795	conserved hypothetical protein	red
orthoMCL0122	RPA4827	conserved hypothetical protein	red
orthoMCL3704	NA	NA	turquoise
orthoMCL3730	NA	NA	turquoise
orthoMCL3185	RPA0087	putative Major Facilitator Superfamily (MFS) transporter	turquoise
orthoMCL2604	RPA0149	possible ABC-type iron-siderophore transport system ATP-binding protein	turquoise
orthoMCL2541	RPA0220	putative cation efflux system protein	turquoise
orthoMCL2421	RPA0370	putative diguanylate cyclase (GGDEF)/phosphodiesterase (EAL) with PAS domains	turquoise
orthoMCL3626	RPA0458	possible fatty acid-CoA ligases.	turquoise
orthoMCL3621	RPA0514	putative efflux pump protein FarB	turquoise
orthoMCL3620	RPA0515	possible efflux pump protein FarA	turquoise
orthoMCL3615	RPA0560	malonyl-CoA decarboxylase	turquoise
orthoMCL2242	RPA0630	DUF423	turquoise
orthoMCL3594	RPA0755	putative oligopeptide ABC transporter, ATP-binding component	turquoise
orthoMCL3115	RPA0765	putative outer membrane receptor for iron transport	turquoise
orthoMCL2178	RPA0785	possible cytochrome P450 family proteins	turquoise
orthoMCL2160	RPA0812	putative transmembrane protein	turquoise
orthoMCL2147	RPA0831	cytochrome c oxidase subunit II	turquoise
orthoMCL4067	RPA0985	putative branched-chain amino acid transport system substrate-binding protein	turquoise
orthoMCL4064	RPA0988	putative branched-chain amino acid ABC transporter, ATP-binding protein	turquoise
orthoMCL4061	RPA1007	possible 2,3-dihydroxyphenylpropionate 1,2-dioxygenase	turquoise
orthoMCL0049	RPA1017	Nitrogen fixation-related protein	turquoise
orthoMCL4059	RPA1059	probable outer membrane protein, TonB-dependent receptor	turquoise
orthoMCL4050	RPA1106	conserved hypothetical protein	turquoise
orthoMCL1938	RPA1168	molybdopterin converting factor, subunit 2	turquoise
orthoMCL1914	RPA1194	putative carboxymethylenebutenolidase	turquoise
orthoMCL1913	RPA1195	short-chain dehydrogenase	turquoise
orthoMCL3064	RPA1213	conserved unknown protein	turquoise
orthoMCL3063	RPA1214	putative ABC transporter ATP-binding protein	turquoise
orthoMCL3062	RPA1215	putative ABC transporter, permease protein	turquoise
orthoMCL3042	RPA1374	Sigma-54 dependent, Vanadium nitrogenase transcriptional regulator, VnfA	turquoise
orthoMCL4015	RPA1468	conserved hypothetical protein	turquoise
orthoMCL0092	RPA1487	conserved hypothetical protein	turquoise
orthoMCL4590	RPA1488	L-carnitine dehydratase/bile acid-inducible protein F	turquoise
orthoMCL3525	RPA1490	Bacteriophytochrome (light-regulated signal transduction histidine kinase), PhyB4	turquoise
orthoMCL4589	RPA1515	conserved unknown protein	turquoise
orthoMCL1703	RPA1703	putative acetyl-CoA acyltransferase	turquoise
orthoMCL3516	RPA1706	putative enoyl-CoA hydratase	turquoise
orthoMCL1677	RPA1763	putative long-chain-fatty-acid CoA ligase	turquoise

orthoMCL2988	RPA1793	branched-chain amino acid transport system permease protein	turquoise
orthoMCL3992	RPA1796	putative Adenylate/Guanylate cyclase	turquoise
orthoMCL3501	RPA1845	putative tonB-dependent receptor protein	turquoise
orthoMCL3987	RPA1846	unknown protein	turquoise
orthoMCL3984	RPA1870	possible transcriptional regulator, MarR/EmrR family	turquoise
orthoMCL4573	RPA1874	hypothetical protein	turquoise
orthoMCL4572	RPA1875	possible uncharacterized iron-regulated membrane protein	turquoise
orthoMCL3981	RPA1876	putative TonB-dependent iron siderophore receptor	turquoise
orthoMCL3492	RPA1908	hypothetical protein	turquoise
orthoMCL4568	RPA1931	methyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL1576	RPA1955	glutathione dependent formaldehyde dehydrogenase	turquoise
orthoMCL4567	RPA1957	alkanal monooxygenase (LuxA-like protein)	turquoise
orthoMCL1545	RPA2011	puative potassium uptake protein Kup	turquoise
orthoMCL4564	RPA2053	conserved hypothetical protein	turquoise
orthoMCL3469	RPA2113	possible nitrate transport system permease protein	turquoise
orthoMCL3468	RPA2114	putative nitrate transport system ATP-binding protein	turquoise
orthoMCL3467	RPA2115	putative cyanate lyase	turquoise
orthoMCL2958	RPA2116	hypothetical protein	turquoise
orthoMCL2957	RPA2117	putative flavodoxin	turquoise
orthoMCL1500	RPA2118	putative ATP-binding protein of ABC transporter	turquoise
orthoMCL4551	RPA2130	DUF81	turquoise
orthoMCL1486	RPA2138	putative acyl-CoA dehydrogenase	turquoise
orthoMCL1475	RPA2155	dihydroxy-acid dehydratase	turquoise
orthoMCL4541	RPA2296	conserved hypothetical protein	turquoise
orthoMCL4537	RPA2312	hypothetical protein	turquoise
orthoMCL3445	RPA2333	putative cation transport ATPase, possible copper transporter	turquoise
orthoMCL4533	RPA2334	unknown protein	turquoise
orthoMCL4532	RPA2335	unknown protein	turquoise
orthoMCL4531	RPA2336	unknown protein	turquoise
orthoMCL4530	RPA2337	hypothetical protein	turquoise
orthoMCL4529	RPA2338	unknown protein	turquoise
orthoMCL4528	RPA2339	transcriptional regulator, FUR family	turquoise
orthoMCL3939	RPA2378	putative tonB-dependent receptor protein	turquoise
orthoMCL3938	RPA2379	probable acetyltransferase	turquoise
orthoMCL3937	RPA2380	probable tonB dependent iron siderophore receptor	turquoise
orthoMCL3936	RPA2381	probable FecR, iron siderophore sensor protein	turquoise
orthoMCL3935	RPA2382	putative iron(III) ABC transporter, ATP-binding protein	turquoise
orthoMCL3933	RPA2384	putative ironIII transport permease protein	turquoise
orthoMCL3932	RPA2385	putative ABC transporter, periplasmic Fe+3 siderophore binding protein	turquoise
orthoMCL4522	RPA2386	conserved hypothetical protein	turquoise
orthoMCL3931	RPA2387	conserved hypothetical protein	turquoise
orthoMCL3930	RPA2388	possible acyl-CoA ligase for activation during siderophore synthesis	turquoise
orthoMCL3929	RPA2389	possible Rhizobactin siderophore biosynthesis protein RhsF	turquoise
orthoMCL3928	RPA2390	possible Rhizobactin siderophore biosynthesis protein rhbC	turquoise
orthoMCL3441	RPA2391	RNA polymerase ECF-type sigma factor, possible FecI	turquoise
orthoMCL1397	RPA2468	DUF59	turquoise

orthoMCL3917	RPA2469	putative lactoylglutathione lyase	turquoise
orthoMCL4492	RPA2704	possible Na+/? antiporter	turquoise
orthoMCL4490	RPA2709	possible FusE-MFP/HlyD family membrane fusion protein	turquoise
orthoMCL3402	RPA2755	possible DNA-binding stress protein	turquoise
orthoMCL2920	RPA2756	hypothetical protein	turquoise
orthoMCL1251	RPA2758	possible cytochrome c oxidase assembly protein	turquoise
orthoMCL2917	RPA2786	hypothetical protein	turquoise
orthoMCL3397	RPA2877	putative glycosyltransferase family protein	turquoise
orthoMCL3385	RPA3046	conserved hypothetical protein	turquoise
orthoMCL0953	RPA3241	30S ribosomal protein S17	turquoise
orthoMCL0940	RPA3255	30S ribosomal protein S12	turquoise
orthoMCL3369	RPA3257	probable transcriptional regulator, AraC family	turquoise
orthoMCL2875	RPA3263	conserved hypothetical protein	turquoise
orthoMCL4432	RPA3359	O-antigen polymerase	turquoise
orthoMCL3348	RPA3458	possible TrapT family, dctP subunit, C4-dicarboxylate periplasmic binding protein	turquoise
orthoMCL0855	RPA3500	fumarate hydratase C	turquoise
orthoMCL0848	RPA3508	ATP-binding component of ABC transporter	turquoise
orthoMCL0814	RPA3548	possible serine protease/outer membrane autotransporter	turquoise
orthoMCL0771	RPA3637	putative 6-phosphogluconolactonase	turquoise
orthoMCL0725	RPA3693	putative cytochrome c	turquoise
orthoMCL2843	RPA3722	putative branched-chain amino acid transport system permease protein	turquoise
orthoMCL3832	RPA3839	putative transcriptional regulator PchR, AraC family	turquoise
orthoMCL3830	RPA3846	GCN5-related N-acetyltransferase	turquoise
orthoMCL0513	RPA4156	3-oxoadipate CoA-transferase subunit A	turquoise
orthoMCL3803	RPA4165	conserved hypothetical protein	turquoise
orthoMCL4380	RPA4279	hypothetical protein	turquoise
orthoMCL0012	RPA4449	pmethyl-accepting chemotaxis receptor/sensory transducer	turquoise
orthoMCL2768	RPA4468	conserved unknown protein	turquoise
orthoMCL2767	RPA4469	conserved unknown protein	turquoise
orthoMCL3794	RPA4481	methyl-accepting chemotaxis sensory transducer	turquoise
orthoMCL0281	RPA4567	putative phosphate acetyltransferase	turquoise
orthoMCL0277	RPA4571	hypothetical protein	turquoise
orthoMCL0270	RPA4604	electron transfer flavoprotein alpha chain protein fixB	turquoise
orthoMCL0266	RPA4608	nitrogenase cofactor synthesis protein nifS	turquoise
orthoMCL0247	RPA4629	ferredoxin 2[4Fe-4S], fdxN	turquoise
orthoMCL0246	RPA4630	nitrogen fixation protein nifB	turquoise
orthoMCL2748	RPA4635	ferrous iron transport protein B	turquoise
orthoMCL2743	RPA4696	conserved unknown protein	turquoise
orthoMCL0167	RPA4760	unknown protein	turquoise
orthoMCL0137	RPA4793	cytochrome bd-quinol oxidase subunit I	turquoise
orthoMCL0136	RPA4794	putative cytochrome bd-quinol oxidase subunit II	turquoise
orthoMCL0117	RPA4836	30S ribosomal protein S20	turquoise
orthoMCL4158	NA	NA	yellow
orthoMCL2664	RPA0028	bifunctional purine biosynthesis protein	yellow
orthoMCL2657	RPA0037	phenylalanyl-tRNA synthetase, alpha-subunit	yellow

orthoMCL2656	RPA0038	ribosomal protein L20	yellow
orthoMCL2655	RPA0039	50S ribosomal protein L35	yellow
orthoMCL2650	RPA0044	bacitracin resistance protein	yellow
orthoMCL2535	RPA0227	beta-isopropylmalate dehydrogenase	yellow
orthoMCL2526	RPA0241	50s ribosomal protein L19	yellow
orthoMCL2522	RPA0245	signal recognition particle protein	yellow
orthoMCL2449	RPA0329	ribonuclease PH	yellow
orthoMCL2439	RPA0339	dihydrodipicolinate reductase	yellow
orthoMCL2410	RPA0392	argininosuccinate synthase	yellow
orthoMCL2409	RPA0393	PA-phosphatase related phosphoesterase	yellow
orthoMCL3167	RPA0395	Metal dependent phosphohydrolase	yellow
orthoMCL2406	RPA0397	conserved unknown protein	yellow
orthoMCL2374	RPA0439	DUF150	yellow
orthoMCL2337	RPA0493	50S ribosomal protein L28	yellow
orthoMCL2325	RPA0508	acetyl-CoA carboxylase carboxyltransferase alpha subunit	yellow
orthoMCL2292	RPA0566	conserved unknown protein	yellow
orthoMCL3116	RPA0762	possible oligopeptide ABC transporter, periplasmic binding protein component	yellow
orthoMCL4077	RPA0786	putative Adenylate/Guanylate cyclase	yellow
orthoMCL2116	RPA0867	Endoribonuclease L-PSP	yellow
orthoMCL3101	RPA0951	Nucleoside 2-deoxyribosyltransferase	yellow
orthoMCL3563	RPA1070	hypothetical protein	yellow
orthoMCL1998	RPA1092	Carboxymuconolactone decarboxylase	yellow
orthoMCL1993	RPA1097	DUF28	yellow
orthoMCL1963	RPA1141	chaperonin GroES1, cpn10	yellow
orthoMCL1869	RPA1278	GatB/Yqey	yellow
orthoMCL4597	RPA1415	possible branched-chain amino acid transport system substrate-binding protein	yellow
orthoMCL1790	RPA1560	ribulose-bisphosphate carboxylase small chain	yellow
orthoMCL1782	RPA1577	hypothetical protein	yellow
orthoMCL1774	RPA1589	30S ribosomal protein S4	yellow
orthoMCL1726	RPA1658	putative 6-pyruvol tetrahydrobiopterin synthase	yellow
orthoMCL1702	RPA1704	probable transcriptional regulator, TetR family	yellow
orthoMCL1446	RPA2200	inosine monophosphate dehydrogenase	yellow
orthoMCL1444	RPA2202	SUN-family protein, putative RNA methyltransferase	yellow
orthoMCL1430	RPA2285	peptide chain release factor 3	yellow
orthoMCL1410	RPA2453	translation peptide releasing factor RF-2	yellow
orthoMCL1405	RPA2460	tyrosyl-tRNA synthetase	yellow
orthoMCL1359	RPA2557	Glycosyl transferase, family 2	yellow
orthoMCL3908	RPA2643	putative ABC transporter ATP-binding protein	yellow
orthoMCL3907	RPA2645	putative ABC transporter permease protein	yellow
orthoMCL1279	RPA2703	DNA topoisomerase IV subunitA	yellow
orthoMCL1267	RPA2728	riboflavin synthase, beta chain	yellow
orthoMCL1163	RPA2906	glutamyl-tRNA synthetase	yellow
orthoMCL1151	RPA2920	uridylate kinase	yellow
orthoMCL1149	RPA2922	30S ribosomal protein S2	yellow
orthoMCL1117	RPA2962	putative trigger factor	yellow

orthoMCL1077	RPA3053	cold shock DNA binding protein	yellow
orthoMCL1076	RPA3056	nucleoside-diphosphate-kinase	yellow
orthoMCL1057	RPA3077	possible 30S ribosomal protein S6	yellow
orthoMCL1056	RPA3078	30S ribosomal protein S18	yellow
orthoMCL3343	RPA3476	possible energy transducer TonB	yellow
orthoMCL0790	RPA3609	ABC transporter, ATPase component	yellow
orthoMCL0712	RPA3726	conserved unknown protein	yellow
orthoMCL0703	RPA3742	putative poly-isoprenyl transferase	yellow
orthoMCL0642	RPA3833	tRNA/rRNA methyltransferase	yellow
orthoMCL2818	RPA3982	Glycosyl transferase, family 4	yellow
orthoMCL0558	RPA3985	ADP-L-glycero-D-mannoheptose-6-epimerase	yellow
orthoMCL0396	RPA4353	conserved hypothetical protein	yellow
orthoMCL0394	RPA4355	putative peptidyl-tRNA hydrolase	yellow
orthoMCL0393	RPA4356	putative 50S ribosomal protein L25	yellow
orthoMCL0325	RPA4440	putative cyclohexadienyl dehydrogenase	yellow
orthoMCL0302	RPA4501	phnA-like protein	yellow
orthoMCL3780	RPA4597	possible seryl-tRNA synthetase	yellow
orthoMCL3779	RPA4598	possible acyl-CoA dehydrogenase	yellow
orthoMCL0258	RPA4616	nitrogenase reductase-associated ferredoxin, nifN	yellow
orthoMCL0257	RPA4617	nitrogenase molybdenum-cofactor synthesis protein nifE	yellow
orthoMCL0256	RPA4618	nitrogenase molybdenum-iron protein beta chain, nifK	yellow
orthoMCL0253	RPA4621	conserved hypothetical protein	yellow
orthoMCL0252	RPA4623	conserved hypothetical protein	yellow
orthoMCL0251	RPA4624	hypothetical protein	yellow
orthoMCL3244	RPA4671	putative maleylacetoacetate isomerase	yellow
orthoMCL0218	RPA4686	possible ABC transporter, periplasmic amino acid-binding protein	yellow
orthoMCL0203	RPA4715	molybdate transport system ATP-binding protein	yellow
orthoMCL0202	RPA4716	molybdate transport system permease protein	yellow
orthoMCL0201	RPA4717	molybdate transport system substrate-binding protein	yellow
orthoMCL0157	RPA4773	putative acetylornithine aminotransferase	yellow
orthoMCL0128	RPA4815	heat shock protein HtpG	yellow

Module members are sorted by module colors and strain CGA009's gene numbering (RPA number).

APPENDIX

Contributions to other works performed during the course of doctoral training.

Chugani *et al.*, 2012. I assisted in the design of selective primers for *Pseudomonas aeruginosa* and processed RNA-seq data.

Hirakawa *et al.*, 2011. I processed RNA-seq data for the study.