Title
Engineering Escherichia coli to grow on methanol as the sole carbon source

Author
Chen, Yu-Hsiao

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Engineering *Escherichia coli* to grow on methanol as the sole carbon source

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemical Engineering

by

Yu-Hsiao Chen

2020

ABSTRACT OF THE DISSERTATION

Engineering *Escherichia coli* to grow on methanol

as the sole carbon source

by

Yu-Hsiao Chen

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2020

Professor James C. Liao, Chair

Electron-rich and potentially renewable, methanol is a promising alternative feedstock for chemical production. Methanol bioconversion further benefits by boasting mild reaction conditions and higher specificity. Though methylotrophs can utilize methanol natively, engineering on these organisms had been challenging due to limited genome editing tools, sensitivity and comparatively low growth. Thus, many has attempted to enable methanol utilization in industrial-friendly organisms. Unfortunately, converting a non-methylotrophic organism to a synthetic methylotroph that can grow to a high cell density has been strenuous, despite the fact that typical sugar metabolism, namely glycolysis (the Embden–Meyerhof–Parnas pathway), and the ribulose monophosphate (RuMP) cycle which methylotrophs use for methanol assimilation differs by only three enzymes.

Here, our first aim is to reprogram an *Escherichia coli* to be dependent on methanol for growth. We sought to design an *E. coli* that survives only when it is able to co-utilize methanol with other carbon sources. We selected to disrupt two genes from the pentose phosphate pathway, ribose-5-phosphate isomerase (*rpi*) and ribulose-phosphate 3-epimerase (*rpe*) to construct two different strains that can grow on methanol with xylose and ribose respectively. We then used adaptive evolution and eventually isolated a stable *ΔrpiAB* auxotroph that can co-utilize xylose with methanol in a 1 to 1 molar ratio and grow to OD_{600} 4.0 in less than 30 hours. This strain was also characterized by genomic sequencing and subsequently engineered to produce ethanol and n-butanol, demonstrating the usefulness of the strain.

Our next aim is establishing completely methylotrophy, and engineering an *E. coli* to become a synthetic methylotroph using methanol as the sole carbon source. We began from adding the *rpi* gene back into the auxotroph strain, and took advice from an in-house theoretical flux prediction program, ensemble modeling robustness analysis (EMRA), to determine two enzymes in glycolysis should be tuned down, namely phosphofructokinase (PFK) and glceraldehyde-3phosphte dehydrogenase (GAP). After extensive adaptive evolution, a strain achieved to utilize methanol as the sole carbon source, reaching an OD_{600} 2.0 in 30 hours. This synthetic methylotroph SM1 then revealed a major hurdle that was expected but underestimated: DNA-protein crosslinking (DPC). SM1 eventually overcome this formaldehyde-induced toxicity by metabolic flux balancing through mutations and insertion sequence (IS) mediated copy number variations (CNV). Not only SM1 can grow at comparable rates against natural methylotrophs, but also in a wide-range of different methanol concentrations. While demonstrating how rational-based genome editing and adaptive evolution can steer a major tropism shift, this synthetic methylotroph also expands the horizon of C1 bioconversion.

The dissertation of Yu-Hsiao Chen is approved.

Yvonne Y. Chen

Robert T. Clubb

Yi Tang

James C. Liao, Committee Chair

University of California, Los Angeles

2020

To my beloved family

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my utmost gratitude to my advisor, Dr. James C. Liao (now President Liao) for his exceptional mentoring throughout my PhD studies. He had truly inspired me and brought out my potentials, not to mention the countless discussion sessions beyond day and night, and his 24-hour-available direct hotline whenever we had exciting ideas to share with each other. His broad connections and open-mindedness always stimulate me to think out of the box beyond metabolic engineering and collaborate with experts in other fields and broaden my thoughts. His move from UCLA to Academia Sinica, Taiwan also granted me a rare opportunity to design a lab from scratch, which was also definitely an unforgettable fun experience.

I wish to thank the generous funding from DOE ARPA-E REMOTE project (Award DE-AR0000430), as well as funding from Academia Sinica, Taiwan. The renovation of our laboratory in UCLA was funded by the National Science Foundation under Grant No. 0963183, affiliated to the American Recovery and Reinvestment Act of 2009 (ARRA).

I greatly appreciated the professors and staffs who helped me during my PhD years as well. I would like to especially thank Dr. Meiyeh Lu, who greatly mentored me in using genome sequencing as a powerful diagnostic and discovery tool. I sometimes feel sorry that we are always overexcited in our projects and too immersed in our conversation that I make her husband wait for her downstairs quite often. Thanks to Dr. Shu-Jen Chou, Dr. Sue-Ping Lee, and Dr. Shu-Yu Lin, I got to learn Digital PCR, EM and proteomic techniques. Dr. Steven Lin was a great consultant regarding Crispr technology. Dr. Hsiao-Ching Lin helped me out setting up the lab in the early days of the Taiwan lab establishment. Dr. Mei-Ching assisted me in securing our patent and provided potential future directions of research for commercialization. Miguel and Jo in UCLA helped me a lot in administrative paperwork during my visit in Taiwan.

I am profoundly grateful to my lab members in UCLA. Igor is the mastermind who began the adventurous journey in methanol utilization. Tony is always the go-to person for everything you need to learn and know about the LA lab. Charlie is the organized protein and production expert that I envy. Po-Wei and Matthew helped me out understand EMRA and introduced me to the coding world. Shanshan's CO_2 fixation cycle and her multiple-enzyme assay techniques were stunning. I wished that I could work with her more on our solid-state idea. Discussions and chatting with Paul, Sammy, Alec, Hong, Xiaoqian, Fabienne, Lemuel, Luo and Derrick were always fun and inspiring. I owe a lot to my awesome undergrads, Allen and Aaron. It's hard to find someone who are both diligent and smart like them. Seems that they are shadowing my tortuous PhD path as well, and I wish them good luck (I hope that Aaron's advisor does not relocate during his PhD just like mine or Allen's, or I will be at blame for sure!)

My peers in Academia Sinica are also well-deserved to be credited. Thank you Arong (Hsin-Wei) for helping me out with all the experiments. I couldn't have finished those massive laborious work alone without you. Chao-yin is the most mellow guy I've ever met. He really helped me out when we were stuck with Crispr optimization. Plus, he's an expert in qRT-PCR! I still can't extract such pure RNA than he does. Kuan-Jen is cheerful, elegant, and always optimistic. Her green fingers are not only limited to passion in plants, but also extend to "fruitful" discussions. Willy is the witty guy who is deadpan outside, but funny inside. He always delivers great ideas, ramen and beer. Sandy is the talented artist hidden in our lab who can swiftly create cartoons upon request. I enjoy chatting about life and science with Sonya and Tim. Watching Irene and Jason roasting each other is an everyday comedy not to be missed. Yu-Chieh and Yi-Hsuan are great moodmakers that make everyone happy. Many thanks to May, Felix, Jack, Esther and Tiffany for keeping our lab functional and clean.

I am also crediting Yi-Yun, Marco, Hung-Yi, Chi-Jui, and Charles for their mental support and their professional advice in Organic chemistry, Coding, Modelling, Physical Chemistry, and Chemical Biology. I am truly blessed to have many friends working in different fields that I can consult to.

Another two friends that I definitely am in debt to are Candy and Meimei. They are the most understanding friends who know well the pains and difficulties a PhD student experience, and help dissipate them. The weekly petit-trip to the east craving for music, food and lots of laughter was one of my highlighted moments in LA. I hope that our qin-trio: pine, bamboo and plum friendship never ends.

The members of the Chemistry Fab four: Francis, Marco, Alex were invaluable to me as well. We paced our journey together towards PhD by sharing doctoral-level jokes, foolish consequences and mocking each other. Especially Francis, my best friend whose fate is almost synchronized with mine. We basically experienced the same events in our career path miraculously. Just to name a few: my advisor moving back to Taiwan while his relocated to Hong Kong just a year after, getting a real first author paper at the same time, planning out a press conference consecutively, you name it. Long live our friendship.

From the bottom of my heart, I would like to thank my family. Without them, I couldn't have made it so far. Thank you for being there during my darkest days. I deeply appreciated my mom devoting her whole in me with love, and my dad supportive of any decision I make. Aunt Juliana always prepared food for mw during weekends and Uncle Rex granted me a pair of feet in LA by sharing me his fancy Mercedes-Benz lease. My grandmother from my father's side, Huei-Mei is very caring and always very keen of my health status. I wish her good health. Special thanks to my sister, Alice who drew a cover art for my paper! Though it was not elected, I personally

think that it's the best of all. The scientist and the designer also perceive things very differently, which inevitably invokes numerous quarrels. Nevertheless, they always reconcile easily and cherish every time spent together. Not to mention most of her brother's profile photo was taken by her hands.

Last of all, I'd like to acknowledge my grandparents who are no longer with us. My grandmother from my mother's side, Li-Chun went to a better place before I began my PhD studies, and my grandfather from my father's side, Tsan-Sung passed away during my PhD studies. Having a deep affection with them, I still regret that I couldn't show them my success in person. It was them, along with my family who have made me who I am today. I believe that they are still watching, and hope that they would be very proud of their grandson. May them rest in peace.

Yu-Hsiao Chen

Education

University of California, Los Angeles, USA **2014-2020**
Ph.D. candidate in Chemical Engineering

National Taiwan University, Taiwan **2008-2013**
B.S, Chemistry,
B.S, Biochemical Science and Technology [Double Major]

Fields of Interests

Synthetic Biology, Metabolic Engineering, Genome Editing, Next Generation Sequencing

Research Experience

Visiting Researcher, 2016-2020
President James Liao, Academia Sinica, Institute of Biological Chemistry, Taiwan

- Established an *E. coli* that can grow and utilize methanol as the sole carbon source by implementing theoretical calculations results and adaptive evolution.
- Developed and Optimized Crispr-based genome editing tools in *E. coli*.
- Employed digital PCR, 2nd and 3rd Generation Sequencing methods (Illumina, Pacbio and Nanopore ONT) to discover and depict copy number variations in prokaryotes.
- Visualized and characterized DNA-Protein crosslinking toxicity by Transmission Electron Microscopy and proteomics.

Doctoral Student, 2014-2020
Dr. James Liao, UCLA, Department of Chemical Engineering

- Constructed a methanol-dependent growing *E. coli* strain by performing gene knockouts and directed evolution.
- Built and optimized production pathways of ethanol and n-butanol in *E. coli* with methanol and xylose as substrates.
- Designed a solid-phase based enzymatic CO_2 fixation module.

Research Assistant, 2013-2014
Dr. Steve S.-F. Yu & Dr. Sunney I. Chan, Academia Sinica, Institute of Chemistry, Taiwan

- Designed FRET and fluorescence quenching experiments to locate copper ion sites in particulate methane monooxygenases (pMMO).
- Used ICP-MS to quantify metal ions in metalloproteins, such as pMMO.

Undergraduate research, exchange student, 2012-2013
Dr. Hiroshi Sugiyama, Department of Chemistry, Kyoto University, Japan

- Studied conformation change between B-Z transitions of artificially synthesized DNA by Circular Dichroism spectroscopy.

Undergraduate research, 2010-2012
Dr. Sunney I. Chan, National Taiwan University & Academia Sinica, Taiwan,

- Used site directed mutagenesis to perform biorthogonal chemistry to attach fluorophores on particulate methane monooxygenases.

Publications

- **Frederic Y.-H. Chen***, Hsin-Wei Jung, Chao-yin Tsuei, James C. Liao (2020) Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol. *Cell*, in press.

- Chang-Ting Chen*, **Frederic Y.-H. Chen***, Igor W. Bogorad*, Tung-Yun Wu, Jason Derks, Ruoxi Zhang, Abraxa S. Lee, Matthew E. Jones, James C. Liao. (2018) Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production. *Metabolic Engineering*, *49*, 257–266. ***Contributed equally to this work**.

- **Frederic Y.-H. Chen**, Soyoung Park, Haruka Otomo, Sohei Sakashita & Hiroshi Sugiyama. Investigation of B-Z transitions with DNA oligonucleotides containing 8-methylguanine (2014). *Artificial DNA: PNA & XNA*, *5*(1), e28226.

- Sammy Pontrelli, Tsan-Yu Chiu, Ethan I. Lan, **Frederic Y.-H. Chen**, Peiching Chang, James C. Liao, (2018) *Escherichia coli* as a host for metabolic engineering. *Metabolic Engineering*, *50*, 16–46.

- Lemuel M. J. Soh, Wai Shun Mak, Paul P. Lin, Luo Mi, **Frederic Y.-H. Chen**, Robert Damoiseaux, Justin B. Siegel, and James C. Liao (2017). Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design. *ACS Synthetic Biology*, *6*(4), 610–618.

Patents

- **Frederic Y.-H. Chen***, Hsin-Wei Jung, Chao-yin Tsuei, James C. Liao (2020) Metabolically-modified microorganisms that can grow on an organic C1 carbon source. Provisional application.

Honors and Awards

- **Scholarships for Excellent Students to Study Abroad, Taiwan (2012)**
  For attending exchange or joint degree program in outstanding overseas universities, sponsored by the Ministry of Education, Taiwan
- **Undergraduate Research Assistant Funding (2012)**
  For participation in undergraduate research projects provided by Academia Sinica
- **Government scholarship to study abroad by the Ministry of Education, Taiwan (2014-2016)**
- **Invited speaker at 14th Asian Congress on Biotechnology (2019)**

Chapter 1 : Introduction

1.1 Background

C1 compounds, which refers to molecules consisting of only one carbon atom, have been of great interest to scientists (Bogorad et al., 2014; Chen et al., 2018; Cotton et al., 2019; Gassler et al., 2020; Gleizer et al., 2019; Mesters, 2016). Methanol is a C1 compound that is a common intermediate for chemical and bioconversion step towards carbon-carbon bond formation for other more inert C1 compounds, such as methane and CO_2 (Conrado and Gonzalez, 2014; Kuk et al., 2017; Patel et al., 2019). Trends of increasing natural gas production, composed of mainly methane, and elevated emissions of CO_2 indicate that C1 compounds could be an attractive carbon source for chemical feedstock (Bertau et al., 2014). Moreover, as global warming aggravates, it is necessary to capture or reduce the amount greenhouse gases in the atmosphere such as methane and CO_2 (Smith et al., 2013). Given all the above reasons, researchers have put effort into the development of C1 carbon conversion and C1 chemistry.

However, the bottleneck of C1 reactions lie in the notoriously high activation energy of C1 compounds, including C-H bond of methane at 439 KJ/mol, and the inert C=O bond of CO_2 at 532 KJ/mol. Thus, conventionally, C1 reactions either requires the development of expensive heavy-metal catalysts or reactions at high temperature and high pressures, such as syngas production and Fischer-Tropsch reactions. Both approaches are neither considered environmentally friendly nor economically sustainable (Wood et al., 2012). Moreover, the difficulty to fine-tune oxidation states of the C1 compounds in chemical reactions also causes a major drawback by producing various side products with limited product yields. In contrast, using biological systems such as enzymes enables the entire reaction to be carried out under ambient conditions and also assuring high product specificity.

Hereby, this thesis aims at developing a synthetic organism to perform methanol assimilation and bioconversion efficiently, as this C1 compound has an intermediate oxidation state and can be potentially converted from the two C1 greenhouse gases: methane and CO_2 renewably. Specifically, we picture a scenario of developing a fast-growing *E. coli* to grow by utilizing methanol as the sole carbon source, and thus, enabling it to serve as a biocatalyst that converts methanol to other higher carbon compounds with high efficiency.

1.2 Significance

C1 compounds are closely related to energy infrastructures, making C1 research extremely important. Methane in natural gas represents a common energy feedstock, while CO_2 is the ultimate product after combustion of fuels and coals for providing energy. As petroleum reserves continue to deplete, natural gas production has spurred recently with advanced technologies such as hydraulic fracturing processes and shale gas exploitation. According to the US Energy Information Administration, worldwide gross natural gas production reached a whopping 152 trillion ft^3 in 2015, while 32.303 billion tons of CO_2 was emitted at the same year. Moreover, natural gas resources in the US are estimated at 2×10^3 trillion ft^3 out of 7×10^3 trillion ft^3 globally, based on current drilling and extraction technology (Blohm et al., 2012). These reserves would be able to supply US for 100 years at the current usage rate. Hence, the high availability has made natural gas economically attractive to extend its applications.

Thanks to the increasing availability of methane, the capacity of methanol production also boosted as they are produced through traditional chemical processing from methane. This also allows methanol to be a feasible material for chemical production, given their abundance and low

cost. Moreover, replacing glucose with methanol as an alternative feedstock for bioconversion is also preferred to avoid competition with the food supply chain.

The rationale behind opting to establish methanol bioconversion are briefly explained previously: reactions at ambient temperature and high product specificity. Biology is capable of converting low grade or low concentration of methanol as well.

We also opted to establish a synthetic organism to assimilate methanol instead of using native methylotrophs. There are already several reports in literature found to be able to perform methanol bioconversion by methylotrophic bacteria, for instance, producing polysaccharide in *Methylocystis parvus* (Hou et al., 1979), L-serine in *Methylobacterium sp.* (Hagishita et al., 1996), L-lysine in *Methylobacillus glycogenes* for (Motoyama et al., 2001) and green fluorescent protein in *Methylobacterium extorquens*(Bélanger et al., 2004). Nevertheless, research in methanotrophs encounters several bottlenecks, including ill-studied cell physiology, limited genetic tools and the slow growth rate of cells. Thus, we anticipated to engineer an industrial-friendly *Escherichia coli* to utilize methanol, as its DNA manipulation tools are well established and has a decent growth rate.

There is already literature reporting that incorporation of methanol into central metabolites in engineered E. coli is possible (Müller et al., 2015). In addition, there are alos reports mentioning *E. coli* could produce bioproducts such as nargenine by methanol and other substrates present in rich medium (Whitaker et al., 2016). Most research groups opted to incorporate the Ribulose monophosphate (RuMP) cycle, which includes three heterologous genes of *mdh* (methanol dehydrogenase), *hps* (2-hexulose-6-phosphate synthase) and *phi* (6-phospho-3-hexuloisomerase), into *E. coli* (Müller et al., 2015; Price et al., 2016; Whitaker et al., 2016). Theoretically, expressing these three heterologous genes should enable the strain to grow on methanol already. However,

the simple overexpression approach has shown only little amount of methanol consumption. Cells were unable to utilize methanol as the sole carbon source and still required other rich nutrients to grow, such as yeast extract.

Though asking why non-methylotrophic strains cannot utilize methanol even expressing their required genes may seem philosophical, it is intriguing and meaningful to decipher what kind of metabolic kinetic traps and regulation an organism needs to overcome to utilize a non-native substrate, especially one that is relatively chemically-inactive. Combing laboratory evolution, this initiative brings opportunity to observe how biology evolves to adapt in order to survive.

1.3 Challenges

There are several challenges in engineering a non-methanotroph, like *E. coli* to assimilate methanol efficiently. Firstly, oxidation of methanol to formaldehyde is a tough process in biological systems. Either unique cofactors such as pyrroloquinoline quinone (PQQ) or tetrahydromethanopterin that are not present in *E. coli* are required, or the reaction is thermodynamically unfavorable, in the case of utilizing the NAD-methanol dehydrogenase. In the latter case, optimal expression is required for the cells to ensure sufficient inward methanol flux, but not over-exhausting the strain resources reserved for protein expression.

Secondly, formaldehyde is a toxic molecule as it spontaneously crosslinks with proteins and nucleotides causing universal cell damage(Bolt, 1987; Chaw et al., 1980; Heck et al., 1990). Thus, once formaldehyde is produced from methanol oxidation, it must be rapidly metabolized and maintained at low concentration, presumably micromolar ranges to prevent toxicity hampering cell growth.

Last but not least, all methylotrophic pathways require metabolic recycling, which is a complex process. Specifically, various metabolites must react with formaldehyde to assimilate them into biomass, and the intermediates must be recycled while some carbons serve as intermediates and products. Poorly controlled or unbalanced ratio of metabolite recycling will lead to formation of kinetic traps, which will ultimately drain the cycle's intermediates.

1.4 Proposal & Solution

Hence, to resolve these difficulties, engineering an efficient methanol assimilating, highly self-regulated and formaldehyde tolerant *E. coli* would be the key to develop a cell-based biocatalyst for methanol conversion which can be executed under mild physiological conditions. To execute the task, we first need to create a platform that can optimize methanol assimilation. As mentioned above, previous studies have already shown that researchers have successfully engineered *E. coli* to uptake methanol and even produce chemicals with methanol. However, the methanol consumption rate is relatively low and still requires rich media for growth, which is not economically viable(Müller et al., 2015; Price et al., 2016; Whitaker et al., 2016).

Chapter 2 describes the efforts on converting an industrial-friendly organism *E. coli* to be growth-dependent on methanol. That is, the strain requires methanol for growth when grown on a specific carbon source. The strain would be a "synthetic methanol auxotroph" since methanol becomes an essential nutrient, creating a growth-based selection platform for us to employ laboratory evolution and improve methanol utilization in *E. coli* by optimizing the flux between methanol assimilation pathways and native ones. As strain engineering will be mainly based on expressing or rewiring the Ribulose monophosphate (RuMP) cycle, we sought to construct two strains, Δ*rpe* and Δ*rpiAB* for the development of methanol auxotrophy, with a growth selection

platform and adaptive evolution to enable the cells to fine-tune their metabolic activity by themselves. After the cell completes its task with utilizing methanol along with the carbon source, we first characterized the strain by genome sequencing, where several crucial mutations were revealed, such as the deletion of the formaldehyde detoxification operon (*frmABR*) and adenylate cyclase (cyaA) that tunes the cyclic-AMP concentration in *E. coli*. We then further engineered this strain to produce chemicals including ethanol and n-butanol from methanol and the co-substrate, xylose. The final titers achieved were 4.6 g/L and 2.0 g/L, respectively.

Chapter 3 depicts the success in developing an *E. coli* strain that is capable of utilizing methanol as the sole carbon source. After the cell completes its task with utilizing methanol along with the co-substrate carbon source mentioned in Chapter 2, we then shift heads towards the ultimate goal of complete methylotrophy. The next step proposed is to reconstruct a methanol auxotroph in a *E. coli* K-12 strain BW25113, as K strains have success rates in genome editing. The strain's required *rpi* to complete the entire RuMP cycle was added back along with several genes modified according to a theoretical calculation tool, Ensemble Modeling Robustness analysis (EMRA). We then characterized a discovered phenomenon, DNA-protein crosslinking by Transmission electron microscopy and proteomics, and analyzed the genome structure by whole genome sequencing and also as well. Out of expectation, we also discovered a phenomenon that is uncommon in prokaryotes: copy number variation. Ultimately, the synthetic methylotroph managed to grow efficiently with the help of dynamic copy number tuning and flux balancing through genetic mutations.

1.5 References

Bélanger, L., Figueira, M.M., Bourque, D., Morel, L., Béland, M., Laramée, L., Groleau, D., and Míguez, C.B. (2004). Production of heterologous protein by Methylobacterium extorquens in high cell density fermentation. FEMS Microbiology Letters 231, 197–204.

Blohm, A., Peichel, J., Smith, C., and Kougentakis, A. (2012). The significance of regulation and land use patterns on natural gas resource estimates in the Marcellus shale. Energy Policy 50, 358–369.

Bogorad, I.W., Chen, C.-T., Theisen, M.K., Wu, T.-Y., Schlenz, A.R., Lam, A.T., and Liao, J.C. (2014). Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. Proceedings of the National Academy of Sciences *111*, 15928–15933.

Bolt, H.M. (1987). Experimental toxicology of formaldehyde. J Cancer Res Clin Oncol 113, 305–309.

Chaw, Y.F., Crane, L.E., Lange, P., and Shapiro, R. (1980). Isolation and identification of cross-links from formaldehyde-treated nucleic acids. Biochemistry 19, 5525–5531.

Hagishita, T., Yoshida, T., Izumi, Y., and Mitsunaga, T. (1996). Efficient L-serine production from methanol and glycine by resting cells of Methylobacterium sp. strain MN43. Bioscience, Biotechnology, and Biochemistry 60, 1604–1607.

Chen, C.-T., Chen, F.Y.-H., Bogorad, I.W., Wu, T.-Y., Zhang, R., Lee, A.S., and Liao, J.C. (2018). Synthetic methanol auxotrophy of *Escherichia coli* for methanol-dependent growth and production. Metabolic Engineering *49*, 257–266.

Cotton, C.A., Claassens, N.J., Benito-Vaquerizo, S., and Bar-Even, A. (2019). Renewable methanol and formate as microbial feedstocks. Current Opinion in Biotechnology *62*, 168–180.

Conrado, R.J., and Gonzalez, R. (2014). Envisioning the bioconversion of methane to liquid fuels. Science *343*, 621-623.

Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D., and Steiger, M.G. (2020). The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO2. Nature Biotechnology 1–16.

Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., et al. (2019). Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO2. Cell *179*, 1255–1263.e12.

Heck, H.D., Casanova, M., and Starr, T.B. (1990). Formaldehyde toxicity--new understanding. Crit Rev Toxicol 20, 397–426.

Hou, C.T., Laskin, A.I., and Patel, R.N. (1979). Growth and Polysaccharide Production by Methylocystis parvus OBBP on Methanol. Applied and Environmental Microbiology 37, 800–804.

Kuk, S.K., Singh, R.K., Nam, D.H., Singh, R., Lee, J.-K., and Park, C.B. (2017). Photoelectrochemical Reduction of Carbon Dioxide to Methanol through a Highly Efficient Enzyme Cascade. Angewandte Chemie *129*, 3885–3890.

Mesters, C. (2016). A Selection of Recent Advances in C1 Chemistry. Annu Rev Chem Biomol Eng 7, 223–238.

Motoyama, H., Yano, H., Terasaki, Y., and Anazawa, H. (2001). Overproduction of L-Lysine from methanol by Methylobacillus glycogenes derivatives carrying a plasmid with a mutated dapA gene. Applied and Environmental Microbiology 67, 3064–3070.

Müller, J.E.N., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W.J., Wendisch, V.F., Brautaset, T., Portais, J.-C., et al. (2015). Engineering Escherichia coli for methanol conversion. Metabolic Engineering 28, 190–201.

Patel, S.K.S., Jeon, M.S., Gupta, R.K., Jeon, Y., Kalia, V.C., Kim, S.C., Cho, B.K., Kim, D.R., and Lee, J.-K. (2019). Hierarchical Macroporous Particles for Efficient Whole-Cell Immobilization: Application in Bioconversion of Greenhouse Gases to Methanol. ACS Appl Mater Interfaces *11*, 18968–18977.

Price, J.V., Chen, L., Whitaker, W.B., Papoutsakis, E., and Chen, W. (2016). Scaffoldless engineered enzyme assembly for enhanced methanol utilization. Proceedings of the National Academy of Sciences 113, 12691–12696.

Smith, K.R., Desai, M.A., Rogers, J.V., and Houghton, R.A. (2013). Joint CO2 and CH4 accountability for global warming. Proc. Natl. Acad. Sci. U.S.A. *110*, E2865–E2874.

Whitaker, W.B., Jones, J.A., Bennett, R.K., Gonzalez, J.E., Vernacchio, V.R., Collins, S.M., Palmer, M.A., Schmidt, S., Antoniewicz, M.R., Koffas, M.A., et al. (2016). Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. Metabolic Engineering.

Wood, D.A., Nwaoha, C., and Towler, B.F. (2012). Gas-to-liquids (GTL): A review of an industry offering several routes for monetizing natural gas. Journal of Natural Gas Science and Engineering 9, 196–208.

Zhang, W., Zhang, T., Wu, S., Wu, M., Xin, F., Dong, W., Ma, J., Zhang, M., Jiang, M., 2017. Guidance for engineering of synthetic methylotrophy based on methanol metabolism in methylotrophy. RSC Adv. 7, 4083–4091. https://doi.org/10.1039/ C6RA27038G.

This page is intentionally left blank.

Chapter 2 : Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production

Disclaimer: This chapter is edited from a published article with the same title in *Metabolic Engineering* **49** (2018) 257-266. The figures and tables have been renumbered for ease of reading and compliance of the UCLA thesis format requirements.

Author Contributions:

This publication had equal 1st co-authorship between Yu-Hsiao Chen, Dr. Igor W. Bogorad, and Dr. Chang-Ting Chen. Dr. Igor Bogorad developed the idea of methanol auxotrophy, and achieved methanol dependent growth using *Δrpe* strains. Yu-Hsiao Chen evolved *ΔrpiAB E. coli* to realize a methanol dependent strain without loopholes. Dr. Chang-Ting Chen characterized the genotype of evolved *ΔrpiAB* strains and demonstrated ethanol and 1-butanol production in both *ΔrpiAB* and *Δrpe* strains. Dr. Tung-Yun Wu processed the next-generation sequencing data of evolved *ΔrpiAB* strains. Jason Derks, Ruoxi Zhang, Abraxa S. Lee and Matthew Jones worked on strain phenotype characterization and production optimization.

2.1 Abstract

Methanol is a potentially attractive substrate for bioproduction of chemicals because of the abundance of natural gas and biogas-derived methane. To move towards utilizing methanol as a sole carbon source, here we engineer an *Escherichia coli* strain to couple methanol utilization with growth on five-carbon (C5) sugars. By deleting essential genes in the pentose phosphate pathway for pentose utilization and expressing heterologous enzymes from the ribulose-monophosphate (RuMP) pathway, we constructed a strain that cannot grow on xylose or ribose minimal media unless methanol is utilized, creating a phenotype termed "synthetic methanol auxotrophy". Our best strains were able to utilize methanol for growth to an OD_{600} of 4.0 in 30 hrs with methanol and xylose co-assimilation at a molar ratio of approximately 1:1. Genome sequencing and reversion of mutations indicated that mutations on genes encoding for adenylate cyclase (*cyaA*) and the formaldehyde detoxification operon (*frmRAB*) were necessary for the growth phenotype. The methanol auxotrophic strain was further engineered to produce ethanol or 1-butanol to final titers of 4.6 g/L and 2.0 g/L, respectively. ^{13}C tracing showed that 43% and 71% of ethanol and 1-butanol produced had labeled carbon derived from methanol, respectively.

2.2 Introduction

Methanol is an intermediate in methane utilization, thus a potentially abundant feedstock that can be derived from natural gas or anaerobic digestion. Methanol is highly reduced and can provide three "H_2" equivalents per carbon. As such, it is desirable to explore possibilities of utilizing methanol in microbial production of chemicals and fuel (Clomburg et al., 2017; Conrado and Gonzalez, 2014; Haynes and Gonzalez, 2014). While natural methylotrophs can utilize methanol, the genetic tools for engineering these organisms are not as mature as those for E. coli (Bennett et al., 2018; Zhang et al., 2017). Thus, it will be advantageous if a readily editable organism such as *E. coli* can be engineered to utilize methanol for production and growth.

At least three natural pathways for methanol assimilation have been discovered up to date: the ribulose-monophosphate (RuMP) pathway, the xylulose monophosphate (XuMP) pathway, and the serine pathway (Zhang et al., 2017). In addition, a synthetic methanol condensation Cycle (MCC) has been demonstrated (Bogorad et al., 2014). These pathways all assume a cyclic configuration to assimilate methanol and generate an output while recycling intermediate metabolites. The RuMP pathway shares many enzymes with the pentose phosphate pathway (PPP), and is our choice here for integration with *E. coli* metabolism. The only three unique enzymes are methanol dehydrogenase (Mdh) which oxidizes methanol to formaldehyde, 3-hexulose-6-phosphate synthase (Hps), which ligates formaldehyde with Ru5P to form hexulose 6-phosphate (H6P), and 6-phospho-3-hexuloisomerase (Phi), which isomerizes H6P to fructose 6-phosphate (F6P).

However, there are several challenges in engineering methanol assimilation into *E. coli*. First, oxidation of methanol to formaldehyde requires either the thermodynamically unfavorable NAD-linked reaction catalyzed by the NAD-dependent methanol dehydrogenase (MDH),

pyrroloquinoline quinone (PQQ)-dependent MDH, or oxygen-dependent methanol oxidase. The latter two cannot recycle electrons easily (Whitaker et al., 2015; Zhang et al., 2017). Second, formaldehyde is extremely reactive and spontaneously reacts with proteins, causing nonspecific cell damage (Gonzalez et al., 2006). Once methanol is oxidized, the formaldehyde produced must be rapidly consumed by reacting with an acceptor metabolite to prevent toxicity. As such, regeneration of the acceptor metabolite must be efficient. Third, since methanol is a non-native and toxic substrate for *E. coli*, modification of cell regulation might be required for the strains to adapt to methanol uptake.

Methanol incorporation to biomass has been demonstrated in *E. coli* (Bennett et al., 2017; Gonzalez et al., 2017; Müller et al., 2015; Whitaker et al., 2016), as well as *Saccharomyces cerevisiae* (Dai et al., 2017), *Corynebacterium glutamicum* (Leßmeier et al., 2015; Witthoff et al., 2015), and *Pseudomonas putida* (Koopman et al., 2009). For *E. coli*, overexpression of RuMP enzymes such as Mdh, Hps, and Phi for methanol assimilation has been explored (Bennett et al., 2017; Gonzalez et al., 2017; Müller et al., 2015; Price et al., 2016; Whitaker et al., 2016). Recently, several new strategies to improve methanol utilization had been reported (Bennett et al., 2017; Gonzalez et al., 2017). Phosphoglucose isomerase (Pgi) deficient *E. coli* was demonstrated with enhanced methanol co-utilization with glucose (Bennett et al., 2017). Deletion of leucine-responsive regulatory protein was shown to promote methanol incorporation to biomass when provided with yeast extract (Gonzalez et al., 2017). Importantly, reports have been published demonstrating methanol dependent production in *E. coli* to specialty (Whitaker et al., 2017) and commodity chemicals (Bennett et al., 2017). However, the production remained to be improved due to the requirement of supplying yeast extract as a feedstock, and low percentage of methanol derived carbon incorporated in the product.

To further improve methanol utilization and methanol dependent production, we designed methanol assimilating *E. coli* strains that couple methanol uptake to growth with a pentose co-substrate. The strain cannot grow on either xylose or ribose minimal media unless methanol is being incorporated (Fig. 2-1). The created phenotype is termed "synthetic methanol auxotrophy". This strain enables the use of evolutionary strategies for methanol metabolism.

During preparation of this manuscript, a report was published using deletion of *rpiAB* and *edd* to develop a methanol essential *E. coli* strain which co-utilized methanol and gluconate for growth (Meyer et al., 2018). Despite the similarity in the strategy in developing methanol co-utilization in *E. coli*, our approach provides the advantage of utilizing a cheaper and more accessible co-substrate (i.e. xylose). Furthermore, in the co-utilization of methanol and gluconate, one carbon is lost in converting gluconate to Ru5P as the form of CO_2. On the other hand, another report (He et al., 2018) described an engineered strain that was able to assimilate xylose and formaldehyde generated from sarcosine for growth. Its key design was *tktAB* deletion, which prevents xylose utilization and therefore rewires the carbon flux for C1 assimilation. However, further engineered strain that grows on methanol and xylose was not demonstrated. Also, the strain was unable to synthesize erythrose 4-phosphate due to *tktAB* deletion. Therefore, it required extra supplement of cellular components that are synthesized from erythrose 4-phosphate.

In this report, we developed a strain that can grow on methanol and xylose as the only carbon sources in minimal medium. We also achieved conversion of methanol to ethanol and 1-butanol and determined the fractional contribution from methanol using ^{13}C labeling, thus demonstrating production of useful compounds by these auxotrophic strains.

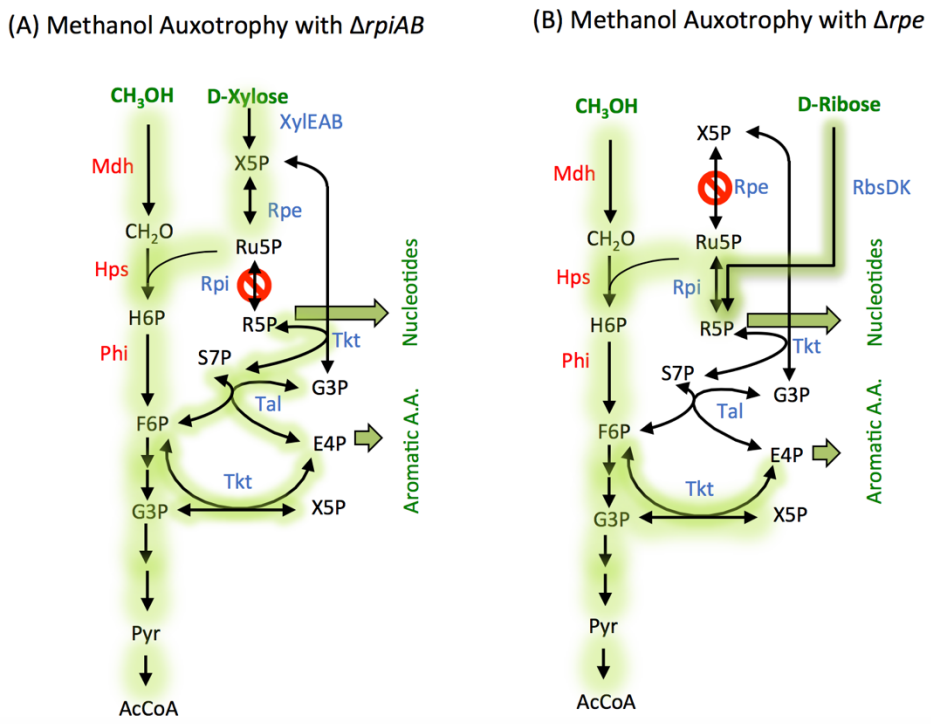


Fig. 2-1. Schematic diagram of synthetic methanol auxotrophy with (A) *ΔrpiAB* or (B) *Δrpe* strain.

An arrow with a slash denotes an eliminated reaction. The green shading indicates the pathways to generate essential metabolites. XylE, D-xylose/proton symporter; XylA, xylose isomerase; XylB, xylulokinase; RbsD, D-ribose pyranase; RbsK, ribokinase; Rpe, ribulosephosphate 3-epimerase; Rpi, ribose-5-phosphate isomerase; Tkt, transketolase; Tal, transaldolase; Mdh, methanol dehydrogenase; Hps, 3-hexulose-6-phosphate synthase; Phi, phosphohexuloisomerase. X5P, xylulose 5-phosphate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7P, sedoheptulose 7-phosphate; G3P, glyceraldehyde 3-phosphate; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; H6P, hexulose 6-phosphate; Pyr, pyruvate; AcCoA, acetyl-CoA.

2.3 Materials and methods

2.3.1 Chemicals and reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, Missouri, USA) unless otherwise stated. ^{13}C methanol was obtained from Cambridge Isotope Laboratories (Tewksbury, Massachusetts, USA). The following enzymes were also purchased from Sigma-Aldrich: hexokinase (*S. cerevisiae*), phosphoglucose isomerase (*S. cerevisiae*), glucose-6-phosphate dehydrogenase (*S. cerevisiae*).

2.3.2 Analytical methods

An Agilent 1100 HPLC was used for measurement of organic acids, sugars (ribose and xylose), and methanol using an Aminex HPX-87H column (Bio-Rad). 30 mM of sulfuric acid was used as mobile phase at a flow rate of 0.6mL/min and run for 30min. For quantification of ethanol and 1-butanol, Agilent Technologies 6890N Network Gas Chromatograph connected with 5973 Network Mass Selective Detector (model G2577A, diffusion pump EI MSD) was used. An initial temperature of 40 °C was held for 2 min, before ramping at 10 °C/min to 60 °C followed by ramping at 100 °C/min to 240 °C with a final two-minute hold. We used a DB-FFAP column (Agilent Technologies 123–3232, 0.32 mm × 30 m × 0.25 μm), with constant pressure of 0.487 bar.

2.3.3 Plasmid construction

Plasmids used in the current research are listed in Table 2-1. DNA fragments for plasmid construction were generated by PCR using Phire Hot Start II DNA polymerase (Thermo Scientific). All plasmids were assembled by Gibson Assembly Master Mix (New England Biolabs) as described by manufacturer's protocol. All of the cloning work was done in *E. coli* XL-1 blue.

Table 2-1. Plasmid list.

Plasmid	Description	Reference
pIB208	P_L*lac*O_1::*mdh2*$_{CT4-1}$ (TIR: 21400)-*hps*$_{BM}$ (TIR: 83400)-*phi*$_{MF}$ (TIR: 200) colE *ori* Carb^r	This study
pCTI2	P_L*lac*O_1::*mdh2*$_{CT2-1}$ (TIR: 82300)-*hps*$_{BM}$ (TIR: 5700)-*phi*$_{MF}$ (TIR: 2400) colE *ori* Carb^r	This study
pTW242	P_L*lac*O_1::*atoB*$_{EC}$-*crt*$_{CA}$-*hbd*$_{CA}$-*ter*$_{TD}$, P_{trc}::*pduP*$_{SE}$ p15A *ori* Kan^r	This study
pTW244	P_L*lac*O_1::*pdc*$_{ZM}$-*adhB*$_{ZM}$ CDF *ori* Spec^r	This study
pCT396	pIB208 with additional Spec^r	This study

mdh2$_{CT4-1}$ and *mdh2*$_{CT2-1}$ indicates variant CT4-1 and CT2-1 of Mdh2 from *Cupriavidus necator* (Wu et al., 2016), respectively. BM, *Bacillus methanolicus*; MF, *Methylobacillus flagellates*; EC, *Escherichia coli*; CA, *Clostridium acetobutylicum*; TD, *Treponema denticola*; SE, *Salmonella enterica*; ZM, *Zymomonas mobilis*. TIR: Prediction of translation initiation rate of synthetic RBS according to RBS calculator (Salis et al., 2010).

2.3.4 Strains and growth conditions

The strains used in this study are summarized in Table 2-2. Both IB730 (Δ*rpiAB*) and IB405 (Δ*rpe*) strains were derived from strain BL21(DE3) using P1 transduction (Thomason et al., 2007) with Keio collection (Baba et al., 2006). For strain IB730, *ΔrpiA* was first constructed and pcp20 (Datsenko and Wanner, 2000) was used to remove kanamycin cassette. *ΔrpiB* was constructed consequently to make strain IB730.

To investigate the effect of mutations acquired during adaptive evolution, a series of strains (CT837 – CT838 and CT850) were constructed with the mutated sequence reverted to its wild-type sequence. Except for CT838, the strains were constructed using lambda red mediated gene replacement describe previously (Datsenko and Wanner, 2000) with chloramphenicol resistance cassette as selection marker.

The *frmRAB* operon in strain CT838 was restored using P1 transduction (Thomason et al., 2007) with donor cell lysate prepared from the *ΔyaiL* strain (JW0345) in Keio collection (Baba et al., 2006).

For methanol dependent growth, all bacterial culture was kept at 37 °C, 250 rpm orbital shaker unless otherwise stated. MOPS minimal medium is composed of 40 mM MOPS, 4 mM tricine, 0.01 mM $FeSO_4$, 9.5 mM NH_4Cl, 0.276 mM K_2SO_4, 0.5 μM $CaCl_2$, 0.525 mM $MgCl_2$, 50 mM NaCl, 0.292 nM $(NH_4)_2MoO_4$, 40 nM H_3BO_3, 3.02 nM $CoCl_2$, 0.962 nM $CuSO_4$, 8.08 nM $MnCl_2$, 0.974 nM $ZnSO_4$, and 1.32 mM K_2HPO_4. For growth tests in MOPS minimal medium, seed cultures for inoculation were grown in Terrific Broth (TB) medium for 16 h. 0.1 mM IPTG for induction of protein expression was added at the beginning of inoculation when necessary. Final concentration of 100 mg/L carbenicillin, 30 mg/L kanamycin, 50 mg/L chloramphenicol, or 250 mg/L spectinomycin were used when appropriate. Notably, we experience huge variation of

lag phase (1 – 3 days) when passing the Δ*rpiAB* auxotrophy strain directly from seed culture to minimal medium for growth on methanol and xylose. This could be a result of dramatic change of growth condition and the stress caused by expressing the methanol assimilation enzyme in high level. Therefore, an intermediate growing stage that buffers the transition was implemented for characterization of methanol-dependent growth and substrate consumption demonstrated in Figs. 2 and 3. Specifically, the seed cultures were washed twice with equal volume of MOPS minimal medium and inoculated (1%) to 3 mL MOPS minimal medium containing 50 mM ribose/xylose and 250 mM methanol with 0.05% casamino acids in BD Vacutainer glass tubes (#366430, BD, Franklin Lakes, NJ, USA) with stopper. The casamino acids were found to shorten the lag phase when the cells are transitioned from growth in rich medium to minimal medium and enabled repeatable growth for the medium transition. The resulting cultures with casamino acids typically reaches OD_{600} ~1.5 in 16 h. For the growth and substrate consumption characterization, the "transition" cultures were then washed twice with MOPS medium and inoculated into 3 mL of identical growth medium but without casamino acids to initial OD of 0.1 in BD Vacutainer glass tubes. The growth experiment was set up under aerobic conditions. However, the stopper of the Vacutainer glass tube was applied at all times (except during culture sampling for substrate analysis) to prevent methanol evaporation. Preliminary test of 250 mM methanol dissolved in 3 mL of MOPS minimal medium showed no measurable evaporation within 3 days under this condition (data not shown). During growth characterization experiment, 300 μL of culture samples were then taken at each desired time point for methanol, xylose, or ribose quantification using HPLC. OD_{600} was monitored by Spectronic 200 (Thermo Scientific) with cultures in the Vacutainer glass tubes.

Table 2-2. Strain list

E. coli Strain	Genotype	Description	Reference
BL21(DE3)	*E. coli* str. B F^{-} *ompT gal dcm lon hsdS$_B$*(r_B^- m_B^-) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	Wild-type	
IB405	BL21(DE3) Δ*rpe*::*kan*	Synthetic methanol auxotrophy strain with Rpe deletion	This study
IB730	BL21(DE3) Δ*rpiA*::*FRT* Δ*rpiB*::*kan*	Initial strain for constructing synthetic methanol auxotrophy strain with Rpi deletion. Cannot grow on ribose plus xylose or methanol plus xylose	This study
CFC65	See Table 3	IB730 evolved to grow in ribose and xylose	This study
CFC133	See Table 3	Evolved strain originated from CFC65, able to grow with methanol and xylose when containing plasmid pIB208	
CT833	CFC133 *add**::*add*(WT)	Replace mutated *add* with WT sequence	This study
CT834	CFC133 *aroK**::*aroK*(WT)	Replace mutated *aroK* with WT sequence	This study
CT835	CFC133 *cyaA**::*cyaA*(WT)	Replace mutated *cyaA* with WT sequence	This study
CT836	CFC133 *deoD**::*deoD*(WT)	Replace mutated *deoD* with WT sequence	This study
CT837	CFC133 *zwf**::*zwf*(WT)	Replace mutated *zwf* with WT sequence	This study
CT838	CFC133 *frmRAB**::*frmRAB*(WT)	Restore truncated *frmRAB* operon	This study
CT850	CFC133 *pykF**::*pykF*(WT)	Replace mutated *pykF* with WT sequence	This study
CFC134	CFC133 Δ*rpiB*::*FRT*	Kanamycin selection marker removed	This study

2.3.5 *ΔrpiAB* directed evolution for substrate adaption

All *E. coli* strains were passed to the next generation when OD_{600} exceeded 1.2, and inoculated initially at OD_{600} = 0.05. The unevolved Δ*rpiAB* strain IB730 was first inoculated in Terrific Broth (TB Sigma) with 7 mM ribose and 7 mM xylose. Strains grew to saturation in two days and were then passed to two different media, Hi Def azure medium (HDA, Teknova), and MOPS EZ Rich Defined Medium (Teknova), both with 100 mM ribose and 100 mM xylose. From Generation 2 to Generation 10, cells were passed from HDA to HDA. Cultures were also passed to MOPS media with same amount of ribose and xylose to compare growth rate on the same basis.

Since Generation 5, ribose and xylose concentrations were reduced to 40mM. Starting from Generation 11, cells were passed from MOPS media to MOPS media. At Generation 20, a single colony strain was isolated and renamed as CFC65. The strain was transformed with plasmid pIB208 for expression of Mdh, Hps, and Phi. Colonies were picked and grown in TB with 7 mM ribose and 7 mM xylose. The strain was then inoculated into MOPS minimal medium with 200 mM methanol, 20 mM xylose, 0.1 mM IPTG, 1x of MEM amino acid (Sigma). Strains were also grown without MEM amino acid as a growth control. MEM amino acid was reduced to 0.5×, 0.25×, 0.125×, 0× at Generation 30, 32, 37, 45 subsequently.

2.3.6 Genome sequencing

Whole genome sequencing of strain CFC65 and CFC133 was performed by GENEWIZ (South Plainfield, NJ, USA) using Illumina MiSeq platform.

2.3.7 Ethanol and 1-butanol production

Engineered strains were cultured in TB medium with appropriate antibiotics for 4 h before adding 0.1 mM IPTG. The strains were grown for another 12–18 h. 500 µl of the cultures were inoculated into 120 mL Hi-Def Azure (HDA) medium (Teknova, CA, USA) supplemented with 0.1 mM IPTG, 50 mM ribose/xylose, and 250 mM methanol until OD_{600} was about 1 (approximately 24–40 h). Appropriate amount of cells were pelleted and washed twice with 30 mL of MOPS minimal medium. The washed pellets were resuspended in 850 µl of MOPS medium, 100 µl of 2 M ribose/xylose, and 50 µl of 10 M ^{13}C methanol to make 1 mL concentrated cultures of OD_{600} = 100. The concentrated culture was prepared under aerobic condition and placed in BD Vacutainer glass tubes with stopper to prevent methanol evaporation. The fermentation was conducted in 37°C, 250rpm orbital shaker. Samples were taken at desired time points and supernatants were isolated for analysis. Ethanol and 1-butanol were measured by GCMS, ^{13}C-labeling was determined by measured mass isotopomer with correction for nature isotope abundance as described (Fernandez et al., 1996). Methanol, xylose, ribose and organic acids byproducts were quantified by HPLC. To calculate cell specific productivities, cell dry weight (CDW; in g/L) was estimated by the conversion factor described previously (Soini et al., 2008): 1 OD_{600} unit = 0.3 CDW.

2.3.8 Enzyme assays

Key enzymes in the pentose phosphate pathway (Tkt, Tal, Rpe, Rpi) were purified from overexpressing strains. Cells were inoculated in 3 mL LB from frozen glycerol stock and grown overnight at 37 °C. The following day, 200 mL LB was inoculated with 1% of overnight culture and grown at 37 °C until optical density was between 0.4 and 0.6, at which a final concentration of 0.1mM IPTG was added. The culture was grown at room temperature overnight, harvested by

centrifuge and broken by sonification. After centrifuging for 1 hour at 18000 rpm and 4 °C, protein solution was vacuum filtered before performing IMAC protein purification with Profinia (Biorad). Final concentrations of purified enzymes are: Tkt at 2.0 μg/μL, Tal at 13.0 μg/μL, Rpe at 5.0 μg/μL, and Rpi at 28.7 μg/μL. Commercial phosphoglucose isomerase and glucose-6-phosphate dehydrogenase were obtained from Sigma-Aldrich. Crude extracts from BL21, IB730, CFC65, CFC65/pIB208, and CFC133 strains were collected. Cells were inoculated in 5 mL LB from frozen glycerol stock and grown overnight at 37 °C with 5000x of 0.5 M IPTG. The following day, cells were harvested by centrifuge and cell pellets were re-suspended in Diglycine Cell Lysis Buffer for sonication. After centrifuging for 30 min at 15000 rpm at 4 °C, solution supernatant was separated and normalized to a concentration of 3.2 μg/μL. Assays were carried out in 100 mM phosphate buffer (pH 7.0) with 20 mM $NADP^+$, 100 mM thiamine pyrophosphate, and 100 mM $MgCl_2$. Tkt, Tal, and Rpe enzyme activity were tested by addition of the substrate, ribose-5-phosphate.

2.4 Results

2.4.1 Design of synthetic methanol auxotrophy

Our design goal was to engineer a strain that cannot grow on a simple co-substrate unless methanol is assimilated. When considering possible co-substrates, we chose precursors of Ru5P to maximize the carbon flux for formaldehyde fixation. Among the possible candidates, D-ribose and D-xylose are both natural substrates for *E. coli* and require only a few enzymatic reactions to produce Ru5P. Theoretically, disrupting any reaction in the non-oxidative part of PPP would abolish the cell's ability to grow on ribose or xylose as a sole carbon source. We focused on Δ*rpe* (coding for ribulose-phosphate 3-epimerase) and Δ*rpiAB* (coding for ribose-5-phosphate isomerase A and B) strains for constructing synthetic methanol auxotrophy. The Δ*rpe* and Δ*rpiAB* strains cannot grow on ribose and xylose minimal medium, respectively. In principle, the growth deficiency of the Δ*rpe* or Δ*rpiAB* strains can be rescued if the first three steps of the RuMP pathway catalzyed by Mdh, Hps and Phi were introduced, forming an alternative route for F6P synthesis (Fig. 2-1). In this strategy, we essentially convert pentose and methanol into a hexose equivalent.

Although transketolase (Tkt) and transaldolase (Tal) deletion mutants can theoretically be used to construct the methanol auxotrophy strain, there were drawbacks in using these strains. The Tal deficient strain was reported to grow on xylose as sole carbon source using a sedoheptulose-1,7-bisphosphate dependent pathway (Nakahigashi et al., 2009). On the other hand, although the Tkt deficient strain cannot use xylose or ribose as sole carbon source, it is unable to synthesize aromatic amino acids (Zhao and Winkler, 1994).

2.4.2 Development of synthetic methanol auxotrophy using Δ*rpe*

We started from using the IB405 (Δ*rpe*) strain to develop synthetic methanol auxotrophy. This strain cannot grow in minimal medium with ribose as a sole carbon source, but can grow in minimal medium with glucose alone or ribose plus xylose as carbon sources (Fig. 2-2A). Growth on ribose plus xylose confirmed the functionality of PPP except for the intended knock-out. Consistent with previous results in the literature (Long et al., 2016; Lyngstadaas et al., 1998), the growth rates in glucose or ribose plus xylose are significantly lower than the wild-type strain, possibly because of the accumulation of PPP intermediates that causes toxicity. Unexpectedly, IB405 can also grow with xylose alone. It may imply the existence of isozyme(s) for Rpe. However, since IB405 strain cannot use ribose as sole carbon source, we still use this strain as our starting point for developing methanol auxotrophy. To construct the synthetic methanol auxotrophy strain, heterologous expression of Mdh, Hps and Phi is required. We initially constructed a plasmid pCTI2 containing synthetic ribosomal binding sites (RBS). Since Mdh has the lowest specific activity among the three enzymes, it was synthesized with an RBS that rendered highest predicted translation initiation rate. Transformation of strain IB405 with pCTI2 resulted in methanol dependent growth around 120 h to reach OD_{600} of 1.4 (Fig. 2-2B). To improve growth, proper balancing of Mdh, Hps, and Phi expression levels is required, as imbalanced enzyme activities would lead to accumulation of toxic intermediate, formaldehyde. To this end, *mdh*, *hps*, and *phi* were constructed as an operon, in which the protein expression level of each gene was altered using an RBS Library designed using the RBS library calculator (Salis et al., 2010) (Fig. 2-2C). The methanol auxotrophy phenotype serves as a platform to enrich the plasmid constructs that confer better methanol dependent growth.

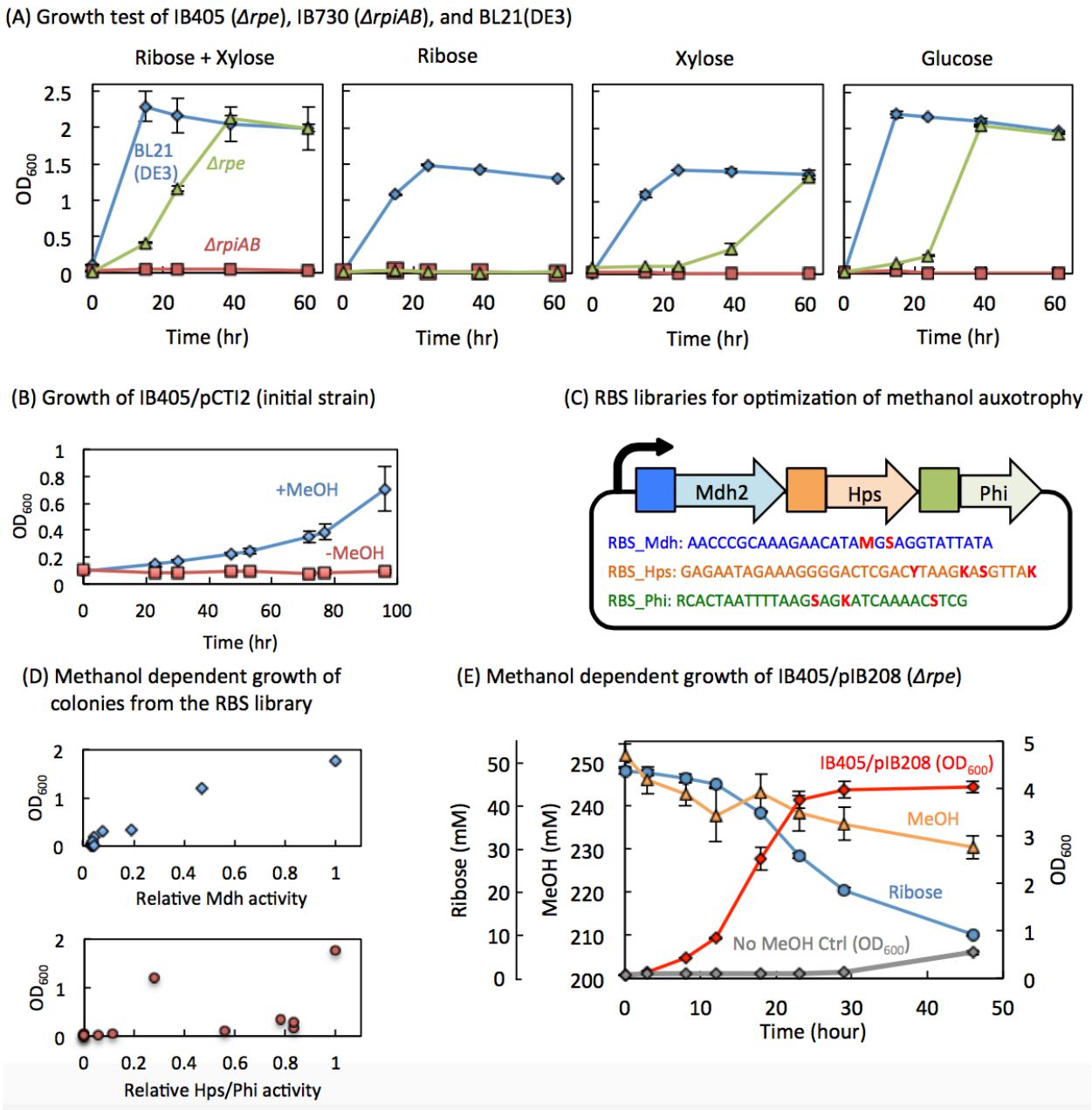


Fig. 2-2. Construction and optimization of methanol auxotrophy in IB405 (BL21(DE3) *Δrpe*).

See figure legends next page.

Fig. 2-2, cont. (A) Growth phenotype of IB730 (*ΔrpiAB*) and IB405(*Δrpe*) was compared to the parental strain BL21(DE3) with 7 mM of ribose, xylose, glucose, or ribose plus xylose (7 mM each). (B) IB405 was transformed with pCTI2 for initial trial of methanol dependent growth. (C) RBS libraries of Mdh, Hps, and Phi for further optimization of methanol assimilation. Each RBS contains diverse expression levels across four orders of magnitude. Degenerate bases: M = A, C; S = C, G; Y = C, T; K = G, T. Strain IB405 was transformed with the RBS library from where (D) 11 colonies were randomly isolated for characterization of methanol dependent growth and Mdh, Hps/Phi activity. OD_{600} was recorded after 48 h. (E) Characterization of strain IB405/pIB208. After transformation of strain IB405 with the RBS library, 4 passages of methanol dependent growth enrichment (grow for 48 h before passing) were carried out. Plasmid pIB208 was isolated from the fastest growing strain, which was identified by first isolating single colonies from the resulting culture on agar (1.5%) plate with MOPS minimal medium plus 7 mM ribose and 200 mM methanol, followed by repeating liquid medium growth to select for the fastest growing strain. Strains were grown in MOPS minimal medium supplemented 250 mM methanol and 50 mM ribose (growth characterization of IB405/ pCTI2 and IB405/pIB208) or 200 mM methanol and 7 mM ribose (all other experiments) at initial OD = 0.05 – 0.1. Cultures were kept at 37 °C, 250 rpm orbital shaker with 0.1 mM IPTG induction. Error bars represent standard deviation (n = 3).

After transformation of the Δ*rpe* strain with the plasmid library, SDS-PAGE showed that the isolated colonies demonstrated diverse expression level of Mdh, Hps, and Phi (Fig. 2-3) and methanol dependent growth phenotype (Fig. 2-2D). The IB405 (Δ*rpe*) strain transformed with the plasmid library was grown in MOPS minimal medium with ribose and methanol for 48 h before re-inoculated into fresh medium. After 4 passages, individual colonies from the final culture were isolated on a MOPS agar plate with methanol and ribose for further growth test. The colony with the fastest growth rate was isolated, its contained plasmid was purified and named as pIB208. Notably, the predicted translation initiation rate of Hps is higher than Mdh on plasmid pIB208 (Table 2-1). The result signified the importance of coordination between formaldehyde production and assimilation for better growth. During initial characterization of substrate consumption, we found that 7 mM of ribose was exhausted during strain growth. Therefore, we increased the ribose concentration to 50mM and simultaneously raised methanol concentration to 250 mM to ensure an amount well above the K_m of Mdh (22 mM) (Wu et al., 2016) throughout the course of growth. With sufficient ribose (50 mM) and methanol (250 mM) concentrations, the IB405/pIB208 strain was able to grow to OD_{600} of 4.0 in 29 h with a specific growth rate of 0.19 ± 0.005 (h^{-1}) (Fig. 2-2E). However, the IB405/pIB208 strain overall consumed 38 mM of ribose but only 21 mM of methanol. Theoretically, methanol and pentose consumptions should be equal. The result possibly indicated an alternative metabolic pathway in the Δ*rpe* methanol auxotrophy that enables ribose metabolism without methanol. Indeed, we observed that the Δ*rpe* strain was able to grow solely on ribose alone after 29 h.

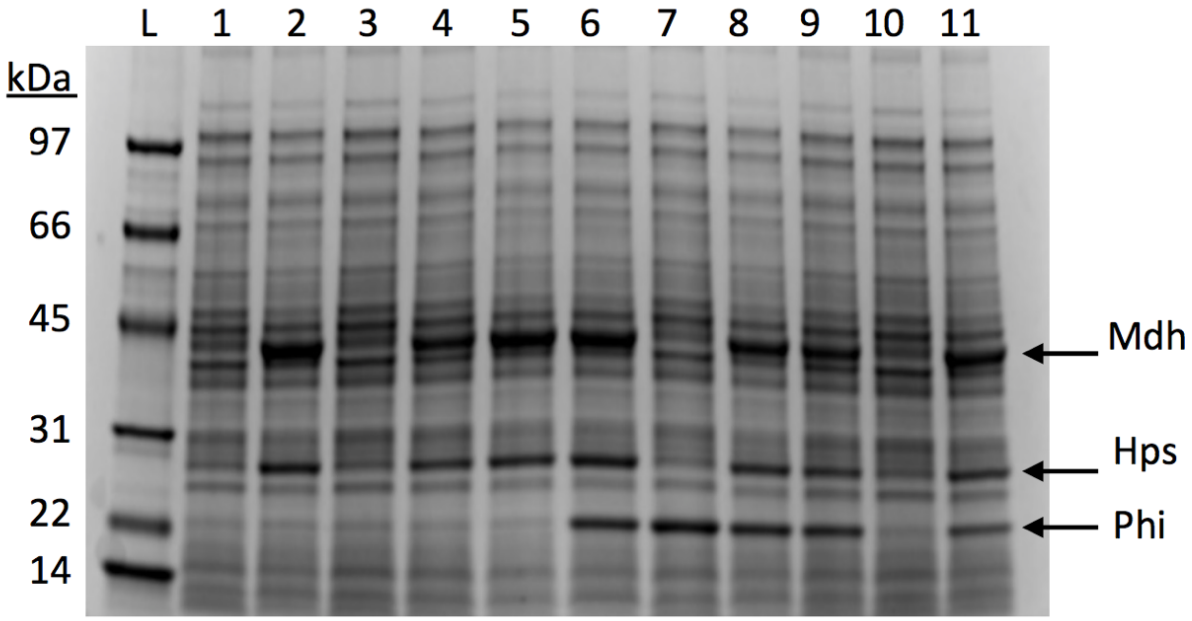


Fig. 2-3. SDS-PAGE of crude extracts from isolated colonies of the Mdh, Hps, and Phi library in strain IB405.

L, protein ladder; number 1 – 11 indicates distinctive colonies.

Previously, an alternative pathway bypassing Rpe through deoxyribonucleosides degradation was proposed. In this pathway, R5P was first used for deoxyribonucleoside biosynthesis, which can be degraded to deoxy-ribose-1-phosphate before being converted to G3P and acetaldehyde by enzymes encoded by *deoB* and *deoC* (Nakahigashi et al., 2009). However, elimination of key genes *deoB* or *deoC* did not affect the growth on ribose (data not shown). We also examined two other genes, *sgcE* and *alsE*, that might encode proteins carrying the Rpe activity (Kim et al., 1997; Nakahigashi et al., 2009). Although individual knock-outs of these two genes did not prevent ribose growth, extended lag phase indicated the contribution of these gene to the ribose growth phenotype (Fig. 2-4). Intriguingly, combination of ΔsgcE and ΔalsE seemed to shorten the lag phase (Fig. 2-4). Further characterization will be required to uncover the underlying pathway(s) of for ribose metabolism in the Δ*rpe* strain.

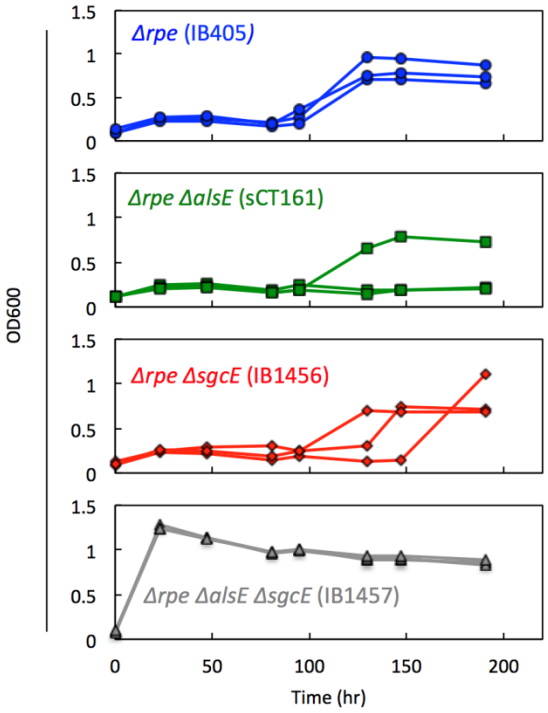


Fig. 2-4. Ribose minimal medium growth of IB405 (Δ*rpe*) and its derivative strains.

Strains were grown in TB with appropriate antibiotics for ~16 hrs before washed twice with MOPS minimal medium and inoculated (initial OD 0.05 - 0.1) to MOPS minimal medium with 7 mM ribose. All strains presented here were constructed using P1 transduction (Thomason et al., 2007) with Keio collection (Baba et al., 2006).

2.4.3 Evolution of Δ*rpiAB* strains allows for a more stringent synthetic methanol auxotrophy strain

The unexpected ribose growth phenotype of the Δ*rpe* (IB405/ pIB208) strain prompted us to explore Δ*rpiAB* for synthetic methanol auxotrophy. Unfortunately, the IB730 (Δ*rpiAB*) strain did not grow in minimal with either glucose alone or ribose plus xylose (Fig. 2-2A) as expected, but was able to grow a rich media, implying a deficiency in metabolism. Therefore, we hypothesized that an unexpected mutation during strain construction might be responsible for the lack of growth. If so, such a problem can be resolved by adaptive evolution.

To do so, the IB730 (Δ*rpiAB*) strain was first grown in Terrific broth (TB) with ribose and xylose to start cell growth in the presence of these two pentoses. The initial culture was followed by passages (0.1% inoculation) in a MOPS medium containing xylose and ribose, but supplemented with low levels of amino acids, vitamins, and nucleotides (Hi-Def Azure, HDA media). After 11 passages, the supplements in the MOPS medium were eliminated and the culture were evolved for another 9 passages. An isolated strain from the 20th passage (CFC 65) was able to reach an OD_{600} of 1.7 in 24 h with xylose and ribose (Fig. 2-5A). CFC65 was then transformed with the plasmid pIB208. However, this initial attempt to demonstrate methanol dependent growth with xylose was still proved unsuccessful. Thus, we once again gradually adapted the strain for methanol dependent growth. Similar to the previous workflow, we grew the strain initially in the HDA medium with methanol and xylose and gradually eliminated the supplements. After another 35 passages, the isolated evolved strain CFC133/pIB208 was finally able to grow on methanol and xylose as sole carbon sources without any amino acid supplements (Fig. 2-5B). Importantly, no growth was observed when methanol was omitted during the course of evolution and characterization (Fig. 2-5A, B, C). Further characterization of CFC133/pIB208 showed a

saturation OD of 4.0 with specific growth rate of 0.17 ± 0.006 (h^{-1}) (Fig. 2-5C). The CFC133/pIB208 strain overall consumed 38 mM of xylose and 43 mM of methanol (Fig. 2-5C), which is very close to the theoretical equimolar consumption. The results implied that most of the methanol consumed was assimilated, since xylose alone cannot be utilized.

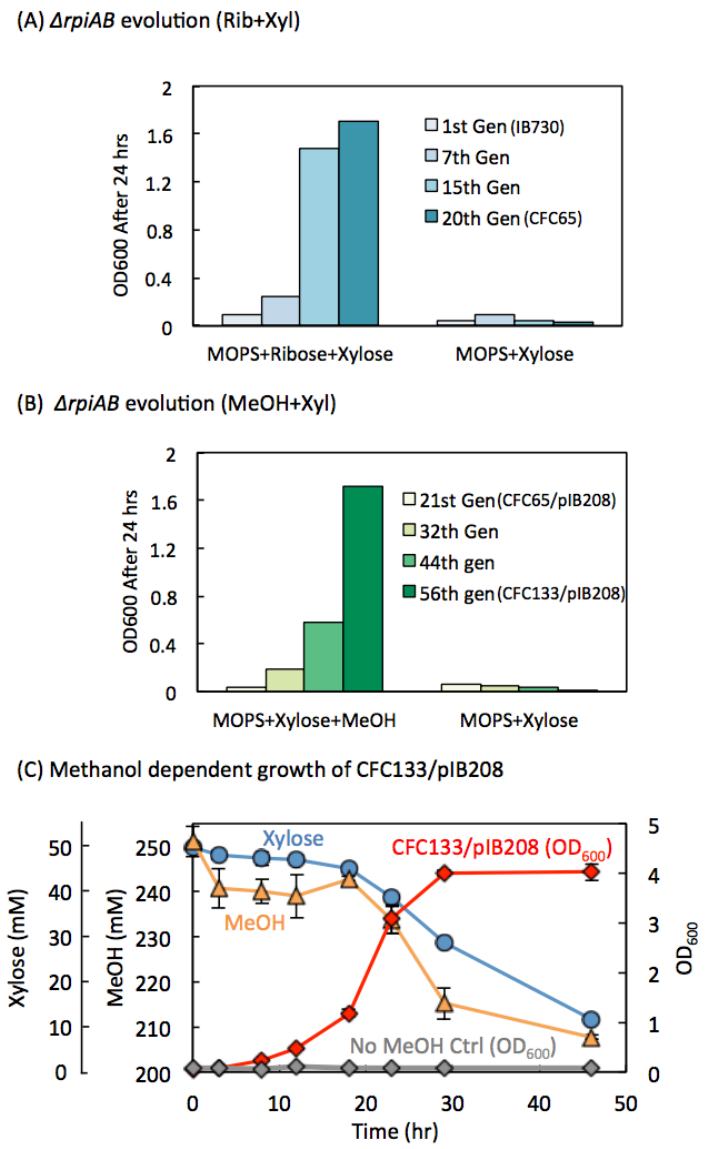


Fig. 2-5. Evolution of IB730 (*ΔrpiAB*) strain and characterization of methanol dependent growth of evolved strain CFC133.

See figure legends next page.

Fig. 2-5, cont. (A) Adaptive evolution of the initial IB730 strain for growth with ribose and xylose. The initial IB730 (*ΔrpiAB*) strain was grown in Terrific broth (TB) medium with 7 mM each of ribose and xylose, followed by passages (0.1% inoculation) of the culture in the Hi-Def Azure media (HDA media, a MOPS based medium supplemented with low levels of amino acids, vitamins, and nucleotides). After the 11th passage, the HDA media was replaced by MOPS minimal medium and the culture was evolved for another 9 passages before strain CFC65 was isolated from the 20th passage. (B) Adaptive evolution for methanol dependent growth with xylose as a co-substrate. CFC65 was transformed with plasmid pIB208 for overexpression of Mdh, Hps, and Phi. Similar to the previous workflow, the CFC65/pIB208 strain was grown initially in the HDA media with 200 mM methanol and 7 mM D-xylose before switching to MOPS minimal medium. After 35 passages, the evolved strain CFC133/pIB208 was finally able to grow on methanol and xylose as sole carbon sources without any amino acid supplements. (C) Characterization of methanol dependent growth and substrate consumption of CFC133/pIB208 in MOPS minimal medium supplemented with 50 mM xylose and 250 mM methanol. Error bars represent the standard deviation (n = 3).

Interestingly, reviving all of the Δ*rpiAB* strains from a LB-based glycerol stock would require 3–4 passes to regain its stable phenotype. Specifically, strains would be inoculated and grown in LB, followed by passing from liquid LB to the desired minimal media. Next, several times of minimal media passing to minimal media passing would be required (either MOPS and ribose plus xylose, or MOPS and methanol plus xylose, data not shown). The reason of the prolonged lag phase remains unknown, but a reasonable explanation would be the requirement of the strain to adapt its metabolism from growing in rich media to a very lean minimal media.

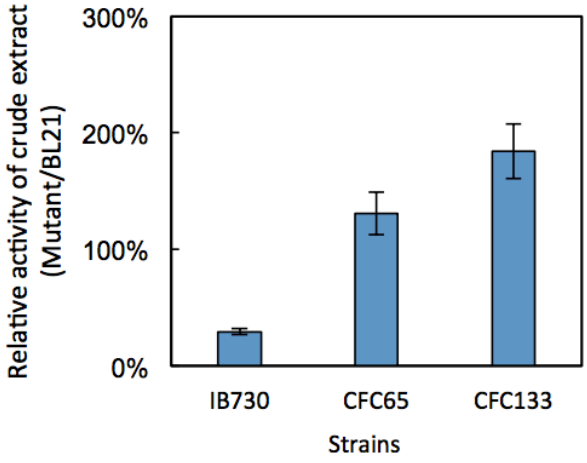
Furthermore, we sequenced the genome of the evolved strains (i.e. IB730, CFC65 and CFC133) to understand the genotype (Table 2-3). After knocking out *rpiAB*, three amino acids of the *tktA* gene were found truncated in strain IB730, starting from position 606 (Fig. 2-6). As evolution process continues, we found out strain CFC65 that successfully grew on ribose and xylose contained TktA with a valine residue recovered among the three deleted amino acids and the new TktA demonstrated higher activity in crude extract (Fig. 2-6). This may have explained the poor initial growth of IB730 strain on ribose and xylose, as Tkt is required to convert the substrates into F6P for growth (Fig. 2-1A).

Table 2-3. Genome sequencing result of strain IB730, CFC65, and CFC133.

R+X: ribose plus xylose; M+X: methanol plus xylose

	Growth Phenotype	Codon Change	Silent Mutation	Non-coding Region	Indel Codon Change	Large Genome Truncation Location
IB730 (Compared to WT BL21)	No growth in both R+X and M+X	None	None	0	*tktA* (2816_2824_delACTGCACGC) (amino acid: 606_608_delTAR)	Δ*rpiA* Δ*rpiB*
CFC65 (Compared to IB730)	Grow in R+X, no growth in M+X	*tktA* (T606V) *glpK* (A55D)	None	4	*tktA* (2819_2824_delGCACGC) (amino acid: 607_608_delAR) *phnE* (405_406_insACGCCAGC)	
CFC133 (Compared to CFC65)	Grow in M+X, no growth in R+X	*insB-9* (M90L) *add* (L157*) *zwf* (Q433*)	None	6	Capsid component B (150_151_insA) [Frameshift] *clpA* (273_274_insA) [Early Stop Codon] pykF (1384_1385_insTAACACC) [Early Stop Codon] *aroK* (510_delC) [Stop Codon Extended] cyaA (1442_1443_insATCAGCC) [Early Stop Codon] deoD (679-680_insATCA) [Frameshift]	347601-350255 The operon *frmR-frmA-frmB* and *yaiO* was deleted 771473-788064 Capsid component B & C was partially deleted

(A)



(B)

Strains	Mutation details	Amino acid sequence of tktA starting from site V605
BL21		VTARVAV
IB730	3 codons trimmed	VV---AV
CFC65	1 mutated, 2 codons trimmed	VV--VAV
CFC133		

Fig. 2-6. Characterization of Tkt throughout evolution

(A) Tkt crude extract activity. The error bars represent the standard deviation, n=3 (B) *tktA* mutation list of Δ*rpiAB* strains.

By comparing CFC133 to CFC65, we noticed seven genes that are known with functions related to either central metabolism (*add, zwf, pykF, aroK, cyaA, deoD*) or formaldehyde detoxification (*frmRAB* operon). These mutations might play a critical role in the methanol dependent growth phenotype. To investigate the contribution of each of these gene mutations to the growth phenotype, we individually replaced the mutated sequence with wild-type sequence using the lambda-red system (Datsenko and Wanner, 2000). For the large truncation of *frmRAB* operon, P1 transduction was used (Thomason et al., 2007). As shown in Fig. 2-7A, the growth phenotype was abolished when the deleted *frmRAB* operon was restored. It showed that deletion of *frmRAB* operon is essential to the methanol dependent growth, possibly by conserving formaldehyde for cell growth (Fig. 2-7B). Indeed, deletion of formaldehyde detoxification had been shown to assist methanol assimilation in previous report (Müller et al., 2015). Another mutation that contributed significantly to the methanol dependent growth phenotype was *cyaA*. Our result showed that when the mutation on the *cyaA* gene was reverted to wild-type sequence, the phenotype of methanol dependent growth cannot be observed (Fig. 2-7A). *cyaA* encodes an enzyme that catalyze the production of important regulator molecule cyclic AMP (cAMP) from ATP (Fig. 2-7B). It is possible that loss of *cyaA* function granted a favorable change of regulation for methanol utilization. We hypothesized that a deleterious mutation of *cyaA* gene resulted in lower transcription level of TCA cycle enzymes, which might be beneficial for methanol utilization since TCA cycle produces NAD(P) H that inhibits methanol oxidation by Mdh. Further details are described in the Discussion section.

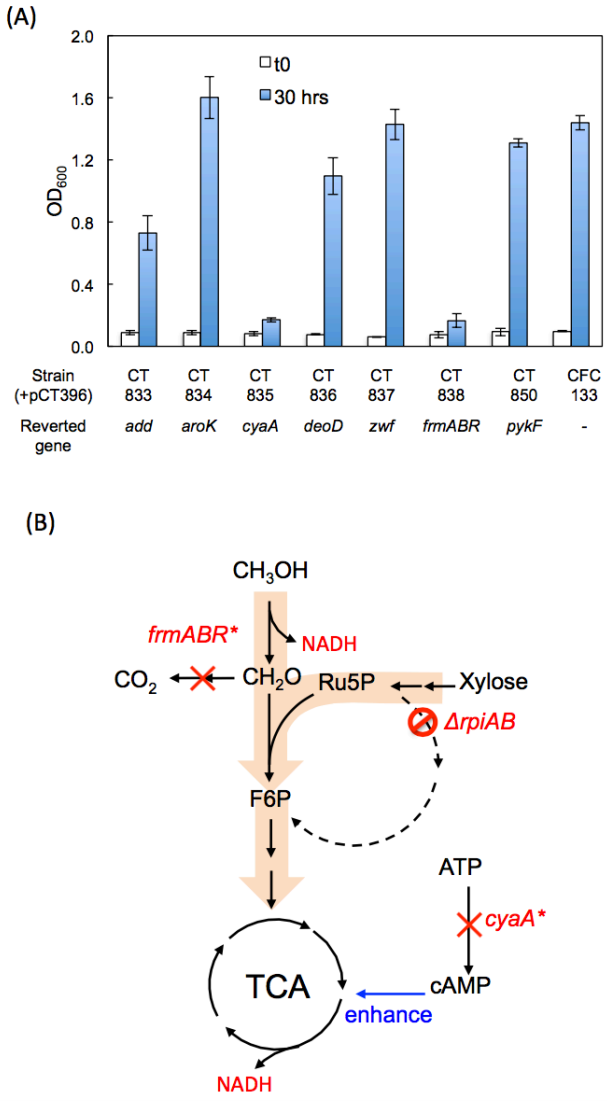


Fig. 2-7. Characterization of key mutations for methanol and xylose growing phenotype of CFC133 (*ΔrpiAB*) strain.

(A) The mutated genes characterized in strain CFC133 (see Table 2-3) were reverted to their wild-type sequence one by one to test their contribution to the growth phenotype. The tested strains were transformed with pCT396 for expression of Mdh, Hps, and Phi. Strains were grown in TB for ~16 h before washed twice with MOPS medium and inoculated to MOPS medium with 0.05% casamino acids, 50mM xylose, and 250mM methanol with appropriate antibiotics. Error bars represent the standard deviation (n = 3). (B) Proposed effect of *frmRAB* and *cya*A mutations to methanol assimilation.

2.4.4 Conversion of methanol to ethanol and 1-butanol

So far, we have established two designs of methanol auxotrophy strains based on *rpe* and *rpiAB* deletions each. Although the Δ*rpe* strains demonstrated methanol auxotrophy, it exhibited a weak phenotype of growth on ribose without methanol. On the other hand, the methanol auxotrophy observed using the Δ*rpiAB* strains was fairly robust. Growth of Δ*rpiAB* strain on xylose only without methanol was never observed in multiple experiments of 30 days or longer cultivation (data not shown).

Xylose is one of the most abundant sugars in cellulosic biomass. Thus, co-utilization of xylose with methanol can potentially "upgrade" the production by providing extra carbon and reducing power. To demonstrate the utility of methanol auxotroph strain for methanol conversion to products, we aimed to produce ethanol and 1-butanol. The ethanol production pathway was constructed by plasmid-based expression of pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (AdhB) from *Zymomonas mobilis* (Fig. 2-8A) as described previously (Ohta et al., 1991). Here we used ^{13}C-methanol to trace the carbon derived from methanol was used with either ribose or xylose. After 3 h of fermentation, the Δ*rpiAB* strain yielded 4.6 g/L (100 mM) of ethanol (Fig. 2-8B) from methanol and xylose (productivity 0.051 g/gDCW/h). The control strain (wild-type BL21(DE3) with the same methanol assimilation and ethanol production plasmids) only showed 2.8g/L (61 mM) of ethanol produced, representing a 64% of increase in titer when the synthetic methanol auxotrophy was implemented for methanol conversion. The Δ*rpe* strain ethanol production using methanol and ribose showed a similar trend (productivity 0.046 g/gDCW/h). The titer was enhanced by 37% when comparing the auxotrophy strain (90 mM, 4.1 g/L) to the control strain (66 mM, 3.0 g/L) (Fig. 2-8C). When comparing the metabolites being produced during the fermentation, it is noticed that BL21(DE3) strains produced more acetate than the Δ*rpiAB* or Δ*rpe*

strains (Table 2-4). It may contribute to the lower titer of ethanol produced from the BL21(DE3) strains.

Both ethanol production using CFC133/pIB208/pTW244 and IB405/pIB208/pTW244 showed significant M+1 labeling product in the mass spectra (Fig. 2-8D). Increment of labeled product (M+1 and M +2) was observed when comparing the BL21/pIB208/pTW244 to CFC133/pIB208/pTW244 (2–43%) or IB405/pIB208/pTW244 (6–31%) auxotrophy strain (Fig. 2-8D). The result demonstrated high level of methanol incorporation to the product. Notably, the theoretical product labeling should be 50% if pyruvate was only synthesized from ^{13}C methanol and non-labeled xylose through EMP glycolysis (Fig. 2-8A). The Δ*rpiAB* auxotrophy strain indeed showed a similar labeling pattern. M +2 signals were only observed in CFC133/pIB208/pTW244, which indicated that Tal and Tkt carbon rearrangement activity were stronger, possibly due to requirement of generation of R5P for nucleotide synthesis. The lower labeling percentage in the Δ*rpe* auxotrophy strain could be attributed to the unknown bypass pathway that enabled ribose utilization without methanol assimilation. Although the labeling patterns clearly showed methanol incorporation, our control experiment with the auxotrophy strains incubating with only the pentoses demonstrated significant consumption (Fig. 2-9). For the Δ*rpe* auxotrophy, it can again be explained by the unknown bypass pathway. However, it was not clear why the Δ*rpiAB* strain also showed continuous consumption over the measure time course (Fig. 2-9A). There was no detection of common fermentation products (e.g. acetate, lactate, succinate). Further investigation such as intracellular metabolite measurement would be required to elucidate the unknown direction of the carbon flux and its product.

Methanol consumption was also enhanced by 11-fold and 5-fold for the Δ*rpiAB* and Δ*rpe* strain (Fig. 2-8E) when compared to BL21 strain, respectively. It is possible that the titer

improvement is not only due to the utilization of methanol as a source of carbon, but also a source of reducing power to enhance overall production of reduced product. The results indicated that the synthetic methanol auxotrophy engineering strategy effectively increase production titer by enabling more efficient methanol utilization.

We further demonstrated 1-butanol production (Lan et al., 2013; Shen et al., 2011) using a plasmid-based construct (pTW242). With the CFC134/pIB208/pTW242 (Δ*rpiAB*) strain, we demonstrated 27 mM (2.0 g/L) of 1-butanol when 87 mM of methanol and 100 mM of xylose was consumed (Fig. 2-8F). The slightly higher xylose consumption was consistent with the xylose consumption in the absence of methanol, as discussed above. Since PduP also has activity towards acetyl-CoA, a significant amount (1.5g/L) of ethanol was produced as byproduct. Similar to the results in ethanol production, significant M+1 and M+2 mass isotopomer of 1-butanol was observed in the mass spectra (Fig. 2-8G). Close to the expectation, M+1 mass isotopomer accounts for 52% of the product 1-butanol while M+2 molecules took up 19%.

In both cases of ethanol and 1-butanol production using the Δ*rpiAB* strain, equimolar consumption of methanol and xylose was observed. The substrate utilization pattern was consistent with that of methanol dependent growth (Fig. 2-5C). It served as an evidence to show that most of the methanol consumed was assimilated.

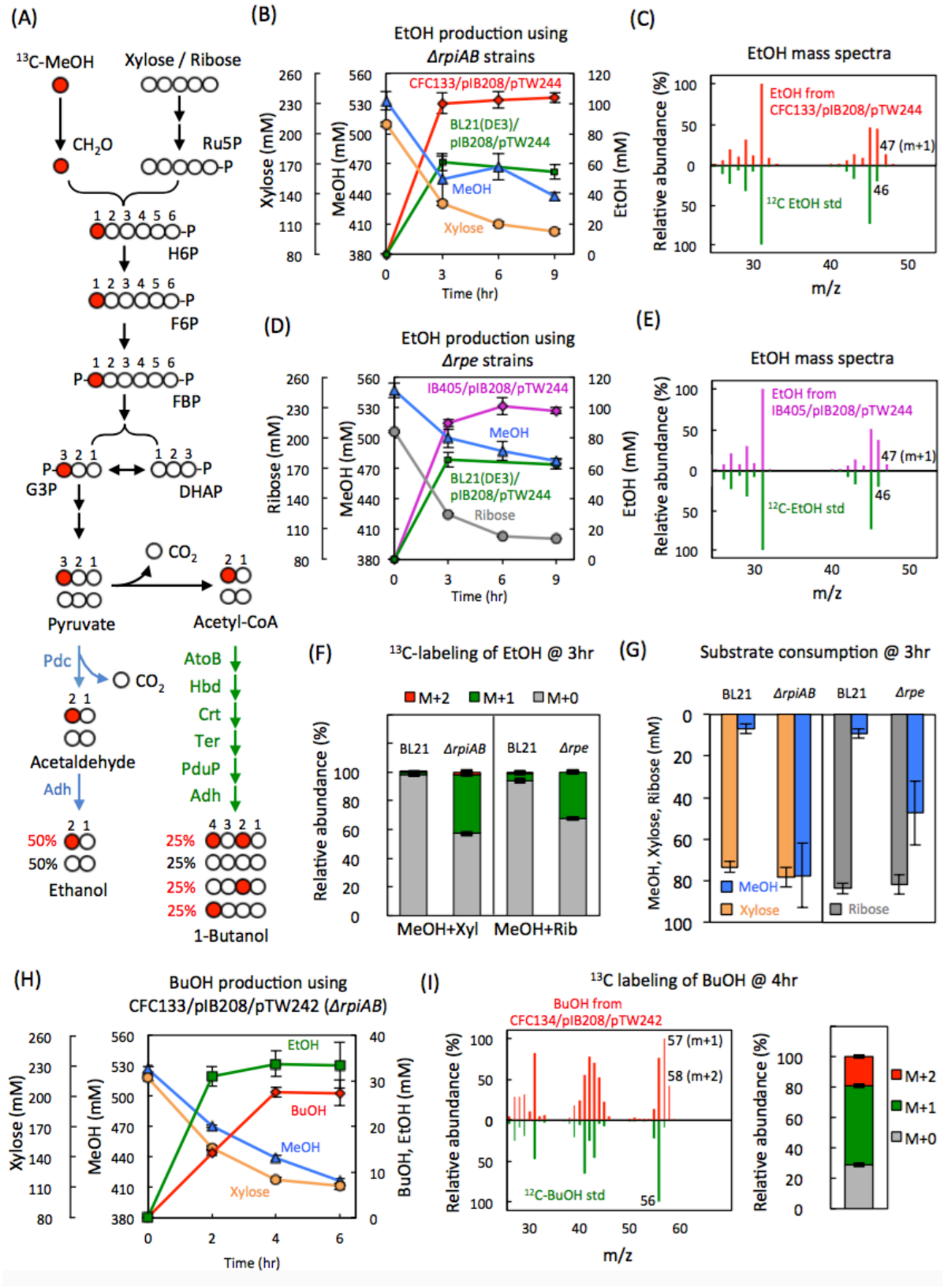


Fig. 2-8. Ethanol and 1-butanol production using methanol auxotrophy strains.

Fig. 2-8, cont. For all production experiments, strains were grown in HDA medium for ~24 h until OD_{600} reaches 2. Cells were then washed twice with MOPS minimal medium before being concentrated to OD_{600} ~100 in the same medium. The concentrated cells were kept in BD Vacutainer glass tube with stoppers to prevent methanol evaporation. 550 mM methanol and 200 mM ribose or xylose (final concentration) were added to start the production. (A) Carbon labeling of ethanol and 1-butanol derived from ^{13}C-methanol with co-utilization of xylose or ribose in the methanol auxotrophy strains. (B) Ethanol production of CFC133/pIB208/pTW244 (*ΔrpiAB*) and BL21(DE3)/pIB208/pTW244 using methanol and xylose. (C) Ethanol production of IB405/pIB208/pTW244 (*Δrpe*) and BL21(DE3)/pIB208/pTW244 using methanol and ribose. (D) Mass isotopomer abundance of ethanol derived from ^{13}C-methanol and (E) methanol and xylose or ribose consumption during ethanol production (3 h sample). *ΔrpiAB*: CFC133/pIB208/pTW244, *Δrpe*: IB405/pIB208/ pTW244, BL21: BL21(DE3)/pIB208/pTW244. (F) 1-Butanol production of CFC134/pIB208/pTW242 (*ΔrpiAB*) using methanol and xylose. Ethanol was a primary byproduct due to promiscuity of PduP to convert acetyl-CoA to acetaldehyde. (G) Mass isotopomer abundance of 1-butanol produced by CFC134/pIB208/pTW242 (4 h sample). Relative abundance of mass isotopomers were calculated as described before (Fernandez et al., 1996). All spectra were normalized to the most abundant internal peak. Error bars represents the standard deviation (n = 3).

Table 2-4. Fermentation products and carbon balance of ethanol production experiment.

M+X: methanol+xylose. M+R: methanol+ribose. Xyl: xylose. Rib: ribose. N.D. not detectable. Errors are indicated in standard deviation (n =3). Data shown here were the result after 3 hr fermentation, as shown in Figure 2-8B and 2-8C.

Strain (carbon sources)		Methanol input	Xyl/Rib input	Ethanol output	Succinate output	Formate output	Acetate output	CO_2 output[a]	Recovery (%)
CFC133	g/L	2.19±0.03	11.7±0.71	4.69±0.18	0.63±0.015	0.010± 0.0014	0.88±0.086		
/pIB208	mM	68.5	78.2	101.9	5.34	0.23	14.7	101.9	
/pTW244 (M+X)	mM carbon	68.5	391	203.8	21.36	0.23	29.4	101.9	77.6
BL21 (DE3)	g/L	0.22±0.073	11.0±0.39	2.8±0.025	0.034±0.003	N.D.	2.97±0.039		
/pIB208	mM	6.9	73.5	61	0.28		49.5	61	
/pTW244 (M+X)	mM carbon	6.9	367.5	122	1.12		99	122	91.9
IB405	g/L	1.51±0.47	12.2±0.14	4.14±0.11	0.16±0.033		1.79±0.08		
/pIB208	mM	47.2	81.6	90	1.38	N.D.	29.8	90	
/pTW244 (M+R)	mM carbon	47.2	408	180	5.52		59.6	90	73.6
BL21 (DE3)	g/L	0.29±0.13	12.5±0.27	3.02±0.22			2.60±0.06		
/pIB208	mM	9	83.6	65.6	N.D.	N.D.	43.4	65.6	
/pTW244 (M+R)	mM carbon	9	418	131.2			86.8	65.6	66.4

[a] CO2 output was estimated as the same as ethanol output, due to the reaction of pyruvate decarboxylase.

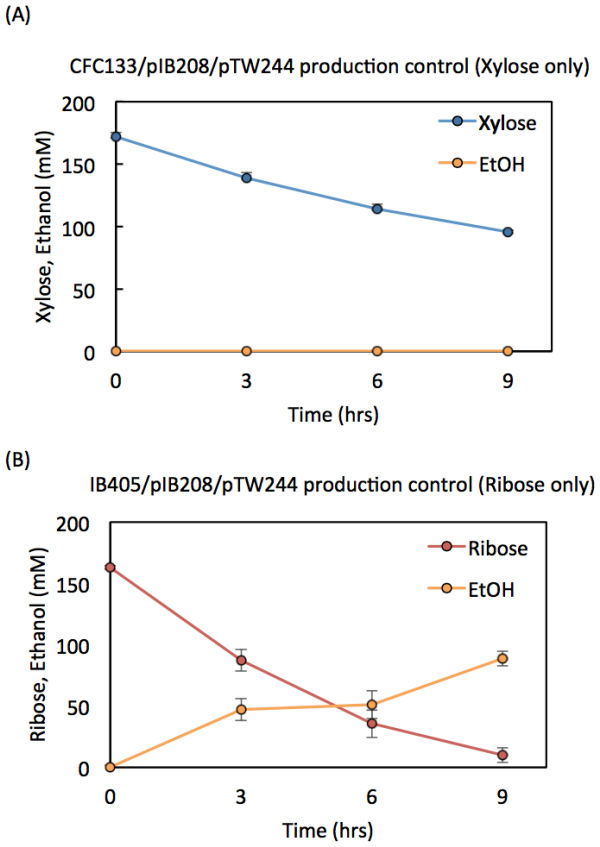


Fig. 2-9. Ethanol production of methanol auxotrophy strains

(A) CFC133/pIB208/pTW244 and (B) IB405/pIB208/pTW244 with only pentose (ribose or xylose) in the absence of methanol. The experiment followed the same procedure as of the production shown in Figure 5. Error bars represents the standard deviation (n = 3).

2.5 Discussion

Existing methanol assimilation pathways involve highly toxic formaldehyde as intermediate (Clomburg et al., 2017). Moreover, overexpressing RuMP enzymes proved to have limited methanol assimilation and cannot support *E. coli* to grow on methanol alone (Müller et al., 2015; Whitaker et al., 2016). In this work, we demonstrated two strategies of "synthetic methanol auxotrophy" based on *rpiAB* and *rpe* deletion strains. The resulting *E. coli* strains cannot grow on xylose or ribose alone, respectively, unless methanol is present. Although the Δ*rpe* strain allowed an easier case for developing the synthetic methanol auxotroph, the strain showed a weak phenotype that bypassed the requirement of methanol assimilation. This hurdle prevented the Δ*rpe* strain from further application as a platform for methanol dependent growth and production. On the other hand, the Δ*rpiAB* strain took more effort to reach methanol auxotrophy, but the resulting strain exhibited a more stringent methanol dependent growth.

The strong phenotype of the Δ*rpiAB* strain enabled the use of evolution approaches to increase methanol utilization. Because of the complexity of metabolic regulation, evolution allowed us to overcome the limitation of rational metabolic pathway design. The outcome of evolution is sometimes counter-intuitive and cannot be predicted rationally a priori (Lin et al., 2018). Moreover, the methanol auxotroph can serve as a platform to screen for methanol assimilating enzymes (e.g. Mdh, Hps, Phi) with higher performance. The methanol auxotro*phi*c strains can also be eventually engineered to achieve growth or production on methanol as a sole carbon source. Since the Δ*rpiAB* strain was forced to evolve for higher activity of PPP to produce essential nucleotides precursor (Fig. 2-1A), a complete RuMP cycle can be achieved by restoring the Rpi function in the evolved Δ*rpiAB* auxotro*phi*c strain.

During adaptive evolution of the Δ*rpiAB* strain to realize methanol dependent growth, *cyaA* was identified as one of the essential mutations for the growth phenotype. In *E. coli*, cAMP binds to cAMP receptor protein (CRP) and the cAMP-CRP affects transcription level of genes involved in TCA cycle, sugar catabolism, stress response, and many others critical pathways (Gosset et al., 2004). It is possible that the early stop codon mutation of *cyaA* mimics the effect of a CRP null mutant shown previously (Gosset et al., 2004), which resulted in lower transcription level of TCA cycle genes. Lower transcription level of TCA cycle could be potentially beneficial since one of the major functions of the cycle is to produce NAD(P)H, which in a high level would inhibit methanol oxidation by Mdh (Fig. 2-7B). Further investigation in transcription and protein activity level would be required to decipher the underlying mechanism of how *cyaA* mutation affects methanol assimilation. Other mutations such as *zwf* or *deoD* may have minor contributions to eliminate or tune down competing pathways such as the Entner-Doudoroff pathway and nucleotide synthesis pathways, but were not essential to the growth phenotype.

The recent work by Meyer et al. (2018), which appeared during the preparation of this manuscript, reported the utilization of the same gene knockout (Δ*rpiAB*) to establish methanol auxotrophy in gluconate minimal media. Gluconate is not an abundant carbon source, and loses one carbon in the form of CO_2 when it is converted in Ru5P via 6-phosphogluconate dehydrogenase, thus impeding the possibility of increasing carbon yield of products by methanol utilization. In contrast, our work utilizes xylose, an abundant sugar in nature as the co-substrate, and can directly supply Ru5P without carbon loss. Not surprisingly, the evolved strains described by Meyer et al. (2018) (strain MeSV2.2) and by the current report (strain CFC133) both shared a common mutation on the formaldehyde detoxification operon (*frmRAB*) as it might increase intracellular concentration of formaldehyde. Interestingly, as Meyer et al. knocked out *edd*

strategically, the strain CFC133/pIB208 in this report also had a mutated *zwf*, suggesting that the Entner-Doudoroff pathway should be blocked for efficient methanol assimilation. Last but not the least, CFC133/pIB208 has a higher growth rate ($0.17 \pm 0.006\ h^{-1}$), with is approximately twice as fast as the previously reported strain reported by Meyer et al. ($0.081 \pm 0.002\ h^{-1}$). In addition, we demonstrated ethanol and 1-butanol formation from methanol. The strategy can be used to upgrade xylose and methanol as fermentation substrates, as both are potentially abundant carbon sources but difficult to use in microbial fermentation.

2.6 Acknowledgements

This work is supported by the Reducing Emissions using Methanotrophic Organisms for Transportation Energy (REMOTE) pro- gram of the Advanced Research Projects Agency-Energy (Award DE- AR0000430). We also thank Seunghwan Lee, Aaron Baugh for their technical support in strain evolutions and enzyme assays.

2.7 References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12 in frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2. https:// doi.org/10.1038/msb4100050.

Bennett, R.K., Gonzalez, J.E., Whitaker, W.B., Antoniewicz, M.R., 2017. Expression of heterologous non-oxidative pentose phosphate pathway from Bacillus methanolicus and phosphoglucose isomerase deletion improves methanol assimilation and metabolite production by a synthetic Escherichia coli methylotroph. Metab. Eng. 45, 75–85. https://doi.org/10.1016/j.ymben.2017.11.016.

Bennett, R.K., Steinberg, L.M., Chen, W., Papoutsakis, E.T., 2018. Engineering the bioconversion of methane and methanol to fuels and chemicals in native and synthetic methylotrophs. Curr. Opin. Biotechnol. 50, 81–93. https://doi.org/10.1016/j.copbio. 2017.11.010.

Bogorad, I.W., Chen, C.-T., Theisen, M.K., Wu, T.-Y., Schlenz, A.R., Lam, A.T., Liao, J.C., 2014. Building carbon–carbon bonds using a biocatalytic methanol condensation cycle. Proc. Natl. Acad. Sci. USA 111, 15928–15933. https://doi.org/10.1073/pnas. 1413470111.

Clomburg, J.M., Crumbley, A.M., Gonzalez, R., 2017. Industrial biomanufacturing: the future of chemical production. Science 355, aag0804. https://doi.org/10.1126/ science.aag0804.

Conrado, R.J., Gonzalez, R., 2014. Envisioning the bioconversion of methane to liquid fuels. Science. https://doi.org/10.1126/science.1246929.

Dai, Z., Gu, H., Zhang, S., Xin, F., Zhang, W., Dong, W., Ma, J., Jia, H., Jiang, M., 2017. Bioresource Technology Metabolic construction strategies for direct methanol utilization in Saccharomyces cerevisiae. Bioresour. Technol. 245, 1407–1412. https:// doi.org/10.1016/j.biortech.2017.05.100.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645. https://doi.org/10.1073/pnas.120163297.

Fernandez, F.A., Des Rosiers, C., Previs, S.F., David, F., Brunengraber, H., 1996. Correction of ^{13}C mass isotopomer distributions. J. Mass Spectrom. 31, 255–262.

Gonzalez, C.F., Proudfoot, M., Brown, G., Korniyenko, Y., Mori, H., Savchenko, A.V., Yakunin, A.F., 2006. Molecular basis of formaldehyde detoxification: characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. J. Biol. Chem. 281, 14514–14522. https://doi.org/10.1074/jbc.M_{600}996200.

Gonzalez, J.E., Bennett, R.K., Papoutsakis, E.T., Antoniewicz, M.R., 2017. Methanol assimilation in Escherichia coli is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. Metab. Eng. 45, 67–74. https://doi.org/10. 1016/j.ymben.2017.11.015.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A., Saier, M.H., 2004. Transcriptome Analysis of Crp-Dependent Catabolite Control of Gene Expression in Escherichia coli 186, 3516–3524. 〈http://dx.doi.org/10.1128/JB.186.11.3516-3524.2004〉.

Haynes, C.A., Gonzalez, R., 2014. Rethinking biological activation of methane and conversion to liquid fuels. Nat. Chem. Biol. 10, 331–339. https://doi.org/10.1038/ nchembio.1509.

He, H., Edlich-Muth, C., Lindner, S.N., Bar-Even, A., 2018. Ribulose monophosphate shunt provides nearly all biomass and energy required for growth of *E. coli*. ACS Synth. Biol. 7, 1601–1611. https://doi.org/10.1021/acssynbio.8b00093.

Kim, C., Song, S., Park, C., Al, K.I.M.E.T., 1997. The D -Allose operon of Escherichia coli K-12. J. Bacteriol. 179, 7631–7637.

Koopman, F.W., De Winde, J.H., Ruijssenaars, H.J., 2009. C1 compounds as auxiliary substrate for engineered Pseudomonas putida S12. Appl. Microbiol. Biotechnol. 83, 705–713. https://doi.org/10.1007/s00253-009-1922-y.

Lan, E.I., Ro, S.Y., Liao, J.C., 2013. Oxygen-tolerant coenzyme A-acylating aldehyde dehydrogenase facilitates efficient photosynthetic n-butanol biosynthesis in cyanobacteria. Energy Environ. Sci. 6, 2672. https://doi.org/10.1039/c3ee41405a.

Leßmeier, L., Pfeifenschneider, J., Carnicer, M., Heux, S., Portais, J., Wendisch, V.F., 2015. Production of carbon-13-labeled cadaverine by engineered Corynebacterium glutamicum using carbon-13-labeled methanol as co-substrate. Appl. Microbiol. Biotechnol. 10163–10176. https://doi.org/10.1007/s00253-015-6906-5.

Lin, P.P., Jaeger, A.J., Wu, T.-Y., Xu, S.C., Lee, A.S., Gao, F., Chen, P.-W., Liao, J.C., 2018. Construction and evolution of an Escherichia coli strain relying on nonoxidative glycolysis for sugar catabolism. Proc. Natl. Acad. Sci. USA 115, 3538–3546.

Long, C.P., Gonzalez, J.E., Sandoval, N.R., Antoniewicz, M.R., 2016. Characterization of physiological responses to 22 gene knockouts in Escherichia coli central carbon metabolism. Metab. Eng. 37, 102–113. https://doi.org/10.1016/j.ymben.2016.05. 006.

Lyngstadaas, A., Sprenger, G.A., Boye, E., 1998. Impaired growth of an Escherichia coli *rpe* mutant lacking ribulose-5-phosphate epimerase activity. Biochim. Biophys. Acta Gen. Subj. 1381, 319–330. https://doi.org/10.1016/S0304-4165(98)00046-4.

Meyer, F., Keller, P., Hartl, J., Gröninger, O.G., Kiefer, P., Vorholt, J.A., 2018. Methanolessential growth of Escherichia coli. Nat. Commun. 9, 1508. https://doi.org/10. 1038/s41467-018-03937-y.

Müller, J.E.N., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W.J., Wendisch, V.F., Brautaset, T., Portais, J.-C., Vorholt, J.A., 2015. Engineering Escherichia coli for methanol conversion. Metab. Eng. 28, 190–201. https://doi.org/10.1016/j.ymben.2014.12.008. Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y.,

Honma, M., Mori, H., Tomita, M., 2009. Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. 5, 1–14. https://doi.org/10.1038/msb.2009.65.

Ohta, K., Beall, D.S., Mejia, J.P., Shanmugam, K.T., Ingram, L., 1991. Genetic Improvement of Escherichia coli for ethanol production: chromosomal integration of zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. Appl. Environ. Microbiol. 57, 893–900.

Price, J.V., Chen, L., Whitaker, W.B., Papoutsakis, E., Chen, W., 2016. Scaffoldless engineered enzyme assembly for enhanced methanol utilization. Proc. Natl. Acad. Sci. USA 113, 12691–12696. https://doi.org/10.1073/pnas.1601797113.

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2010. Automated design of synthetic ribosome binding sites to precisely control protein expression. Nat. Biotechnol. 27, 946–950. https://doi.org/10.1038/nbt.1568.Automated.

Shen, C.R., Lan, E.I., Dekishima, Y., Baez, A., Cho, K.M., Liao, J.C., 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. Appl. Environ. Microbiol. 77, 2905–2915. https://doi.org/10.1128/AEM.03034-10.

Soini, J., Ukkonen, K., Neubauer, P., 2008. High cell density media for Escherichia coli are generally designed for aerobic cultivations – consequences for large-scale bioprocesses and shake flask cultures. Microb. Cell Fact. 7, 26. https://doi.org/10.1186/ 1475-2859-7-26.

Thomason, L.C., Costantino, N., Court, D.L., 2007. *E. coli* Genome manipulation by P1 transduction. Curr. Protoc. Mol. Biol. 1.17.1–1.17.8. https://doi.org/10.1002/ 0471142727.mb0117s79.

Whitaker, W.B., Jones, J.A., Bennett, R.K., Gonzalez, J., Vernacchio, V.R., Collins, S.M., Palmer, M.A., Schmidt, S., Antoniewicz, M.R., Koffas, M.A.G., Papoutsakis, E.T., 2016. Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. Metab. Eng. https://doi.org/10.1016/j.ymben.2016.10.015.

Whitaker, W.B., Jones, J.A., Bennett, R.K., Gonzalez, J.E., Vernacchio, V.R., Collins, S.M., Palmer, M.A., Schmidt, S., Antoniewicz, M.R., Koffas, M.A., Papoutsakis, E.T., 2017. Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. Metab. Eng. 39, 49–59. https://doi.org/10.1016/j.ymben.2016.10. 015.

Whitaker, W.B., Sandoval, N.R., Bennett, R.K., Fast, A.G., Papoutsakis, E.T., 2015. Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. Curr. Opin. Biotechnol. 33, 165–175. https://doi.org/10.1016/j.copbio.2015.01.007.

Witthoff, S., Schmitz, K., Niedenführ, S., Nöh, K., Noack, S., Bott, M., Marienhagen, J., 2015. Metabolic engineering of Corynebacterium glutamicum for methanol metabolism. Appl. Environ. Microbiol. 81, 2215–2225. https://doi.org/10.1128/AEM. 03110-14.

Wu, T.Y., Chen, C.T., Liu, J.T.J., Bogorad, I.W., Damoiseaux, R., Liao, J.C., 2016. Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. Appl. Microbiol. Biotechnol. 100, 4969–4983. https:// doi.org/10.1007/s00253-016-7320-3.

Zhang, W., Zhang, T., Wu, S., Wu, M., Xin, F., Dong, W., Ma, J., Zhang, M., Jiang, M., 2017. Guidance for engineering of synthetic methylotrophy based on methanol metabolism in methylotrophy. RSC Adv. 7, 4083–4091. https://doi.org/10.1039/ C6RA27038G.

Zhao, G., Winkler, M.E., 1994. An Escherichia coli K-12 *tkt*A *tkt*B mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth. J. Bacteriol. https://doi.org/10.1128/jb.176.19.61346138.1994.

This page is intentionally left blank.

Chapter 3 : Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol

Disclaimer: This chapter consists of an accepted manuscript in press, scheduled be published in *Cell* **182** (2020) on August 20, 2020. The figures and tables have been renumbered for ease of reading and compliance of the UCLA thesis format requirements.

Authors contributions:

F.Y.C. (Yu-Hsiao Chen) and J.C.L (Dr. James C. Liao) conceived the study and designed the experiments. F.Y.C. and C.T. (Chao-yin Tsuei, Academia Sinica) performed strain construction and cloning. F.Y.C and H.J. (Dr. Hsin-Wei Jung, Academia Sinica) performed the evolution experiments, DPC extraction, strain isolation and characterization. H.J. extracted genomic DNA for NGS sequencing. C.T extracted RNA and performed qRT-PCR experiments. F.Y.C analyzed all NGS DNA sequencing, RNA-seq and Proteomics data. F.Y.C and J.C.L composed the manuscript. H.J and C.T assisted on preparing the manuscript.

3.1 Summary

Methanol, being electron-rich and derivable from methane or CO_2, is a potentially renewable one-carbon (C1) feedstock for microorganisms. Although the ribulose monophosphate (RuMP) cycle used by methylotrophs to assimilate methanol differs from the typical sugar metabolism by only three enzymes, turning a non-methylotrophic organism to a synthetic methylotroph that grows to a high cell density has been challenging. Here, we reprogrammed *E. coli* using metabolic robustness criteria followed by laboratory evolution to establish a strain that can utilize methanol as the sole carbon source efficiently. This synthetic methylotroph alleviated a heretofore uncharacterized hurdle, DNA-protein crosslinking (DPC), by insertion sequence (IS) mediated copy number variations (CNV) and balanced the metabolic flux by mutations. Being capable of growing at a rate comparable to natural methylotrophs in a wide-range of methanol concentrations, this synthetic methylotrophic strain illustrates genome editing and evolution for microbial tropism changes, and expands the scope of biological C1 conversion.

3.2 Introduction

One-carbon (C1) compound assimilation by microorganisms has emerged as a promising approach in abating climate change (Bogorad et al., 2014; Chen et al., 2018; Cotton et al., 2019; Gassler et al., 2020; Gleizer et al., 2019). Among all C1 compounds, methanol is the most electron-rich in the liquid form and can be derived from greenhouse gases methane or CO_2 (Conrado and Gonzalez, 2014; Kuk et al., 2017; Patel et al., 2019). It also avoids the diffusion barrier compared to those gaseous C1 compounds. In addition, methanol is already an industrial feedstock, thus being potentially attractive as a substrate for bioconversion (Bertau et al., 2014). The native methanol utilization and conversion pathways in natural methylotrophs such as *Methylobacterium extorquen*s and *Bacillus methanolicus* have been well-characterized (Bennett et al., 2018b; Brautaset et al., 2007; Smejkalová et al., 2010). These organisms typically utilize the RuMP cycle or the serine pathway for methanol assimilation (Chistoserdova et al., 2009). In particular, the enzymes involved in the RuMP cycle overlap with those used in the typical sugar metabolism (Fig. 3-1A and Table 3-1), except three enzymes (methanol dehydrogenase, Medh; hexulose-6-phosphate synthase, Hps; 6-phospho-3-hexuloisomerase, Phi). Thus, significant efforts have been reported to convert sugar heterotrophs to methylotrophs for both scientific and industrial interests by overexpressing these three enzymes (Chen et al., 2018; Gonzalez et al., 2018; Meyer et al., 2018; Tuyishime et al., 2018). If this can be achieved, the industrial biotechnology sector can include methanol as an alternative non-food carbon source using a common industrial microbe.

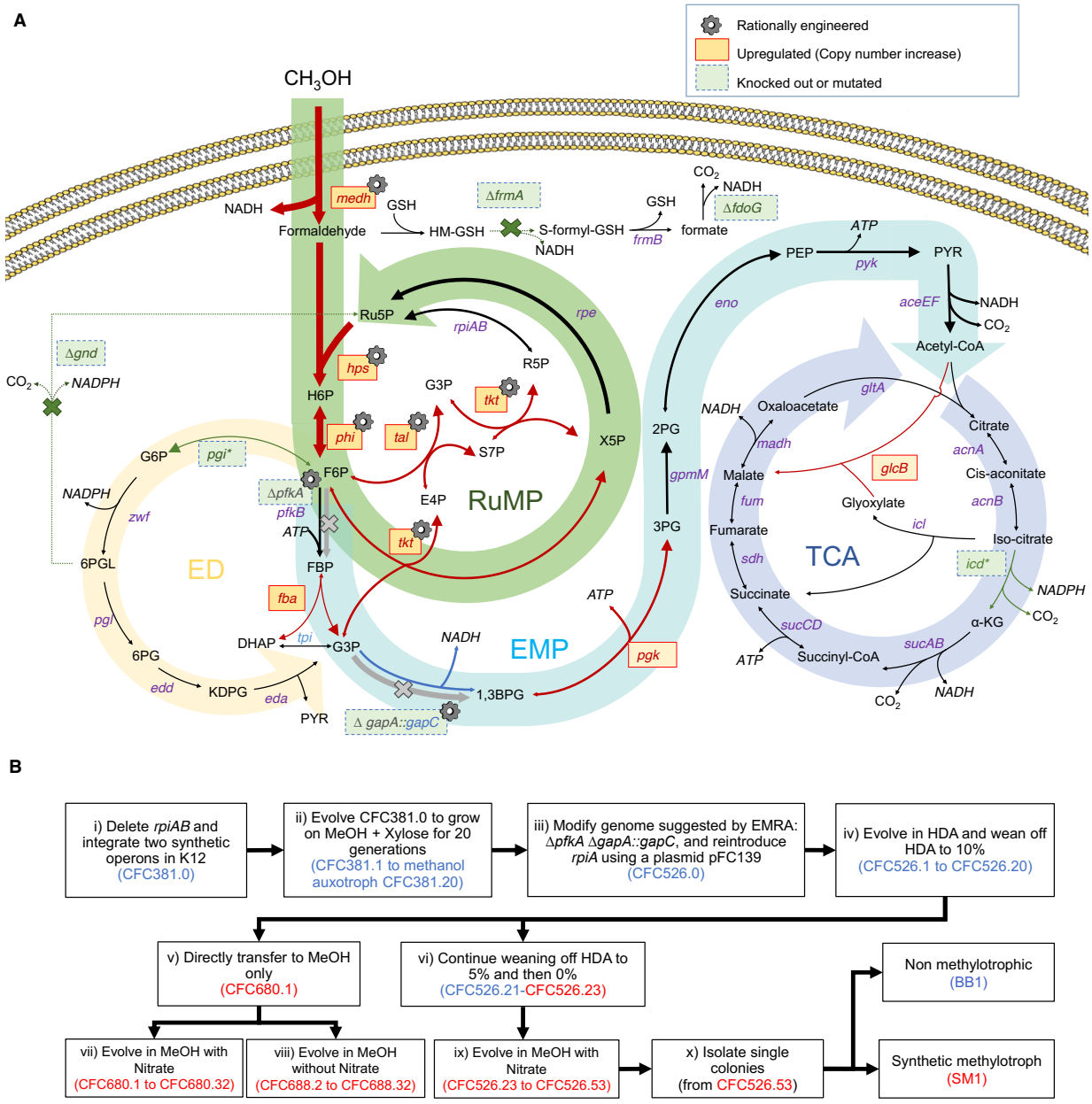


Fig. 3-1. Build and evolve a synthetic methylotrophic *E. coli* strain.

(A) Pathway and mutations relevant to synthetic methylotrophic *E. coli* SM1. The cog icons represent rationally designed and engineered gene modifications. Red solid boxes indicate up-regulated or high copy number genes, while green dashed boxes represent genes knocked out or mutated. **(B)** Flowchart for construction and evolution of synthetic methylotroph. Red letters indicate cultures or strains that grew on methanol as the sole carbon source. Abbreviations are defined in Table S1. See also Table 3-1.

Table 3-1. Metabolites and genes list

Metabolite Acronym	Full metabolite name	Gene	Encoding enzyme
G6P	Glucose 6-phosphate	*zwf*	$NADP^+$-dependent glucose-6-phosphate dehydrogenase
6PGL	6-phospho D-glucono-1,5-lactone	*gnd*	6-phosphogluconate dehydrogenase
6PG	D-gluconate 6-phosphate	*pgl*	6-phosphogluconolactonase
KDPG	2-dehydro-3-deoxy-D-gluconate 6-phosphate	*edd*	Phosphogluconate dehydratase
PYR	pyruvate	*eda*	2-keto-3-deoxygluconate 6-phosphate aldolase
DHAP	Dihydroxyacetone phosphate	*tpi*	Triose-phosphate isomerase
HM-GSH	hydroxymethyl glutathione	*medh*	methanol dehydrogenase
S-formyl-GSH	S-formylglutathione	*frmA*	S-(hydroxymethyl)glutathione dehydrogenase
H6P	Hexulose-6-phosphate synthase	*frmB*	S-formylglutathione hydrolase
F6P	fructose 6-phosphate	*phi*	6-phospho-3-hexuloisomerase
FBP	Fructose-1,6-bisphosphate	*fdoG*	formate dehydrogenase-O
G3P	3-phospho-D-glycerate	*pfkA*	ATP-dependent 6-phosphofructokinase
1,3BPG	1,3-Bisphosphoglycerate	*pfkB*	ATP-dependent 6- phosphofructokinase isozyme
3PG	3-phosphoglycerate	*gapA*	glyceraldehyde-3-phosphate dehydrogenase A
2PG	2-phosphoglycerate	*gapC*	glyceraldehyde-3-phosphate dehydrogenase
PEP	Phosphoenolpyruvate	*pgk*	phosphoglycerate kinase
G3P	Glycerol-3-phosphate	*rpiAB*	ribose-5-phosphate isomerase
E4P	Erythrose 4-phosphate	*gpmM*	cofactor-independent phosphoglycerate mutase
S7P	Sedoheptulose-7-phosphate	*eno*	enolase
R5P	Ribose-5-phosphate	*pyk*	pyruvate kinase
Ru5P	Ribulose 5-phosphate	*aceEF*	pyruvate dehydrogenase

α-KG	Alpha-Ketoglutarate	*tal*	transaldolase
		tkt	transketolase
		rpe	ribulose-phosphate 3-epimerase
		gltA	citrate synthase
		acnA	aconitate hydratase A
		acnB	aconitate hydratase B
		icd	isocitrate dehydrogenase
		sucAB	2-oxoglutarate dehydrogenase
		sucCD	succinyl-CoA synthase
		sdh	succinic dehydrogenase
		fum	fumarase
		madh	malate dehydrogenase
		glcB	malate synthase
		icl	isocitrate lyase

Despite initial successes in engineering sugar heterotrophs to assimilate methanol, it has not been possible to convert such heterotrophs to methylotrophs that utilize methanol as the sole carbon and energy source efficiently. Reported examples either required other carbon sources or nutrients in the medium to support growth (Chen et al., 2018; Gonzalez et al., 2018; Meyer et al., 2018; Tuyishime et al., 2018), or demonstrated minimal growth with a doubling time of 55 hours and a maximum OD_{600} of 0.2 with methanol alone (Kim et al., 2020). Apparently, successful expression of three heterologous genes is insufficient to turn *E. coli* into a methylotroph. Here, we report a major problem involving DPC that prevented *E. coli* from growing in methanol as the sole carbon source, and how genome editing, copy number variations (CNVs), and mutations introduced by laboratory evolution overcame this hurdle, resulting in a synthetic methylotrophic *E. coli* that grows to a high optical density (OD) efficiently with a doubling time of 8 hours.

3.3 Results

3.3.1 Methanol auxotrophy as a starting point

To develop a synthetic methylotroph, we began from our RuMP cycle-based methanol auxotrophy strategy (Chen et al., 2018) (Fig. 3-1B and Fig. 3-2A). It calls for a disruption of the pentose phosphate pathway by deleting *rpiAB* and installing the methanol utilizing genes (*medh, hps, phi*), such that the cell can grow on methanol plus xylose in minimal media, but not on xylose alone. Thus, methanol assimilation can be used as a selection pressure during evolution. Instead of the previously established BL21 strain, we opted to reconstruct an auxotrophic *E. coli* BW25113 *ΔrpiAB* strain for its higher success rate of genome manipulation. We integrated two synthetic operons (Fig. 3-2B) for stable expression, designated as CFC381.0. The first operon consists of the three heterologous genes, *medh* (CT4-1, engineered *from Cupriavidus necator* (Wu et al., 2016)*, hps* (from *Bacillus methanolicus*) and *phi* (from *Methylobacillus flagellatus*). The second operon includes the same *medh* and *phi*, but different *hps* (from *Methylomicrobium buryatense 5GB1S*) (Fig. 3-2C) along with *tkt* (encoding transketolase from *Methylococcus capsulatus*) and *tal* (encoding transaldolase from *Klebsiella pneumoniae* (Lin et al., 2018)). As enzymes from different organisms differ in Km, we intentionally expressed various isofunctional enzymes simultaneously to maximize the flexibility of metabolic flux balance. After 20 liquid transfer cycles, which we define as "passages", the evolved strain CFC381.20 could grow from OD_{600} 0.1 to OD_{600} 1.0 in a minimal medium containing 400 mM of methanol and 20 mM xylose in 48 hours (Fig. 3-2D and 3-2E), but could not grow without methanol. As approximately 3 to 4 generations elapsed per passage, CFC381.20 demonstrated the methanol auxotrophy phenotype as desired within 80 generations.

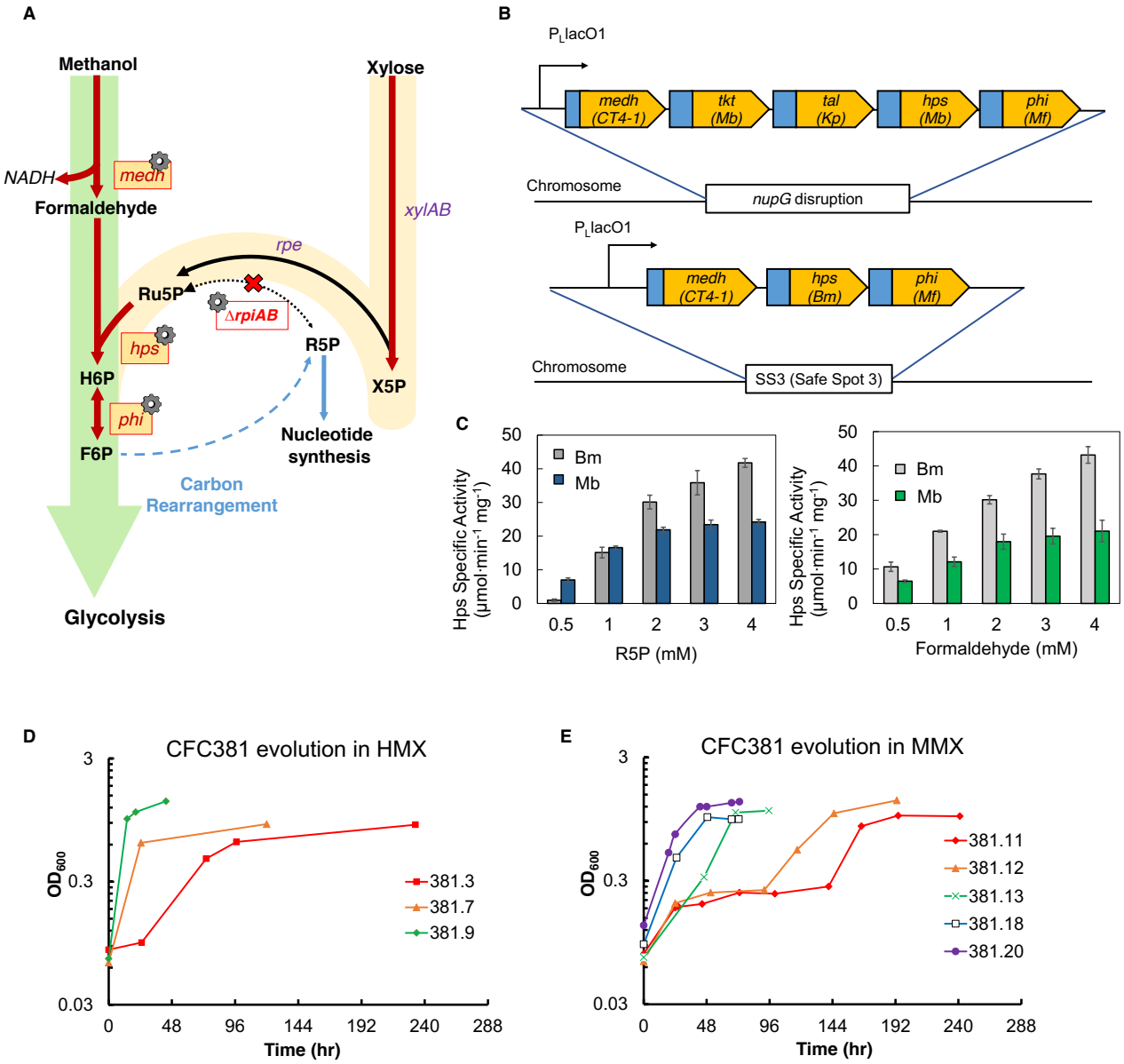


Fig. 3-2. Construct and evolve a methanol auxotroph strain.

See figure legends next page.

(A) Methanol auxotrophy scheme. (B) Two synthetic operons integrated in CFC381.0. “SS3” refers to a safe spot for genome integration. (C) Bioprospecting Hps. Other than Bacillus methanolicus Hps, we bioprospected and found another Hps from Methylomicrobium buryatense 5GB1S. We tested specific activity with a coupled assay with rpiA, feeding a fixed amount (2 mM) of either formaldehyde or R5P. Noticeably, the Hps (Mb) has higher activity under low concentrations of R5P, though performs worse in reacting with formaldehyde. The bars represent biologically independent triplicate mean value with error bars as the standard deviation. (D) Growth curve showing the evolution of CFC381 in in HDA media with 400 mM MeOH and 20 mM xylose (HMX). (E) Growth curve showing the evolution of CFC381 in MOPS with 400 mM MeOH and 20 mM xylose (MMX), after evolution in HMX for 10 passages.

Whole genome sequencing of CFC381.20 (Table 3-2) revealed a 4bp-insertion in the *frmA* gene. The inactivation of *frmA,* which is responsible for detoxifying formaldehyde to formic acid, was also seen previously in similar efforts (Chen et al., 2018; Gonzalez et al., 2018; Meyer et al., 2018). This suggests that the formaldehyde flux must be directed to biosynthetic pathways for efficient methanol-dependent growth. Other significant mutations included truncation of *gnd* (encoding 6-phosphogluconate dehydrogenase, Gnd) and a frameshift in *fdoG* (encoding formate dehydrogenase. Gnd forms a non-productive cycle with Hps, Phi, Pgi and Zwf, with a net reaction to convert formaldehyde to CO_2 and NADPH. Similarly, *frmA* and *fdoG* drains formaldehyde to CO_2 while producing excess NADH. These mutations indicated that through evolution the methanol auxotrophic strain reduced competing flux away from the productive RuMP cycle for efficient biosynthesis and biomass accumulation. This methanol auxotroph strain demonstrated that the methanol assimilation branch of the RuMP cycle was functional. However, the replenishment of ribulose-5-phosphate (Ru5P) was still supplied by xylose because the regeneration pathway was disrupted by the *rpiAB* deletion.

Table 3-2. Genotype of strains and cultures

	Codon Change	Non-coding Region	Indel Codon Change	Large Genome Truncation	Copy number Variation (CNV)
CFC381.0	n/a	n/a	n/a	Δ*rpiA* Δ*rpiB*	*n/a*
CFC381.20 (Same as CFC381.0 except the list to the right)	*araG* (I275R) *rpoA* (G315C) *xylR* (K320Q) Low frequency SNVs: *cybB* (Y69S) *fhu* (M484G) *ydhB* (A45G) *ydhB* (V46G) *ydhB* (P47A)	n/a	*fdoG* (2,194_del G) [Frame shift] *frmA* (383_ins CCCG) [Frame shift] *ugd* (1-51_del) [Early stop codon] *smf* (198_ins IS2) [Early stop codon]	2,093,044-2,095,229 *ugd* N-terminal and operon *gnd*, *wbbL* deletion *stfP*_291-*stfE*_3 transversion	70k (*yggE* to *yghO*) duplicate
CFC526.0 (Same as CFC381.20 except the list to the right)	n/a	n/a	n/a	Δ*gapA::gapC* Δ*pfkA*	n/a
SM1 (Same as CFC526.0 except the list to the right)	*rpoC* (S733F) *proQ* (E12*) *icd* (D398E) *icd* (D410E) absence of the low frequency SNVs	*gltA* upstream (IS2 insertion at 750,112) [Disrupts promoter] *ptsH* upstream (IS2 insertion at 2,527,069) [Disrupts promoter]	*pgi* (705_del GTTGCAAAACAC) [Codon:236-239_delVAKH] *ptsP* (1673_ins C) [Frame shift] *relE* (202_ins IS2) [Early stop codon] *elfG* (949_ins IS3) [Early stop codon] *yhcM* (904_ins IS2) [Early stop codon]	1,191,676-1,206,868 cryptic prophage e14 deletion	70k (*yggE* to *yghO*) 4 fold 130k (*rrsA* to *rrlB*) duplicate 240k (*yqiG* to *smf*) duplicate
BB1 (Same as CFC526.0 except the list to the right)	*dacC* (V298M) *glpR* (T22I) *rpoA* (A267T)	*yihL* upstream (IS4 insertion at 4053628) [Disrupts promoter]	*alsC* (59_ins A) [Early Stop Codon] *frmA* (384_del T) [Back to in-frame] *ompF* (106_del T) [Frame shift]		70k (*yggE* to *yghO*) 1 fold 7k (osmC to *dosP*) 85 fold

3.3.2 Rational design and evolution for creating a synthetic methylotroph

Next, we attempted to close the RuMP cycle by transforming a plasmid (pFC139) that carries an RBS library expressing *rpiA*, so that CFC381.20 could utilize methanol as the sole carbon source. Unfortunately, the strain could only acquire limited growth advantage after series of evolution in the presence of methanol while supplying additional carbon sources, such as amino acids or xylose. We then hypothesized that kinetic traps in the system curtailed the flux during methanol assimilation. To identify them, we used Ensemble Modeling for Robustness Analysis (EMRA) (Lee et al., 2014; Rivera et al., 2015), which examines a large number of models with different kinetic parameters, and perturbs them by varying enzyme V_{max}, which is largely proportional to the expression levels. It then detects the models that become unstable after increasing or decreasing each V_{max} up to tenfold and reports the percentage of stable models after each perturbation. If the system becomes unstable sharply after a small perturbation in the V_{max} of an enzyme, then this enzyme may be involved in a kinetic trap. This analysis provides a qualitative assessment to suggest enzymes that require up or down expression in order to avoid system instability caused by kinetic traps.

Results revealed that high activities of phosphofructokinase (Pfk) and glyceraldehyde 3-phosphate dehydrogenase (Gapdh) tend to destabilize the system by diverting the flux away from the RuMP cycle (Fig. 3-3) and preventing the replenishment of the cycle intermediates. Accordingly, we reduced these enzyme activities by knocking out *pfkA,* which accounts for 90% of the Pfk activity (Kotlarz et al., 1975), and replacing *gapA* with *gapC* from *E. coli* BL21 (Wu et al., 2012) that possesses around 40% of K12 BW25113 *gapA* activity (Fig. 3-4A). The resulting strain, CFC526.0, was then subject to laboratory evolution with different strategies of nutrient weaning (Fig. 3-1B).

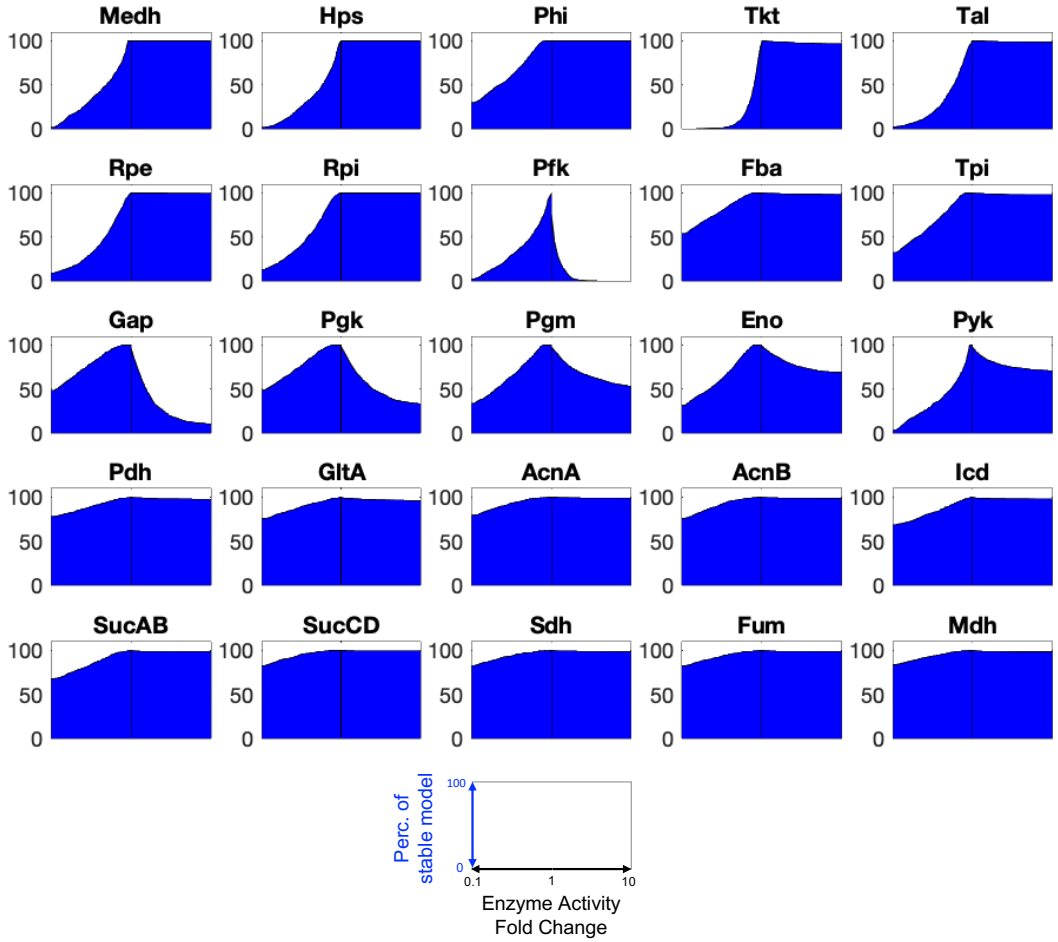


Fig. 3-3. Ensemble-Modelling Robust Analysis (EMRA) of RuMP-EMP-TCA cycle.

The x-axis represents the fold change of a specific enzyme activity, while the y-axis refers to the ratio of the 100 parameter sets that are robust at the specific perturbed enzyme activity. Results indicate that high level expression of *pfk*, *gapA*, *pgk*, *gpmM*, *eno*, and *pyk* may cause a system stability problem because of kinetic traps. This result indicates that high activities of Pfk and enzymes in the lower glycolysis may be detrimental to the system. Hypothesizing that *E. coli* possesses high glycolytic activity natively, we knocked out *pfkA*, and knocked down *gapA*, which is the first gene in the lower glycolysis that is unstable, by replacing it with a functional BL21 *gapC*.

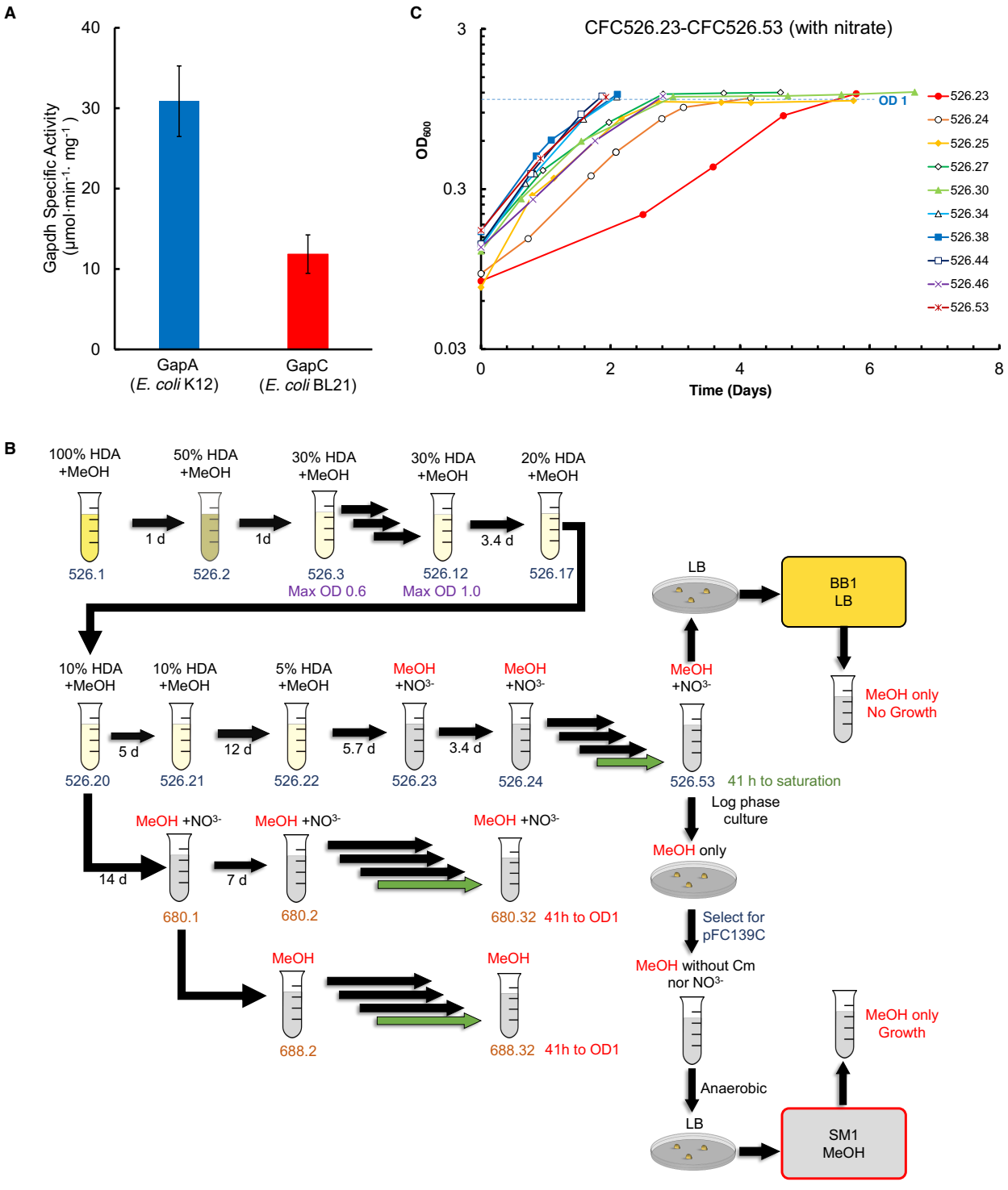


Fig. 3-4. Evolving a synthetic methylotrophic strain

See figure legends next page.

Fig. 3-4, cont. (A) *E. coli* K12 BW25113 GapA activity and BL21 GapC specific activity. The two enzymes were his-tag cloned and purified before enzyme assay. n=3, error bars represent the standard deviation. **(B)** Detailed flowchart of the entire evolution process to enable *E. coli* to grow on methanol as the sole carbon source. Note that aside from the methylotrophic strain SM1, we also isolated a non-methylotrophic strain BB1 in the final mixed culture that can grow on methanol. **(C)** Growth curve that shows the evolution of CFC526.23-53 in 400 mM methanol with nitrate.

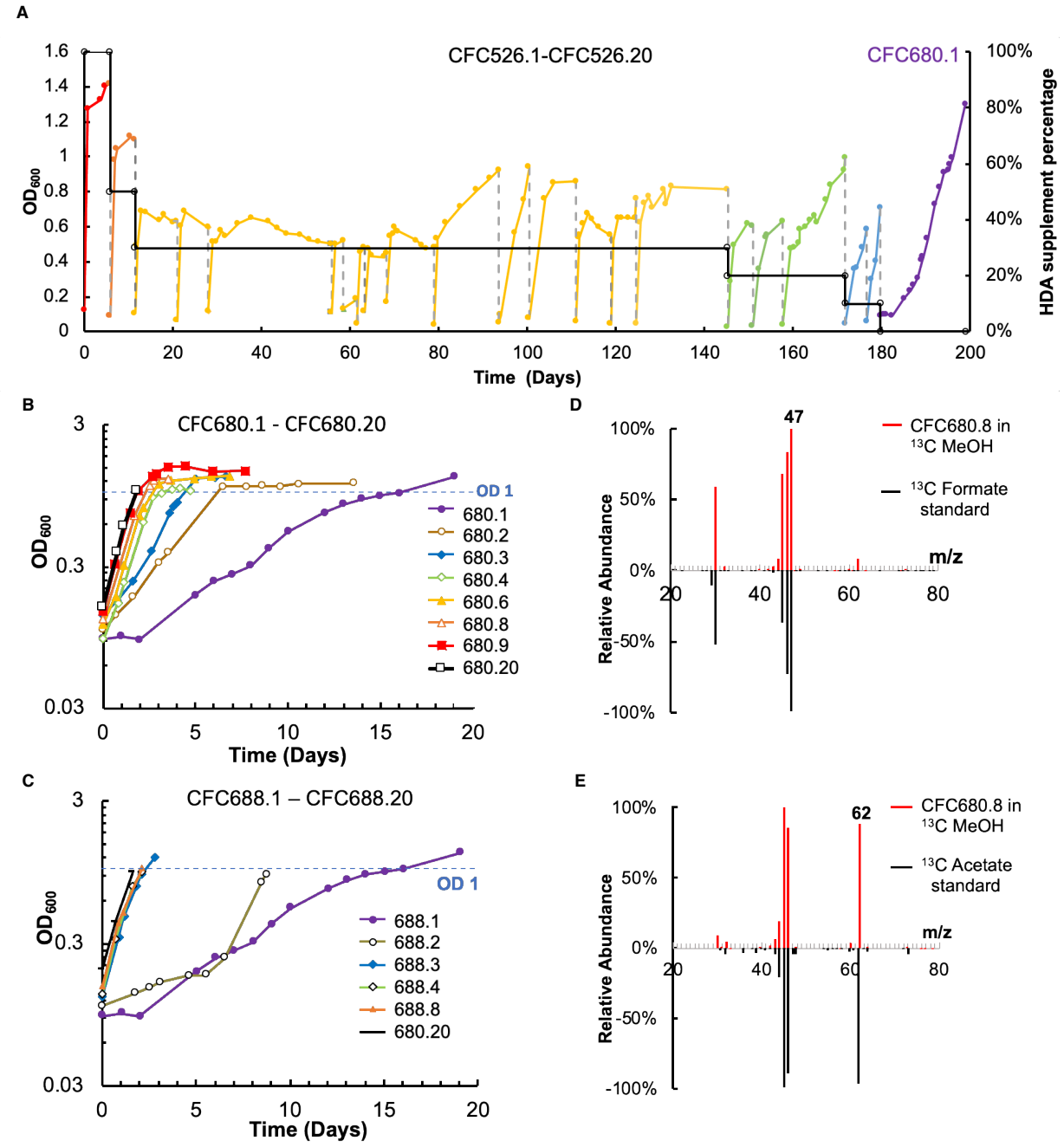


Fig. 3-5. Evolution results and verification of *E. coli* growing on methanol as the sole carbon source.

See figure legends next page.

Fig. 3-5, cont. (A) The evolution trajectory (step iv in Fig. 1B) of CFC526.1-20. The media consisted of a decreasing portion of an amino acid mixture (HDA) in MOPS, while keeping methanol at 400 mM. The last passage (purple line) was in methanol only (step v in Fig. 1b). The black solid line represents HDA percentage in the media. The other colors represent growth curve of cultures in different media **(B)** Growth curves of CFC 680.1-20 throughout evolution in methanol MOPS (MM) media with nitrate. **(C)** Growth curve that shows the evolution process of CFC688.1-20 culture with serial inoculation in MM without nitrate. **(D) and (E)** ^{13}C labelling patterns of acetate and formate from CFC680.8. The red lines represent the sample, while the black line illustrates a ^{13}C standard.

Specifically, we grew CFC526.0 in a medium containing methanol and a defined semi-minimal medium, Hi-Def azure (HDA) that contained amino acids. The HDA amount was sequentially reduced and replaced by the methanol MOPS (MM) minimal medium until the culture could grow on methanol as the sole carbon source. Extra vitamins were supplied for better cell metabolism. Nitrate was initially supplied as an extra electron acceptor in addition to oxygen, since methanol is an electron-rich substrate and oxygen transfer may be limiting in shaking-flasks. After about 180 days and 21 passages, the culture could finally grow on methanol without any amino acid supplement (Fig. 3-5A and Fig. 3-4B). This initial methylotrophic culture CFC680.1 that grew solely on methanol required 20 days to grow to saturation at OD_{600} = 1. After 20 more passages, CFC680.20 grew to OD_{600} = 1 within 41 hours (Fig. 3-5B). We evolved the culture without nitrate and generated a culture CFC688.20 that could grow without nitrate to reach a similar growth rate (Fig. 3-5C). Independently, we also obtained another methanol-growing strain CFC526.23 by employing a slower nutrient reduction strategy and evolved it to obtain CFC526.53 (Fig. 3-4C).

To ensure that all metabolic products were derived from methanol, ^{13}C labeling experiments were performed. CFC680.8 was passed six times in MM with ^{13}C methanol until all isotopes reach a steady state. As expected, we found that acetate was double-labeled, while formate was single-labeled (Fig. 3-5D and 3-5E). The isotope labeling experiments provided solid evidence that methanol was the sole carbon source for growth.

3.3.3 The DNA-protein crosslinking problem

One noticeable phenotype of these methylotrophic cultures was an exceedingly long lag phase (up to 20 days) if the culture was inoculated from the stationary phase (Fig. 3-6A), but not from the log-phase. Similarly, colonies on a methanol minimal medium plate could not proliferate in a liquid minimal medium. Although microorganisms do exhibit a lag phase when inoculated from a stationary-phase culture, these synthetic methylotrophic *E. coli* cultures seemed to go through a "point-of-no-return," beyond which the exceedingly long lag phase appeared. After monitoring the viability of the cells at stationary phase by flow cytometry, we found that up to 10% of the cells were dead (Fig. 3-6B). We speculated that the strain may have experienced toxicity from intermediate metabolites, mostly likely formaldehyde. This was foreseeable as the inactivation of *frmA* inherited from the auxotrophy strain hindered the entire formaldehyde detoxification pathway.

Surveying the broad spectrum of biomolecules susceptible to formaldehyde reactions, we hypothesized that DNA-protein crosslinking (DPC) was the most likely cause of cell death which may lead to the disruption of DNA replication, transcription, translation and protein function (Stingele and Jentsch, 2015). To test the hypothesis, we purified DPC products from methanol-growing cultures by a modified DNA extraction method (Qiu and Wang, 2009). After de-crosslinking the extractant, we analyzed the protein portion by SDS-PAGE. Results suggested that DPC did occur as the culture reached stationary phase (Fig. 3-7A). We then imaged the isolated DPC products using transmission electron microscopy (TEM) and unveiled the severity of formaldehyde crosslinking (Fig. 3-6C). Typically, DNA could not be seen with negative staining without coating by proteins such as cytochrome C, as the DNA string is too thin for TEM observation (Serwer, 1978).

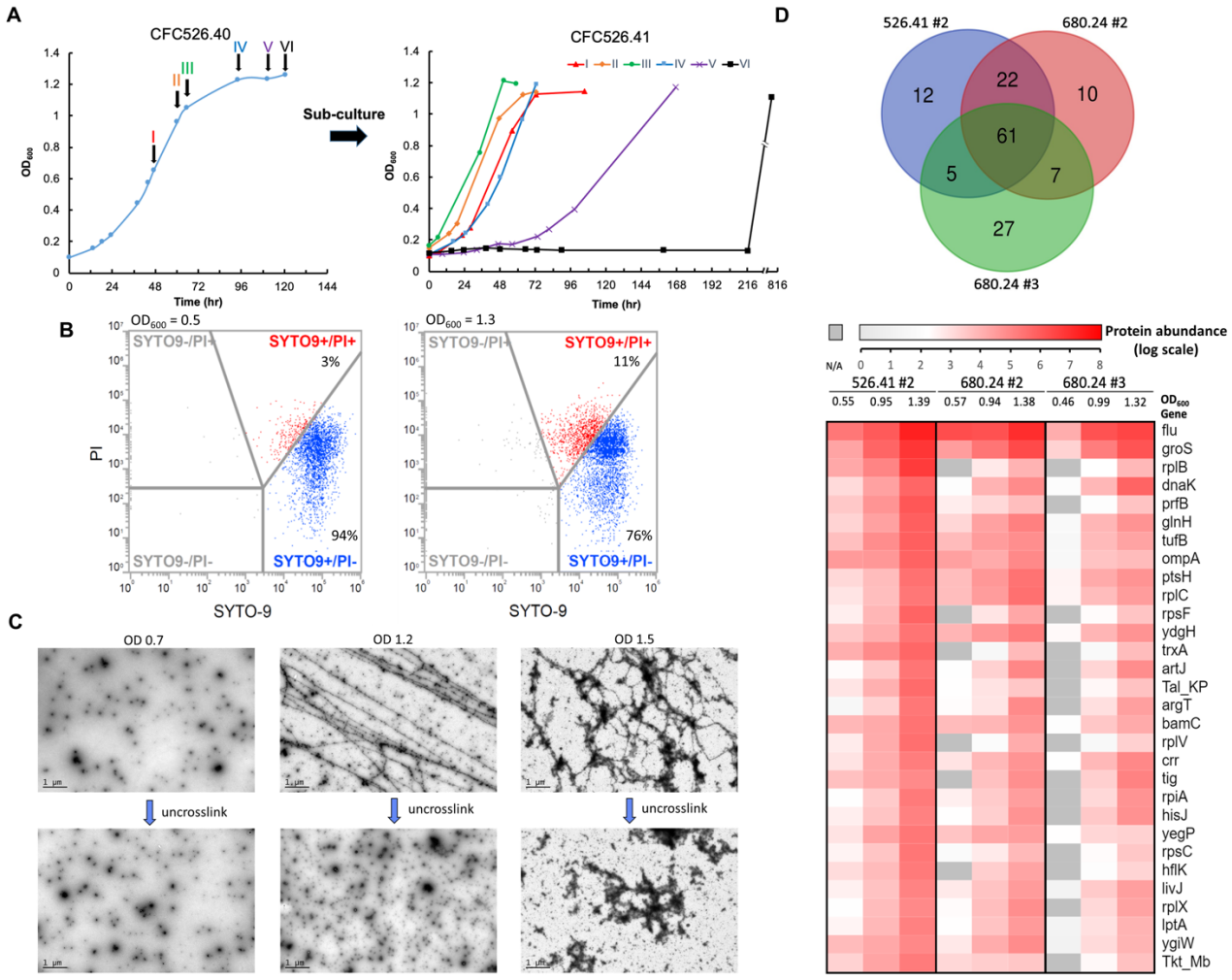


Fig. 3-6. DPC products identified in methylotrophic *E. coli* cultures.

(A) Extended lag phase seen when *E. coli* is subcultured in methanol media MM at stationary phase. CFC526.40 was being passed from time point I~VI and showed various levels of time lag. Note that starting from time point V, the strain experienced a serious lag phase for growth in methanol. **(B)** Flow cytometry-based cell viability test. All cells are stained with SYTO-9, while propidium iodide (PI) only stains dead cells when the cell membrane can be penetrated. The coordinates were defined by control samples, including healthy *E. coli* cells and ethanol-treated dead cells. **(C)** TEM images of DPC products extracted from different growth stages of CFC526.41, and their uncrosslinked forms. **(D)** Quantitative proteomics analysis of the proteins from uncrosslinked DPC samples from CFC 526.41 and CFC680.24. Among 6 samples, CFC526.41#2, CFC680.24 #2 and CFC680.24 #3 were selected for analysis based on their similar growth trends. 30 out of 61 common top hits ranked by average abundance were presented.

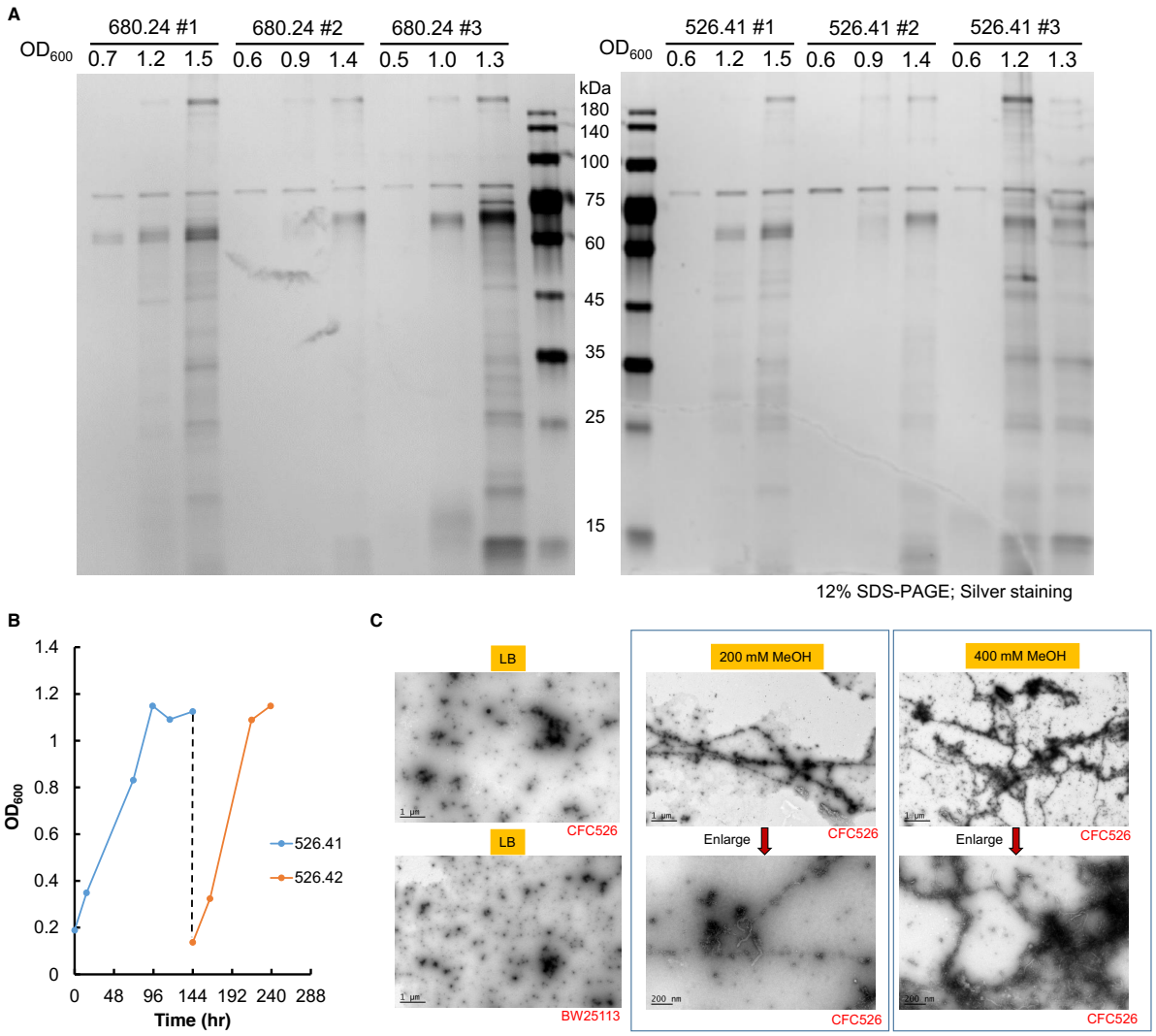


Fig. 3-7. Further Characterization of DPC in methanol growing strains.

(A) SDS-PAGE analysis of proteins extracted from DPC. There is a clear trend that DPC accumulates when OD_{600} increases. Although the pattern of the bands looks similar, the amount of DPCs detected varies among samples. **(B)** Growth curve of CFC526.41 and its offspring CFC526.42 growing in MOPS medium with 200 mM methanol. No lag phase observed after inoculation of 562.42. **(C)** TEM images of DNA/DPCs extracted from *E. coli* cultures grown in different conditions. The LB 526 and LB BW25113 samples are controls for the experiment. Note that lower MeOH concentrations (200 mM) alleviated DPC.

As expected, during the log phase, only free protein particles were observed, most likely attributable to protein leftovers during the salting-out process. In contrast, the DPC level increased when the culture reached the stationary phase (OD_{600} 1.2), causing the entire DNA string to be visible as protein was coated to DNA due to formaldehyde crosslinking. Moreover, we could see protein aggregates along the DNA string. At OD_{600} 1.5, formaldehyde-induced crosslinking became extremely severe, where the DNA started to form a web-like structure by DNA-Protein-DNA crosslinking or even DNA-DNA crosslinking. Noticeably, DNA strings disappeared when we heated and de-crosslinked DPCs, ruling out the possibility of DNA-protein nonspecific binding or image overlapping. DPC was less severe when the cells were growing in lower methanol concentrations (Fig. 3-7B and 3-7C).

We then conducted quantitative proteomics to reveal that more than 500 proteins were crosslinked with DNA. The common 61 hits of the 100 proteins with the highest abundance from 3 independent samples were then visualized with a heat map (Fig. 3-6D and Fig. 3-8). As expected, there is a clear trend of increasing crosslinked proteins as the culture enters the stationary phase, and the protein abundance in DPC products in the same culture could differ up to 7 orders of magnitude between the log phase and late stationary phase. Moreover, gene ontology analysis of the 61 proteins suggested that DPCs mainly consisted of ribosomes and outer membrane proteins, while several metabolic enzymes were also identified, such as Medh, Tkt, Tal, AceA, Eno, Pyk and Pgk. Malfunction of these proteins may cause cell death due to outer membrane porin induced programmed cell death (Noor, 2015), or metabolic flux imbalance. Moreover, the strong presence of ribosomes also suggested that transcription and translation were heavily impacted by DPC as well. The accumulation of DPC could explain why the culture exhibits an exceedingly long lag

phase when inoculated from a stationary phase culture, and may also shed light on the difficulty of evolving a non-methanol-utilizing bacterium to grow on methanol as the sole carbon source.

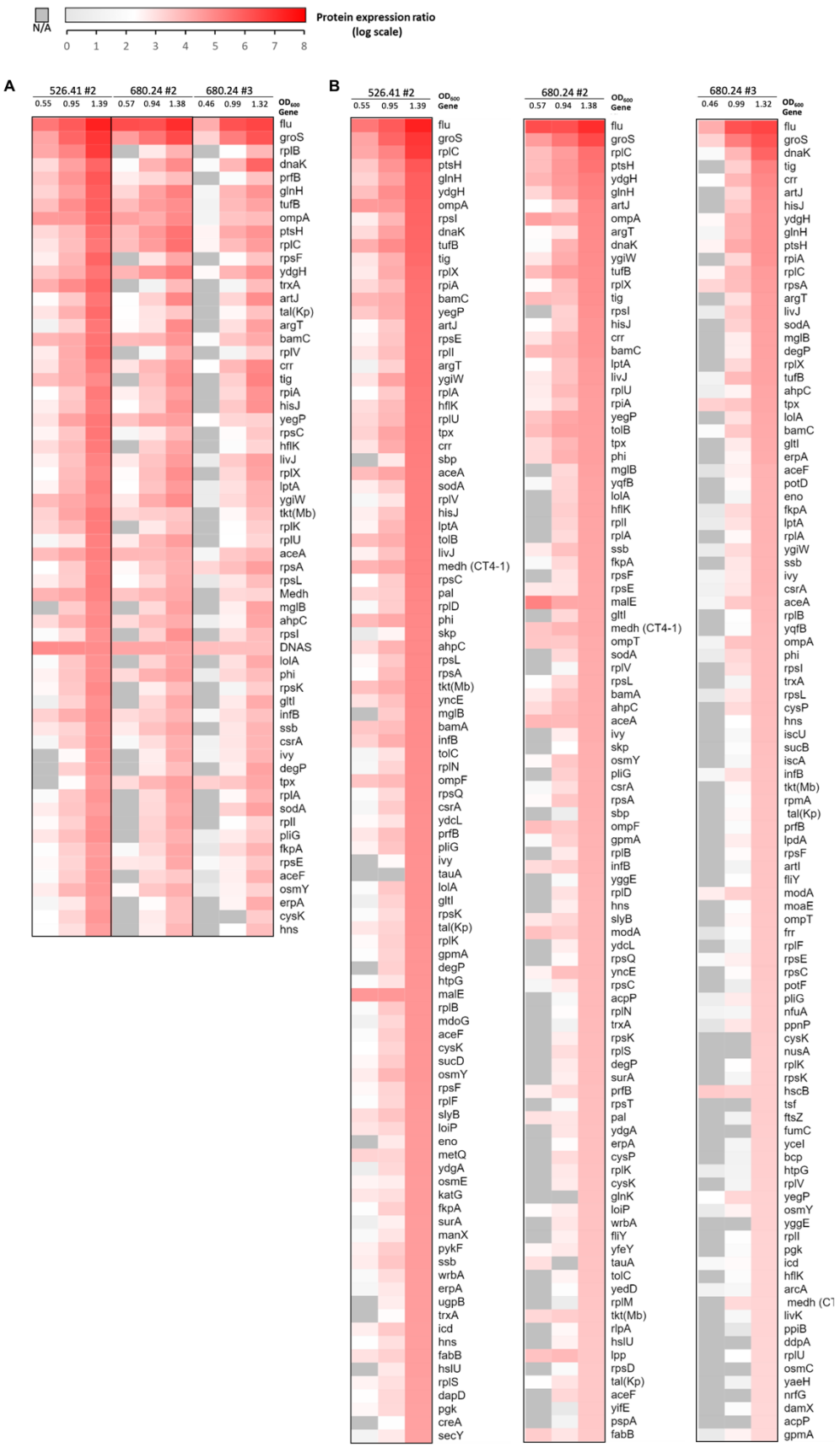

N/A
Protein expression ratio (log scale)
0 1 2 3 4 5 6 7 8
A
526.41 #2
680.24 #2
680.24 #3
OD600
Gene
B
526.41 #2
680.24 #2
680.24 #3

Fig. 3-8. Detailed Proteomics data of proteins extracted from DPCs.

(A) Complete heat map of the common top 61 hits. The map is ranked by average protein abundance at stationary phase. Note that the deoxyribonuclease (DNAS) entry is an externally added enzyme used for DNA clean up and an internal standard. **(B)** Individual top 100 hits. The DNAS data is omitted.

3.3.4 Genome sequencing revealing sub-populations in evolved cultures

Another phenomenon identified was that the cultures evolved to grow on methanol as the sole carbon source initially failed to grow in the same medium after passing through Luria-Bertani (LB) rich medium. This observation implied that the sub-populations emerged during evolution and were enriched in different media. To determine how CFC526.0 evolved to grow in methanol, we sequenced 18 evolved cultures in total along the evolution process (Fig. 3-9 and Fig. 3-10A). Results showed that some mutations appeared but then vanished within a few passages. Along the evolution line, insertion sequence element 2 (IS2) was inserted upstream of two genes, *gltA* and *ptsH* that distances their promoter away from the open reading frame. Accordingly, the TCA-cycle activity may be impeded, while the *ptsH* encoded Hpr protein may be insufficiently expressed, causing a disruption in the phosphotransferase system (PTS). Other mutations included a 12-bp in-frame deletion in *pgi* and truncated *ptsP* and *proQ*. Noticeably, the heterologous genes in the two operons we integrated, which include *medh*, *hps, phi, tkt and tal*, remained unchanged during evolution.

We also identified that the evolved mixed cultures had three high coverage regions flanked by IS elements in their chromosomes: a 70k region spanning from *yggE* to *yghO* (Fig. 3-10B) that contains many glycolytic genes and a synthetic operon P_L*lac*O_1:: *medh-tkt-tal-hps-phi* in the RuMP pathway, a 7k region encoding the dipeptide transporter operon (*ddp*) (Fig. 3-10C), and an 130k region from *rrsA* to *rrlB* containing several 16S RNAs. The high coverage implies that the cells may have increased expression of genes in those regions.

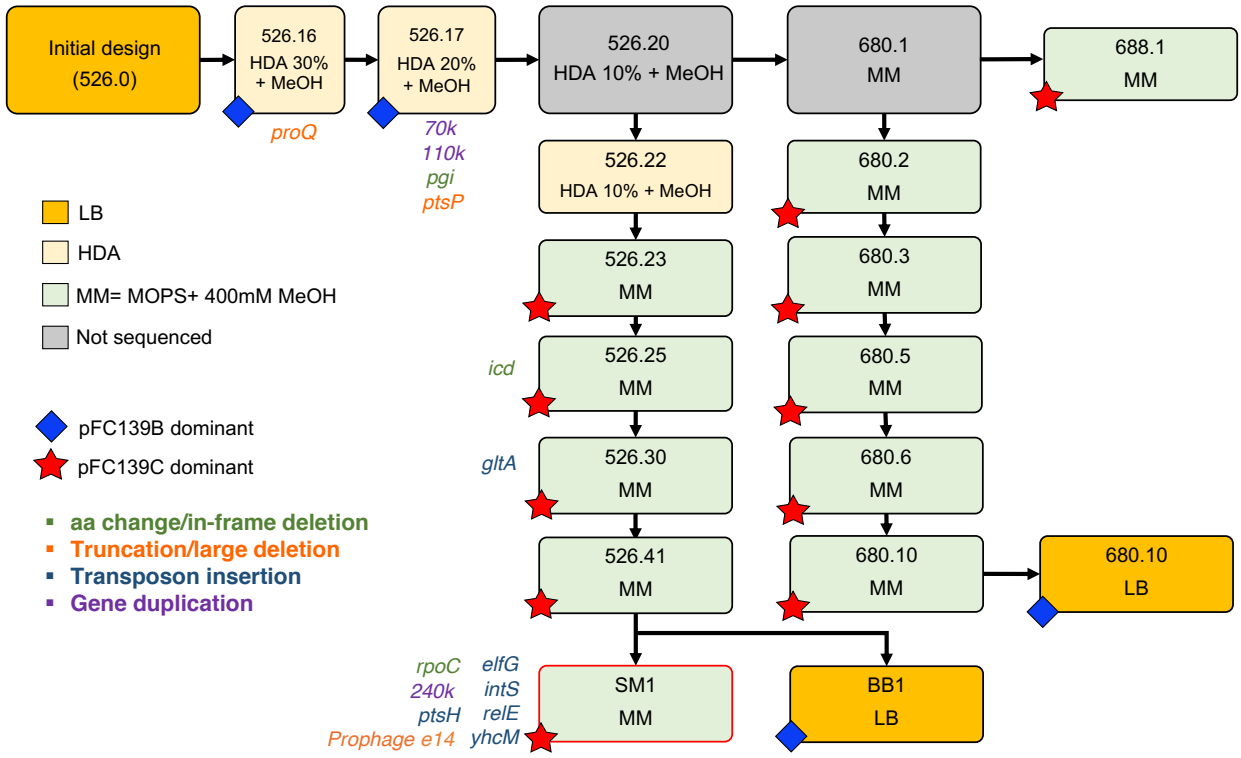


Fig. 3-9. Strain characterization of methylotrophic *E. coli*.

Relationships between evolution cultures that are sequenced by Illumina Miseq/ Hiseq are shown. All strains on this chart except strains in grey boxes were sequenced by Illumina Miseq/ Hiseq. Only mutations that are still preserved in SM1 were annotated. See also Fig. 3-9, 2-13 and 3-15.

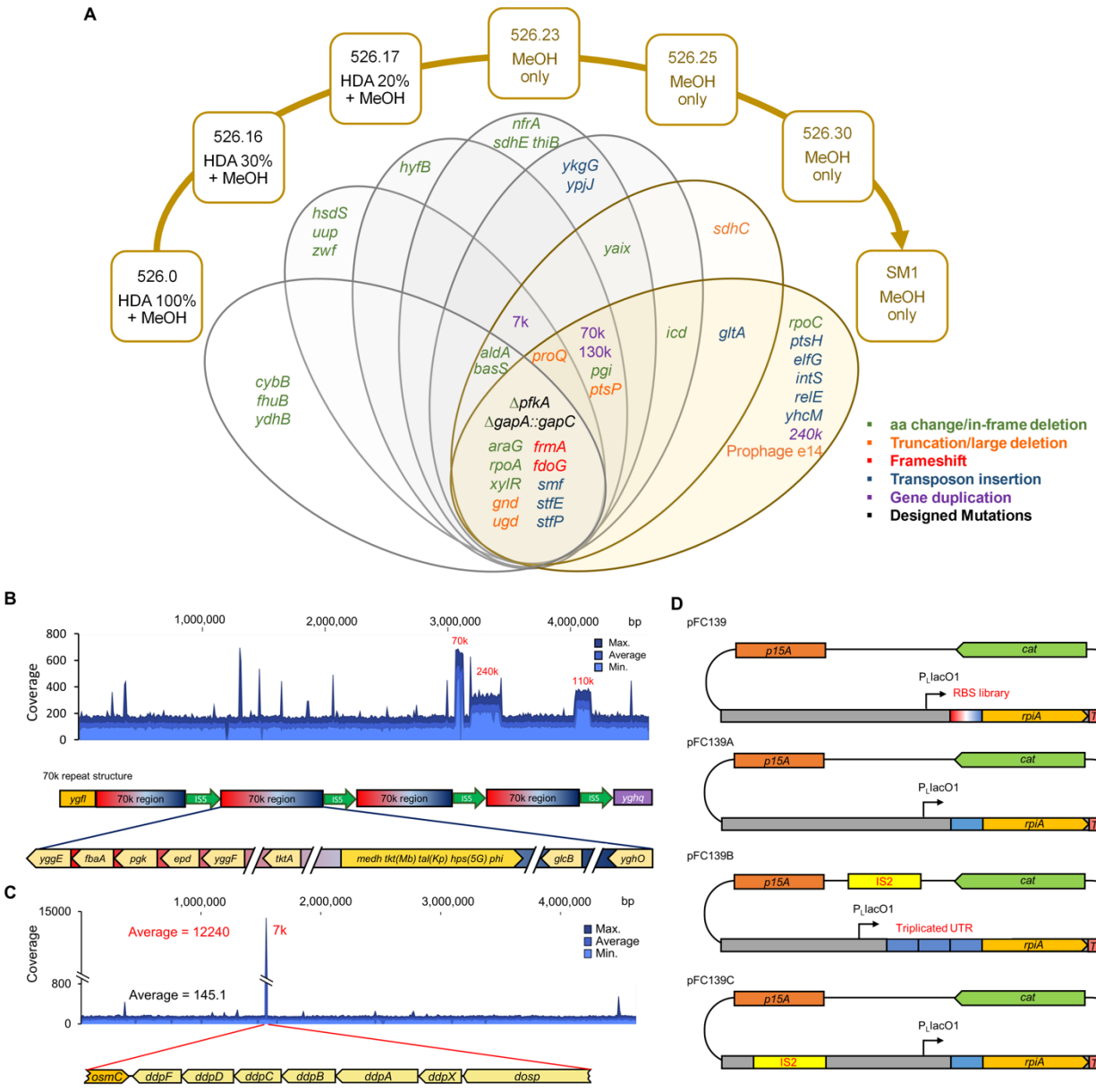


Fig. 3-10. Genomic analysis of Methylotrophic *E. coli*.

See figure legends next page.

(A) Venn diagram of mutations of CFC526 along the laboratory evolution process. To reduce complexity, 4 strains (526.23 to SM1) capable to grow in methanol were selected to represent the genomic changes. Single nucleotide variations (SNVs) that are higher than 30% are reported in the graph. The notation 7k, 70k, 130k, and 240k refer to a region spanning the respective size with high copy numbers. Green text: amino acid change or in-frame deletion; Orange text: truncation or large deletion; Red text: frameshift; blue text: transposon insertion; Purple text: gene duplication; Black text: designed mutation. **(B)** Genome structure of SM1. The top part shows Illumina Hiseq mapping coverage of SM1, while the bottom presents a 70k-tandem repeat in SM1 derived from Pacbio and Nanopore sequencing. Some important metabolic genes including a synthetic operon encoding RuMP cycle genes are illustrated. **(C)** Genome structure of BB1. The 7k region including the *ddp* operon shows about 84-fold increase in read coverage from the Hiseq mapping. **(D)** Schematics of the original designed plasmid pFC139 with a *rpiAB* library, and mutated plasmid pFC139A, B, C emerged during the evolution. See also Fig. 3-13, and Table 3-6.

The plasmid sequence also showed three different versions (Fig. 3-10D): one (pFC139A) that contained a specific RBS from the library; one (pFC139B) that contained a triplicated untranslated region (UTR) upstream to *rpiA*, and an IS2 insertion between the p15A replication origin and the *cat* gene; yet another one (pFC139C) that contained the same RBS as pFC139A, and an additional inserted IS2 before the promoter of *cat* gene.

To evaluate the genome variation through evolution, we found that the copy number of the 70k region gradually increased up to 5.6 copies (Fig. 3-11A). Meanwhile, plasmid pFC139A and pFC139B dominated at the early stage of evolution, but pFC139C eventually dominated at the end of the evolution (Fig. 3-11B).

The coherent increase in the copy number of the 70k region, pFC139C, and some single nucleotide variations (SNVs) implies that there are two main sub-populations in the evolved CFC526 and CFC680 culture series: one real synthetic methylotrophic strain (SM1) containing pFC139C and the 70k multicopy region (Fig. 3-10B), and the other non-methylotrophic strain (BB1) containing pFC139B and the 7k multicopy region but not the 70k repeated region (Fig. 3-10C).

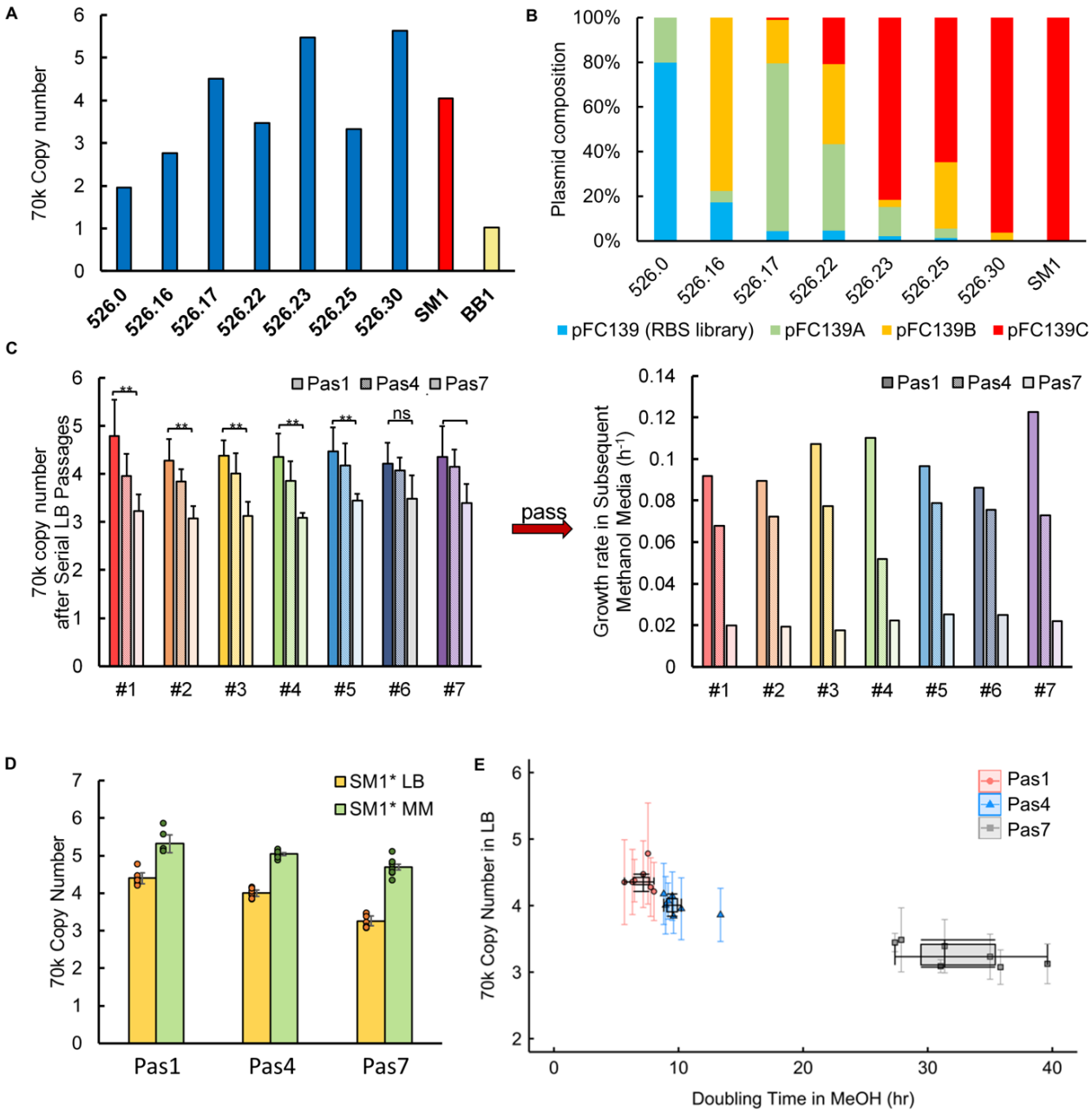


Fig. 3-11. Copy number and plasmid variation in methylotrophic *E. coli.*

(A) Copy number of the multiplicated 70k gene of cultures throughout the evolution process, derived from Illumina Miseq/ Hiseq coverage data. **(B)** Estimated plasmid composition variation in evolved cultures. The plasmids are categorized into the following: pFC139A, pFC139B, and pFC139C, and the rest of the original pFC139 with RBS library. **(C)** 70k region copy number

dynamics experiment. 7 colonies of perturbed SM1 were then picked and were regarded as individual biological repeats (See methods). The colonies were inoculated into LB and was recorded as “Pas1”. They were then passed 3 more times in LB to “Pas4” and another 3 times to “Pas7”. They were then inoculated into MM to calculate growth rates. The copy number of the 70k region in LB was determined by digital PCR. Error bars are calculated from the mean and SD from sampling 4 genes in the 70k region. The statistical significance between Pas1, Pas4, Pas7 was determined by a t-test, n=4. $**p<0.01$, $*p<0.1$, ns= no significance. **(D)** 70k region copy number comparison between LB cultures and their subsequent MM culture. n=7 **(E)** 2D-box plot overlaid with scatter plot. The box plot values were calculated by doubling time in methanol and average values of copy numbers. The error bars on the scattered dots are calculated from the mean and SD from sampling 4 genes in the 70k region. n=7.

3.3.5 Isolation and characterization of a pure synthetic methylotrophic strain

After several attempts, we isolated SM1 and BB1 single colonies by colony PCR verification of unique mutations such as *pgi* 12-bp deletion in SM1. The final SM1 strain retains its ability to grow on MeOH even after culturing in LB (Fig. 3-12A). Moreover, the strain can also grow without any nitrate or vitamin supplementation (Fig. 3-12B).

Illumina HiSeq sequencing of SM1 showed similar SNVs with increased frequency (close to 100%) compared to the last sequenced mixed culture, CFC526.30. The 70k and 130k multi-copy regions remained while another 240k duplicate appeared in SM1 (Fig. 3-10B). In contrast, the high coverage 7k region disappeared, which was later identified as a unique feature of BB1 strain (Fig. 3-10C).

To determine the genome structure, we sequenced SM1 and BB1 with Pacbio Sequel and Nanopore sequencing to seek longer reads. *De novo* assembly of these long-read sequencing results were instrumental for determining the genome structure and polishing the genome sequence. Several previously identified low-frequency SNVs from Hiseq sequencing were actually IS insertions (Table S2). These long sequencing reads (Fig. 3-13A) also revealed that the IS5-flanked 70k region consisted of tandem repeats (Fig. 3-10B). In particular, several ultra-long mapping reads (100~130 kb) from Nanopore sequencing that spanned three tandem repeats appeared (Fig. 3-13A). Comparing SM1 and the wild-type *E. coli* BW25113, we observed several genomic structural variations due to insertion sequences and CNVs (Fig. 3-13B).

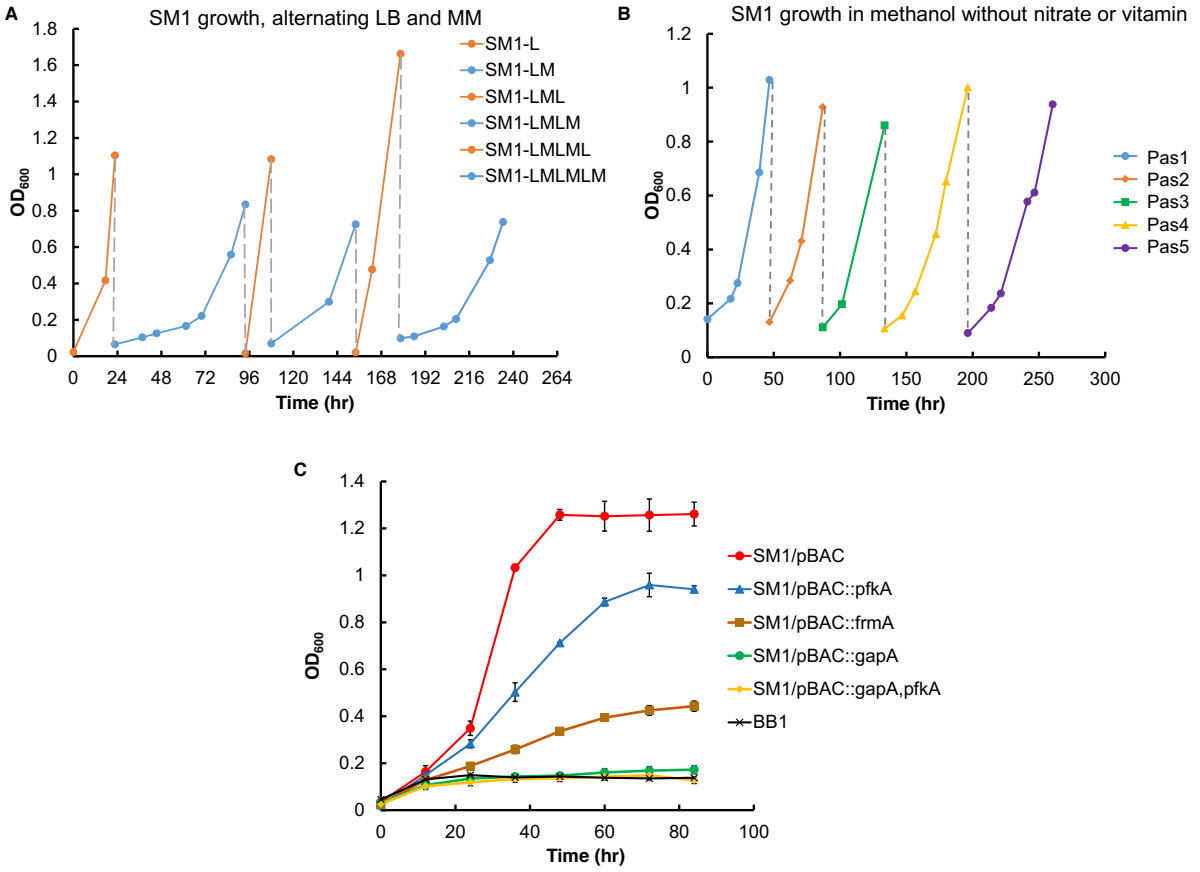


Fig. 3-12. Growth phenotype and characterization of methylotrophic *E. coli.*

(A) SM1 metabolic flexibility in switching LB & Methanol media. The "L" (orange line) and "M" (blue line) represent LB medium and methanol MOPS media respectively. Strains are passed at an inoculation volume of 100ul with initial OD_{600} of 0.05. **(B)** SM1 growing in 400 mM methanol without nitrate or vitamin. SM1 can be stably passed in a minimal media with methanol as the sole carbon source, without any supply of nitrate or vitamin. Strains are passed when it reached OD_{600}=1 with an initial OD_{600}=0.1. **(C)** Evaluation of rationally designed gene modifications. The WT genes were restored into SM1 and their growth phenotype were evaluated. n=3. The error bars represent the standard deviations.

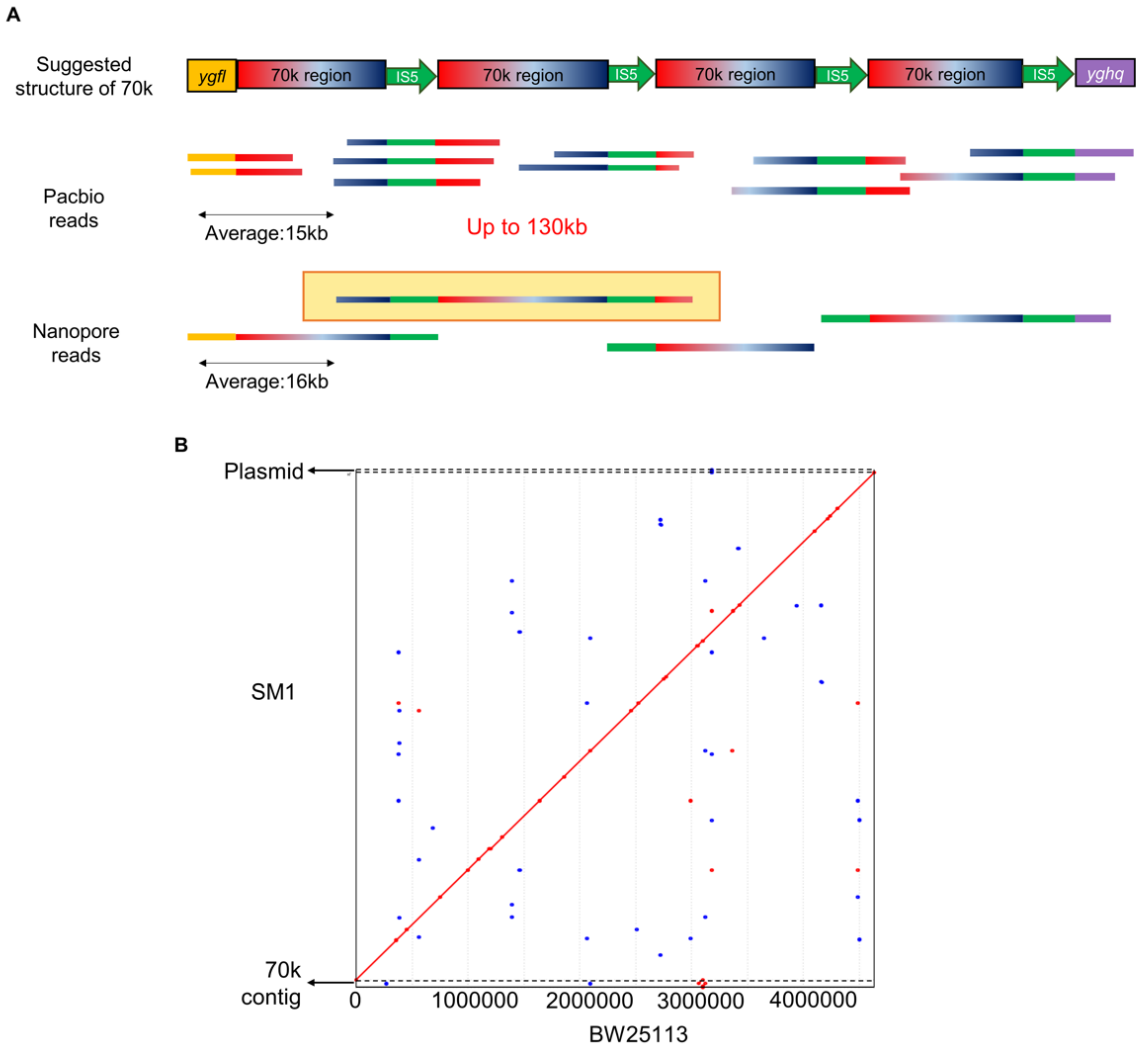


Fig. 3-13. Long-read sequencing of methylotrophic *E. coli* SM1

(A) Pacbio and Nanopore sequencing established the genomic structure of the 70k repeated region. The longest read from Pacbio Sequel that mapped between the 70k tandem repeat is 34k, while the longest read from Nanopore that mapped between tandem repeats is 110 kb-long. The latter proves the presence of an at-minimum triplicated 70k tandem-repeat. **(B)** Mummer plot comparing SM1 and BW25113. The SM1 genome is acquired by *de novo* assembly from Pacbio sequel data. The main contig highly correlates with the WT genome, suggesting that data is reliable. Moreover, there are two more contigs, including the plasmid and the 70k region. Note that the 70k aligns well on the BW25113 with a breaking point, due to the lack of the synthetic promoter that is integrated in SM1. The plasmid mapped to the WT *rpiA* position as excepted.

3.3.6 Beneficial IS-mediated copy number variations

During evolution, the copy number of the 70k tandem repeat increased, leading to 4 copies in the isolated SM1 strain (Fig. 3-11A). The fine-tuning of CNV implies that the 70k-tandem repeats may play an important role in synthetic methylotrophy as they host one of the artificially integrated operon, P_L*lac*O_1:: *medh-tkt-tal-hps-phi*, while also containing glycolysis and gluconeogenesis genes such as *fbaA*, *pgk* and *yggF* (a fructose-bisphosphatase isozyme) (Fig. 3-10B). The upregulation of the RuMP pathway enzymes may have enhanced the efficiency of methanol assimilation. The increase in *yggF* copy number may have further decreased Pfk flux, which is consistent with our EMRA prediction. We confirmed the copy number of the 70k tandem repeat in SM1 by digital PCR, Illumina sequencing, and long-read sequencing coverage data. Noticeably, we found that the copy number of the 70k reduced to 3 when the strain was grown in LB. On the other hand, the copy number of the 240k and 130k duplicated regions did not vary along the evolution path.

To further investigate the correlation between methanol growth and the 70k CNV, and the dynamics of CNV in SM1 strain, we picked a single colony of SM1 strain and passed it in LB 4 times, and then to methanol minimal medium (MM) to generate CNVs. We then passed several isolated single colonies from the last MM culture through additional serial passages in LB, while tracking their 70k CNVs and their methanol growth abilities after LB exposure. We found that the copy number of the 70k region decreased as the strain was more exposed to LB (Fig. 3-11C). After passing the strain back to MM, the strain increased its copy number (Fig. 3-11D) back to more than 4.5. Also, there is a correlation between methanol growth rate and the 70k copy numbers (Fig. 3-11E). On the other hand, the 7k multi-copy region unique to the BB1 strain featured a remarkable 85-fold coverage (Fig. 3-10C). This region hosts the *ddp* operon that is a putative dipeptide

transport and utilization, suggesting that BB1 may be co-evolved for the purpose of utilizing dipeptides derived from the debris of SM1 after cell death. After entering the stationary phase in the MM medium or passing through LB, this strain rapidly took over and dominated the culture. This apparent symbiosis explained the difficulty experienced in isolating SM1 from the evolved mixed culture when we isolate strains directly from LB plates.

3.3.7 Balancing formaldehyde flux

Balancing formaldehyde flux appears to be the most important task to avoid DPC. This task is particularly challenging when the cell needs to replenish Ru5P to react with formaldehyde in methanol-only media. SM1 accomplished this task in the log phase but still exhibited some DPC when it entered the stationary phase. To probe the gene expression changes between the stationary phase and the log phase, we performed RNA-seq analysis of SM1 in the MM medium and compared the mRNA transcript levels at OD_{600} 1.1 to OD_{600} 0.7. Indeed, the mRNA profile in the RuMP pathway was significantly altered in the stationary phase (Fig. 3-14A). The transcript levels of most RuMP genes responsible for the regeneration of Ru5P that reacts with formaldehyde were dramatically decreased, while the formaldehyde-forming gene (*medh*) was down-regulated less. Consequently, the flux imbalance caused the accumulation of formaldehyde. We also used qRT-PCR methods to verify that the expression changes were consistent with the RNA-seq results (Fig. 3-14A). It appeared that the fine balance between formaldehyde-forming and formaldehyde consuming flux was crucial when the cells enter the stationary phase (Fig. 3-14B). The Entner-Doudoroff (ED) pathway provides another route for entering the RuMP pathway to regenerate Ru5P. Interestingly, the ED pathway genes were also down regulated more in the stationary phase than the formaldehyde generation gene, medh, contributing to the DPC formation in the stationary phase.

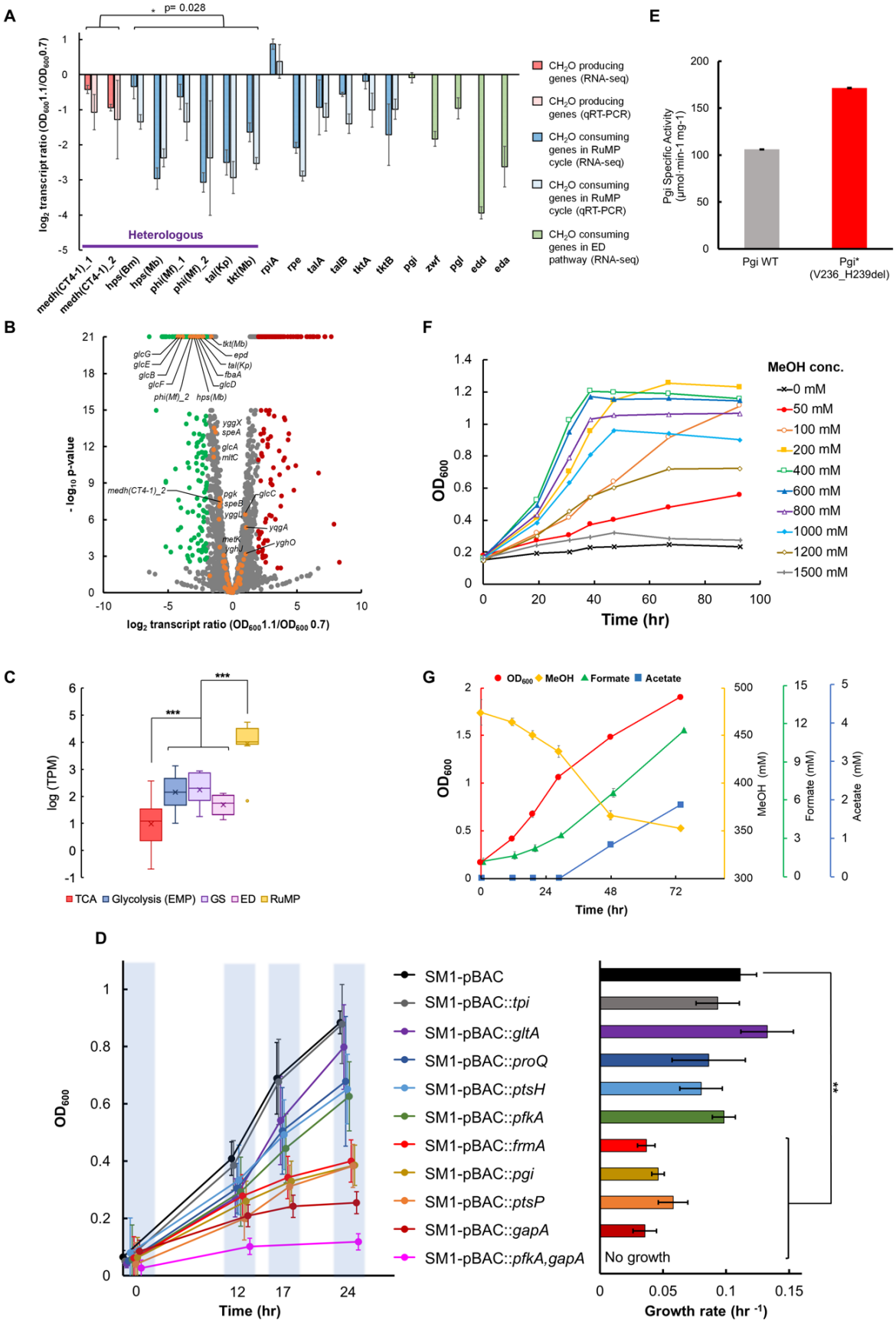

A
p= 0.028
log2 transcript ratio (OD600 1.1/OD600 0.7)
CH2O producing genes (RNA-seq)
CH2O producing genes (qRT-PCR)
CH2O consuming genes in RuMP cycle (RNA-seq)
CH2O consuming genes in RuMP cycle (qRT-PCR)
CH2O consuming genes in ED pathway (RNA-seq)
Heterologous
medh(CT4-1)_1
medh(CT4-1)_2
hps(Bm)
hps(Mb)
phi(Mf)_1
phi(Mf)_2
tal(Kp)
tkt(Mb)
rpiA
rpe
talA
talB
tktA
tktB
pgi
zwf
pgl
edd
eda
E
Pgi Specific Activity (μmol·min-1·mg-1)
Pgi WT
Pgi* (V236_H239del)
B
- log10 p-value
log2 transcript ratio (OD600 1.1/OD600 0.7)
tkt(Mb)
epd
tal(Kp)
fbaA
glcD
hps(Mb)
glcG
glcE
glcB
glcF
phi(Mf)_2
yggX
speA
glcA
mltC
medh(CT4-1)_2
pgk
speB
yggL
glcC
yqgA
metK
yghJ
yghO
F
OD600
Time (hr)
MeOH conc.
0 mM
50 mM
100 mM
200 mM
400 mM
600 mM
800 mM
1000 mM
1200 mM
1500 mM
C
log (TPM)
TCA
Glycolysis (EMP)
GS
ED
RuMP
G
OD600
MeOH
Formate
Acetate
Time (hr)
MeOH (mM)
Formate (mM)
Acetate (mM)
D
OD600
Time (hr)
SM1-pBAC
SM1-pBAC::tpi
SM1-pBAC::gltA
SM1-pBAC::proQ
SM1-pBAC::ptsH
SM1-pBAC::pfkA
SM1-pBAC::frmA
SM1-pBAC::pgi
SM1-pBAC::ptsP
SM1-pBAC::gapA
SM1-pBAC::pfkA,gapA
No growth
Growth rate (hr -1)

Fig. 3-14. Characterization of SM1 strain.

(A) Core methanol production/ consumption gene transcript ratios (OD_{600} 1.1/0.7) in 400 mM methanol MOPS medium measuring by RNA-seq and qRT-PCR. The RNA-seq results of the ED pathway genes are also shown in green. **(B)** The volcano plot of RNA-seq (log_2 transcript ratio of OD_{600} 1.1/ 0.7, 400 mM methanol). Red: $^{**}p<0.01$, log_2 ratio>2; Green: $^{**}p<0.01$, log_2 ratio<-2; Grey: $p>0.01$, |log2 ratio|< 2; Orange: Genes involved in multi-copy 70k region with $^{***}p<0.001$. **(C)** Expression Profile of SM1 sorted by metabolic pathways. TPM (Transcripts per million) was deduced by RNA-seq of SM1 during log phase growth (OD_{600} =0.7). **(D)** Growth phenotype of SM1 strain re-expressing *tpi*, *gltA*, *proQ*, *ptsH pfkA*, *frmA*, *ptsP*, *pgi*, *gapA* in 400 mM MeOH. n=3. **(E)** Specific Activity of Pgi and Pgi mutant (V236_H239del). **(F)** Growth of SM1 strain in various methanol concentrations. **(G)** Fermentation profile of SM1. Lines represent growth (red), methanol consumption (brown), formate (green) and acetate (blue). All error bars are defined as standard deviations, n=3. * p< 0.05, ** p<0.01, ***p<0.001.

3.3.8 Beneficial mutations for synthetic methylotrophy

An important reason for the success in evolving SM1 was the rational design guided by EMRA that involved the deletion of *pfkA* and *gapA* and expression of *gapC*. These genome changes were designed to direct more flux to replenish Ru5P to assimilate formaldehyde. To verify the importance of these genome edits along with other mutations introduced during laboratory evolution, we reversed these changes in SM1 and tested their phenotypes. We cloned *frmA, pfkA, gapA, pgi, gltA*, *ptsH*, *ptsP* and *proQ* on a bacterial artificial chromosome (pBAC) under their native promoters. Results showed that reinstalling the wild-type versions of these genes all caused a negative effect on methanol growth (Fig. 3-14D). Specifically, *frmA*, *gapA, pgi and ptsP* showed the most significant effects, indicating that these mutations were particularly beneficial to SM1 growth. Moreover, when both *pfkA* and *gapA* were simultaneously reintroduced, the strain almost stopped growing, requiring a 7-day recovery to grow back to OD_{600} 1 (Fig. 3-14D and 13C). Therefore, the rationally designed *pfkA* and *gapA* genome edits effectively created a path for genomic evolution towards efficient growth in methanol.

As mentioned previously, we found an IS2 insertion in the promoter region of *gltA*. Re-expressing a copy of *gltA* on pBAC slightly reduced the growth rate, suggesting that the IS2 insertion was instrumental for optimal SM1 growth. Moreover, RNA-seq data indicated that TCA cycle genes transcripts per million (TPM) were much lower than other major metabolic pathways such as glycolysis and RuMP cycle in SM1 (Fig. 3-14C).

We also expressed and His-tagged purified the Pgi variant coded by the 12bp-deleted *pgi* gene in SM1. Interestingly, this Pgi variant resulted in a higher specific activity, presumably increasing the flux through Zwf to produce NADPH for growth (Fig. 3-14E). NADPH in the wild-type *E. coli* mainly comes from three sources: Icd in the TCA cycle, Gnd, and Zwf in the oxidative

pentose phosphate pathway. Since Gnd is deleted and the TCA cycle activity is low as deduced from the RNA-seq data (Fig. 3-14C), Zwf may have become the major NADPH source for growth.

3.3.9 Growth characterization of SM1 strain

Finally, we determined the methanol growth characteristics of SM1. This strain could grow in a wide concentration range of methanol from 50 mM to 1.2 M as the sole carbon source, free of nitrate (Fig. 3-14F). Optimal growth was observed around 400 mM methanol, as the strain grew from OD600 0.1 to 1.0 in 30 hours with a doubling time of 8 hours and consumed around 120 mM of MeOH to reach a final OD600 of 1.9. Formate and acetate were the major products (Fig. 3-14G).

3.4 Discussion

This work demonstrated the tropism change of a microorganism. With only three missing genes, the metabolic rewiring turns out to be unexpectedly complicated. We started from a methanol auxotrophy strategy that established the working pathway for methanol assimilation, but the regeneration of the co-substrate, Ru5P, for formaldehyde conversion was supplied from an external carbon source, xylose. This methanol auxotrophy strain was evolved to grow every well with one sixth of its carbon derived from methanol. The remaining task was to wean off xylose and regenerate Ru5P by diverting part of the glycolytic flux to the RuMP cycle. Unexpectedly, this task has transpired to be the most challenging, and yet most revealing one in converting *E. coli* to a synthetic methylotroph. In the early stage of evolution for methanol auxotrophic growth (CFC381.20), the formaldehyde detoxification gene, *frmA*, was inactivated by a frameshift mutation to direct the formaldehyde flux to the productive RuMP pathways. This gene was also artificially deleted by other groups in efforts to turn *E. coli* into a synthetic methylotroph (Bennett et al., 2018a; Gonzalez et al., 2018; Woolston et al., 2018), or mutated during evolution in our previous work and another group (Chen et al., 2018; Meyer et al., 2018), suggesting that formaldehyde needs to be conserved for productive biosynthetic pathways for methylotrophic growth.

Since methylotrophic growth on methanol requires a proper balance between RuMP cycle, glycolysis, pentose phosphate pathway, and the ED pathway, imbalance among these pathways causes the shortage of either Ru5P for formaldehyde assimilation, pyruvate for building blocks, or NADPH for biosynthesis. Shortage of Ru5P will result in formaldehyde-induced DPC and then cell death. Shortage of pyruvate or NADPH will hamper growth. Such a fine balance could not be designed *a priori*, and will be very difficult, if not impossible, to accomplish by evolution within

the laboratory time scale. The EMRA provided a guideline to identify key enzymes that need to be up or down regulated. According to the EMRA result, Pfk and Gapdh need to be down regulated in order to avoid severe imbalance among different pathways. Pfk catalyzes a major metabolic step involved in ATP consumption and tunes glycolysis and gluconeogenesis, while Gapdh is a key metabolic node involved in NADH generation and is a junction among glycolysis, RuMP cycle and the pentose phosphate pathway. After implementing the EMRA-suggested genomic changes, the cells were able to gain growth advantage in methanol and evolve towards methylotrophic growth. Without these genomic edits at the initial stage, the cells seem to be trapped by DPC and not be able to evolve at the time scale of interest.

The DPC problem was visualized by TEM, clearly demonstrating the difficulty in turning *E. coli* to grown in methanol. The DPC phenomenon was most significant in the stationary phase. Even when the cells were able to grow in methanol, DPC kills the cells in the stationary phase. It was common to observe weak growth in methanol to a low cell density in a long time frame, but such colonies or cultures were not able to evolve further, possibly because of the DPC issue. Avoiding formaldehyde accumulation appears to be the only solution, as DPC is a problem on the global proteome level. Typical microbes detoxify formaldehyde by oxidizing it to CO_2, but this strategy wastes the biosynthetic carbon source. For methylotrophy, the organism needs to achieve a fine balance among formaldehyde generation and formaldehyde consumption flux. Native methylotrophs presumably have achieved this fine regulation through natural evolution. We solved this problem by introducing two genome edits followed by laboratory evolution.

Throughout evolution, we found divergence that created sub-populations identified by genome sequencing. Eventually, we identified two main populations, the methylotrophic SM1 and the non-methylotrophic BB1 strain (Table 3-2). SM1 grows on methanol and produces acetate in the late growth phase, which may feed the BB1 strain for growth. In the stationary phase, SM1 suffers from DPC and dies. Meanwhile, BB1 may also have enjoyed the acetate along with nutrients from dead SM1, such as amino acids or proteins, implicated from the increased copy numbers of the *ddp* operon. This symbiotic co-culture hypothesis explains the difficulty of isolating a pure synthetic methylotrophic strain, even after several rounds of plate streaking. Whether BB1 is helpful in the evolution of SM1 or just a fortuitous by-stander remains unproven. This result suggests that divergent evolution is likely during laboratory evolution, and that sequencing the evolving cultures and isolating single colonies is important after evolution.

One of the most intriguing features that laboratory evolution used to solve the DPC problem is CNV. In the SM1 strain, the copy number of the 70K repeated region increased as the evolution proceeded. The isolated SM1 strain showed that the copy of the 70K region decreased when the strain was cultured in LB, but increased when changing from LB to methanol minimal medium (Fig. 3-11D). This phenomenon was observed in all colonies we tested. It appears that SM1 uses CNV dynamically to adapt to new environment. The co-evolved non-methylotrophic BB1 strain does not contain the high copy 70k region, but acquired an extremely high (85) copy of the 7K region. It implies that CNV plays an important role in laboratory evolution for adapting to a challenging environment. Although CNV are common in eukaryotes studies like cancer cells (Shlien and Malkin, 2009) and fruit flies (Zhou et al., 2011), CNV in monoploid bacteria has not been well characterized with some notable exceptions (De Meur et al., 2018; Tyo et al., 2009). How *E. coli* dynamically tuned CNV along with environmental changes remains to be explored.

Moreover, the copy number for the 70k region in the initial CFC526.0 strain was already 2, indicating that CNV may have occurred since methanol auxotrophy evolution. This may also explain why our stepwise evolution strategy is effective. Without this auxotroph strategy to prepare the genomic background, we probably would not have obtained this critical 70k region available for further copy number increase and optimization.

The synthetic methylotroph SM1 strain also benefitted from other mutations, such as the 12-bp deletion in *pgi*, which increased its activity (Fig. 3-14E). It's likely that Pgi activity diverts part of the flux to the oxidative pentose phosphate pathway and generate NADPH for growth. The carbon flux then feeds into the ED pathway to generate pyruvate for growth and G3P for the RuMP pathway (Fig. 3-1A).

After extensive evolution, the final synthetic methylotroph strain exhibits an 8.5 h doubling time (t_D) and methanol tolerance (up to 1.2 M) comparable to native methylotrophs such as *Methylobacterium extorquen*s AM1, (t_D = 4 h (Nayak and Marx, 2014), *Methylobacterium extorquen*s TK0001 (t_D = 4~6 h (Belkhelfa et al., 2019)) and *Pichia pastoris* (t_D = 8.2 h (Moser et al., 2017)) in methanol only conditions.

Future engineering efforts can be implemented for balancing flux at the stationary phase to avoid DPCs. Such a synthetic methylotrophic strain may be engineered to produce a variety of chemicals from methanol, expanding the substrate range of *E. coli* in industrial microbiology. Moreover, the methanol condensation cycle (MCC) [Bogorad et al., 2014] can be implemented to avoid carbon loss when producing acetyl-CoA derived compounds such as ethanol and higher alcohols. Additionally, successful expression of methane monoxygenases (MMOs) may further extend the substrate range to methane as the sole carbon source. Because of its genetic tractability,

the synthetic methylotrophic *E. coli* presented here opens a wide possibility for further scientific studies as well as engineering applications for the production of chemicals from C1 compounds.

3.5 Acknowledgements

We thank the following colleagues from Academia Sinica, Taiwan: Dr. Meiyeh Jade Lu for fruitful discussion in genomics data analysis; Chia-che Lee for NGS data assistance; Dr. Shu-Jen Chou for digital PCR experiments; Dr. Sue-Ping Lee and Wen-Li Peng for TEM assistance; Dr. Shu-yu Lin and Dr. Yi-yun Chen for proteomics discussion; and Dr. Tzi-Yuan Wang for Nanopore sequencing. We appreciate previous UCLA Liao lab members: Dr. Tung-yun Wu for bioprospecting enzymes; Dr Powei Chen for EMRA support; Dr. Chang-Ting Chen and Seunghwan Lee for methylotrophy work; and Aaron Baugh for Crispr optimization efforts. We also thank the following core facilities in Academia Sinica: NGS High Throughput Genomics Core Facility at Biodiversity Research Center, the Imaging Core facility at the Institute of Molecular Biology Center, the Genomic Technology Core at the Institute of Plant and Microbial Biology, and the Academia Sinica Common Mass Spectrometry Facilities for Proteomics and Protein Modification Analysis at the Institute of Biological Chemistry. The NGS core facility and mass spectrometry Facility are supported by Academia Sinica Core Facility and Innovative Instrument Project (Grant ID AS-CFII-108-114 and AS-CFII-108-107, respectively).

3.6 Declaration of interests

Academia Sinica and UCLA are currently in the process of applying for patents based on this work.

3.7 Materials & Methods

3.7.1 Resource availability

3.7.1.1 Lead Contact

Additional Information and requests including materials and resources is available by contacting the Lead Contact, James C. Liao (liaoj@gate.sinica.edu.tw)

3.7.1.2 Materials Availability

All materials are available upon request. This study did not incorporate new unique reagents. Refer to 3.7.1.4 Key resources Table

3.7.1.3 Data and Code Availability

A copy of the EMRA code is available upon request. The Whole Genome Sequencing data and RNA-seq Data were deposited in the National Center for Biotechnology Information (NCBI) databases. They can be accessed via Bioproject: PRJNA643164 and Gene Expression Omnibus (GEO): GSE153477.

3.7.1.4 Key Resource Table (Table 3-3)

See next page.

3.7.2 Experimental and subject details

3.7.2.1 *Escherichia coli*.

We used *E. coli* K-12 BW25113 as the experimental model in this study. See Key Resource Table (Table 3-3) and Table 3-4.

Table 3-3. Key Resource Table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli strain BW25113	Coli Genetic Stock Center (CGSC)	CGSC #: 7636
E. coli strain BW25113 Δ*rpiA*	CGSC	CGSC #: 11414
E. coli strain Dh5alpha	CGSC	CGSC #: 14231
Chemicals, Peptides, and Recombinant Proteins		
Media		
Luria-Bertani (LB)	BD	BD 244610
Bacto Agar	BD	BD 214010
Hi Def azure media (HDA)	Teknova	Cat. #3H5000
MOPS EZ buffer	Teknova	Cat. #M2101
Terrific Broth (TB)	Sigma-Aldrich	Cat. #T0918
Antibiotics		
Carbenicillin disodium	AKsci	Cat. # C435
Kanamycin Sulfate	Amresco	Cat. # 97062-956
Chloramphenicol	Sigma-Aldrich	Cat. # C0378
Spectinomycin	Sigma-Aldrich	Cat. # S4014
Chemicals		
D-(-)-Ribose	Sigma-Aldrich	Cat. #R7500
D-Xylose	Sigma-Aldrich	Cat. #X1500
MeOH	Sigma-Aldrich	Cat. #646377
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Amresco	Cat. #0487-10G
Nicotinamide	Sigma-Aldrich	Cat. #N0636
Thiamine hydrochloride	Sigma-Aldrich	Cat. #T1270
Riboflavin	Sigma-Aldrich	Cat. #R9504
Calcium pantothenate	Sigma-Aldrich	Cat. #C8731
Biotin	Sigma-Aldrich	Cat. #B4639
Folic acid	Sigma-Aldrich	Cat. #8758
Vitamin B12	Sigma-Aldrich	Cat. #V6629
Sodium nitrate	Alfa Aesar	Cat. #AF-14493
Methanol-^{13}C,	Sigma-Aldrich	Cat. #277177
Sodium acetate-$^{13}C_2$	Sigma-Aldrich	Cat. #282014
Sodium formate-^{13}C	Sigma-Aldrich	Cat. #279412
Lysozyme	Bioshop	Cat. #LYS702
DNAzol reagent	ThermoFisher	Cat. #10503027
Ethanol	Fisher	Cat. #AC615090010

Urea	Sigma-Aldrich	Cat. #15604
Sodium dodecyl sulfate (SDS)	Amresco	Cat. #AM-0227
Sodium hydroxide	Amresco	Cat. #AM-E584
Tris	VWR	Cat. #PT-0826
Hydrochloric Acid	J.T.Baker	Cat #9673-00
HindIII	New England Biolabs	Cat # R0104
L-(+)-Arabinose	Sigma-Aldrich	Cat. #A3256
Critical Commercial Assays		
Qiagen Miniprep kit	Qiagen	Cat. #27106
Qiagen Puregene kit	Qiagen	Cat. #69506
NEBuilder kit	New England BioLabs	NEB #C2987
Pierce™ silver staining kit	Thermo Scientific	Cat. #24612
Pierce™ Coomassie Plus (Bradford Assay)	Thermo Scientific	Cat. #23236
KAPA Hyper Prep Kit	Roche	Cat. 07962363001
Ligation Sequencing Kit	*Oxford Nanopore technologies*	Cat. SQK-LSK109
RNeasy Plus mini kit	Qiagen	Cat. #74136
QuantiNova reverse transcription kit	Qiagen	Cat. #205413
QuantiNova SYBR green RT-PCR kit	Qiagen	Cat. #208152
Deposited Data		
Whole genome sequencing data	This study	Bioproject: PRJNA643164
RNA-seq data	This study	GEO: GSE153477
Experimental Models: Organisms/Strains		
CFC381	This study	See Table S3
CFC381.20	This study	See Table S3
CFC526.0	This study	See Table S3
CFC526.53	This study	See Table S3
CFC680.1	This study	See Table S3
CFC680.32	This study	See Table S3
CFC688.1	This study	See Table S3
CFC688.32	This study	See Table S3
SM1	This study	See Table S3
BB1	This study	See Table S3
Oligonucleotides		
See Table 3-5 for primers used	Purigo	N/A
Recombinant DNA		
See Table 3-6 for all plasmids used and constructed	See Table 3-6	See Table 3-6
Software and Algorithms		

Quantasoft Analysis Pro	Bio-Rad	https://www.bio-rad.com/en-id/product/qx200-droplet-digital-pcr-system
Microsoft Excel	Microsoft	www.microsoft.com
R software	R Core Team	www.r-project.org
CLC Genomics Workbench 20	Qiagen	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench/
Geneious 2020	Geneious	www.geneious.com/
Matlab 2019b	Matlab	mathworks.com
MUMmer4	MUMmer4	www.mummer4.github.io/

Table 3-4. Strain list.

Name	Genotype	Description	Reference
Dh5alpha	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ$^-$	Wild-type used for plasmid construction.	
BW25113	F$^-$ LAM$^-$ *rrnB3* Δ*lacZ4787 hsd*R514 Δ*(araBAD)*567 Δ*(rhaBAD)*568 *rph-1*	Wild-type used for establishing synthetic methanol growth strain.	
CFC381	BW25113 Δ*rpiA::FRT* Δ*rpiB* Δ*nupG*::$P_L$lacO_1::*medh/tkt(Mb)/tal(Kp)/hps(Mb)/phi* SS3(intergenic site)::$P_L$lacO_1::*medh/hps(Bm)/phi*	Initial strain to evolve for methanol auxotrophy	This study
CFC381.20	Refer to Table 3.2	CFC381 evolved to grow in methanol and xylose after 20 passages. A methanol auxotroph.	This study
CFC526.0	CFC381.20 Δ*pfkA* Δ*gapA::gapC* with pFC139	Initial strain to evolve for synthetic methylotrophy	This study
CFC526.53	Undetermined (Mixed culture)	Evolved from CFC526.0 to grow in methanol plus nitrate, with slower nutrient reduction.	This study
CFC680.1	Undetermined (Mixed culture)	Evolved from CFC526.0 to grow in methanol plus nitrate with faster nutrient reduction.	This study
CFC680.32	Undetermined (Mixed culture)	CFC680.1 evolved to grow in methanol after 32 passages.	This study
CFC688.1	Undetermined (Mixed culture)	CFC680.1 grown in methanol without nitrate	This study
CFC688.32	Undetermined (Mixed culture)	CFC688.1 to grow in methanol without nitrate after 32 passages.	This study
SM1	Refer to Table 3-2, with pFC139C	Isolated single colony from CFC526.53 that can grow in methanol without nitrate or vitamin	This study
BB1	Refer to Table 3-2, with pFC139B	Isolated single colony from CFC526.53 that cannot grow in methanol	This study

Table 3-5. Primer List

Description	Sequence
rpiA knockout validation forward primer	5'- CGCCTTCTACCAGCAGAAAC -3'
rpiA knockout validation reverse primer	5'- CCCAGACCGTTGTATGCTTT -3'
rpiB knockout validation forward primer	5'- GGAAGCGCTGAATCAAACTC -3'
rpiB knockout validation reverse primer	5'- GCTCTTCATCCTCCAGTTGC -3'
nupG knock-in validation forward primer	5'- ATATGCCATTTGCCACACCA -3'
nupG knock-in validation reverse primer	5'- CTTATATTCGCGGTGACGTG -3'
SS3 site knock-in validation forward primer	5'- TGTTAATTAGCGGGCAATTGTACC -3'
SS3 site knock-in validation reverse primer	5'- GATACCTACAGCGCAGAAAAACAA -3'
gapA knockout/gapC knock-in validation forward primer	5'- TGCTTCGATATTATGGCGGGCTT -3'
gapA knockout/gapC knock-in validation reverse primer	5'- GCCAGATGTGCAGGTTTCTCTTT -3'
pfkA knockout validation forward primer	5'- ATCAATCTTATGGACGGCTGGTC -3'
pfkA knockout validation reverse primer	5'- TGCTGATCTGATCGAACGTACCG -3'
frmA frameshift validation forward primer	5'- TATTTTGCCAGCCGCCAAAG -3'
frmA frameshift validation reverse primer	5'- CGAAATGACTGCTACAGCCG -3'
pgi deletion validation forward primer	5'- GAAGTCAACGCGGTGCTG -3'
pgi deletion validation reverse primer	5'- CCCTGGTGGATCAGCTGG -3'
pFC139 plasmid variation validation forward primer	5'- ATCCTACTGCTTTTTTCAATTCATC -3'
pFC139 plasmid variation validation reverse primer	5'- CAAGGGTGAACACTATCCCA -3'

Table 3-6. Plasmid List

Name	Description	Reference
pCas9	Plasmid for Cripsr-Cas9 with lambda-red recombinase system	Yu Jiang, et al.
pFC98	pTarget for Crispr knockout of Δ*rpiB* *Spec*^{R}	This study
pFC99	pTarget for Crispr editing of ΔSS3::$P_L$lacO_1::*medh/hps(Mb)/phi* *Spec*^{R}	This study
pCT309	pTarget for Crispr editing of Δ*nupG*::$P_L$lacO_1::*medh/tkt(Mb)/tal(Kp)/hps(Mb)/phi* *Spec*^{R}	This study
pFC139	P_L*lac*O_1::RBS library::*rpiA* p15A *ori* *Cm*^{R}	This study
pFC139A	Selected from pFC139 library with a specific RBS 5'-GACTAAAAACATTCGGAGGCTTAAGCAGTCATCGT-3'	This study
pFC139B	Same as pFC139, but with IS2 sequence between p15A and *cat,* Triplicated UTR before *rpiA*	This study
pFC139C	Same as pFC139A, but with an additional IS2 between *cat* and *rpiA*	This study
pFC144	pTarget for Crispr knockout of Δ*pfkA* *Spec*^{R}	This study
pFC149	pTarget for Crispr editing of Δ*gapA*::*gapC* *Spec*^{R}	This study
pCY96	Same as pBAC (oriV *ori* *Amp*^{R}), empty plasmid	This study
pCY153	pBAC::Native BW25113 *pfkA* operon (4,096,980-4,098,472)	This study
pCY154	pBAC::Native BW25113 *frmA* operon (373,388-375,577)	This study
pCY156	pBAC::Native BW25113 *gapA* operon (1,856,478-1,858,053)	This study
pCY161	pBAC::*gapA* operon (pCY156), *pfkA* operon (pCY153)	This study

3.7.2.2 Media and growth conditions

All strains were grown in at 37°C 250 rpm in a New Brunswick Scientific Innova 44, unless specified otherwise. LB (Becton Dickinson) was used for cloning purposes and the priority media when rich media was required. Antibiotics were used when required, at a final concentration of 100 mg/L for carbenicillin, 30 mg/L for kanamycin, 50 mg/L for chloramphenicol, or 250 mg/L for spectinomycin. Hi Def azure media (HDA, Teknova) was used as a preliminary-stage medium for adaptive evolution with limited nutrients. MOPS EZ buffer (MOPS, Teknova) was modified and utilized as a minimal medium, which consisted of 40 mM MOPS, 50 mM NaCl, 9.5 mM NH_4Cl, 0.525 mM $MgCl_2$, 4 mM tricine, 1.32 mM K_2PO_4, 0.276 mM K_2SO_4, 0.01 mM $FeSO_4$, 0.5 μM $CaCl_2$, 40 nM H_3BO_3, 8.08 nM $MnCl_2$, 3.02 nM $CoCl_2$, 0.962 nM $CuSO_4$, 0.974 nM $ZnSO_4$, and 0.292 nM $(NH_4)_2MoO_4$. OD_{600} was monitored by a G30 spectrometer (Thermo Scientific).

3.7.2.3 Strain construction and adaptive evolution of the MeOH auxotrophy strain

Strains used in this study are summarized in Table S2. The initial auxotrophy strain, CFC381.0 was constructed from a ΔrpiA strain included in Keio collection (Baba et al., 2006), followed by kanamycin cassette removal with a pCP20 plasmid encoding the FLP recombinase (Cherepanov and Wackernagel, 1995). Subsequently, rpiB removal and insertion of two operons, PLlacO1::medh-hps(Mb)-phi in the SS3 site (between ompW and yciE) (Bassalo et al., 2016), and PLlacO1::medh-tkt(Mb)-tal(Kp)-hps(Mb)-phi in the nupG site, were performed by a modified Crispr/Cas9 system from Jiang et al. (Jiang et al., 2015). The operons were integrated in SS3 and nupG respectively. pCas9-transformed strains were grown overnight, and reinoculated with an initial OD=0.1 and grown for 4 hours in a 30°C shaker with LB and 100mM arabinose. The strains are then electroporated with the pTarget

plasmid and plated on a LB plate at 30°C overnight. Successful gene editing targets were confirmed with colony PCR. The pTarget plasmid was then removed by growing cell on LB with 0.1mM IPTG, while the pCas9 was almost ultimately removed by growing the variant in LB 37°C and extensive screening. Methanol auxotrophy was evolved similarly as in previous studies (Chen et al., 2018). All E. coli strains were grown in a 3ml PP tube with the cap sealed to prevent evaporation and transferred to the next passage when OD600 exceeded 1.2, or reached a stationary phase. Strains are typically inoculated at an initial OD600 at 0.05 to 0.2. Accordingly, the population size at the bottleneck between transfers are approximately 1.5~6x108 cells. Thus, effectively 3~4 generations elapsed per passage.

The unevolved *ΔrpiAB* strain CFC381 was first inoculated in Terrific Broth (TB, Sigma) with 20 mM ribose and 20 mM xylose. Strains grew to saturation in two days and were then passed to HDA and MOPS separately, both with 400 mM MeOH and 20 mM xylose induced by 1 mM IPTG (media named as HMX and MMX, respectively). From Passage 2 to Passage 10, cells were passed from HMX to HMX. Beginning from Passage 11, cells were passed into MMX up until Passage 21 (CFC381.20), where the strain was subjected to further gene modifications after incorporating results suggested from theoretical calculations (See EMRA in the Methods Details Section).

3.7.2.4 Strain construction and adaptive evolution of MeOH growth strain

After methanol auxotroph was achieved from adaptive evolution, further knockout of *pfkA* and replacement of *gapA* with *gapC* were implemented by Crispr/Cas9 considering EMRA results. The Crispr protocol is identical to the one in the previous section. A plasmid pFC139 consisting *rpiA* with an RBS library was transformed into the strain by electroporation. Adaptive evolution was carried out by mixing a ratio of HDA and MOPS

with 400 mM of MeOH. In addition, a vitamin mix was added regardless of the ratio of MOPS and HDA medium, where the following final concentration is reached: 40.94 μM nicotinamide, 14.82 μM thiamine hydrochloride, 13.29 μM riboflavin, 10.49 μM calcium pantothenate, 8.19 μM biotin, 4.53 μM folic acid, 0.07 μM vitamin B12. Moreover, 10 mM of $NaNO_3$ was incorporated and 1 mM of IPTG was used for induction at the point of inoculation. The evolution began with 100:0 of the HDA: MOPS media (After adding MeOH and all supplement vitamins, the actual ratio of HDA medium was diluted to 95.2%). At passage 2 (CFC526.2), HDA: MOPS was adjusted to 50%. From passage 3 (CFC526.3) to passage 16 (CFC526.16), HDA was reduced to 30% HDA ratio was further reduced to 20% and 10% from passage 17 (CFC526.17) to 19 (CFC526.19), and passage 19 (CFC526.19) to 20 (CFC526.20), respectively. Finally, HDA was completely eliminated where MOPS was used on Passage 21, which was renamed to CFC 680.1. CFC680.1 was then further evolved on solely MOPS and 400 mM MeOH for 31 passages until CFC680.31. CFC688.2 was simultaneously grown from CFC680.1 with MOPS and 400 mM MeOH without nitrate, and then evolved for 30 passages until CFC 688.32. Last but not least, at CFC526.20, a slower HDA approach was done. Specifically, 10% and 5% HDA supplement was provided until passage 21 (CFC526.21) and passage 22 (CFC526.22) respectively. HDA was completely omitted at passage 23 (CFC526.23). CFC526.23 was then evolved for 30 passages until CFC526.53.

A final single colony of strain SM1 was obtained by first streaking out CFC526.53 on a MOPS plus 400 mM MeOH agar plate. The single colony was then inoculated into MOPS plus 400 mM MeOH liquid culture again, followed by streaking out on a LB plate in anaerobic conditions. The SM1 was finally retrieved by growing single colonies in LB with

colony-PCR confirmation. The other single colony strain BB1 was simply isolated by growing CFC526.53 in LB liquid and LB plate.

The final SM1 strain is grown in the 1x MOPS EZ media (10X stock from Catalog No. M2101, Teknova) along with 400mM MeOH, the previously mentioned vitamin mix, 1mM IPTG and 50mg/L chloramphenicol. Note that the chloramphenicol was dissolved in pure methanol as well, and was added to the media by a 1000x stock.

3.7.3 Method Details

3.7.3.1 Plasmid Construction

All plasmids that were implemented in this study are summarized in the Key Resources Table. All of the plasmids were constructed by Gibson Assembly with the NEBuilder kit (New England Biolabs) while DNA fragments were amplified by KODone (Toyobo). *E. coli* DH5alpha was used as the cloning host.

3.7.3.2 Robustness of RuMP-EMP-TCA cycle by EMRA.

EMRA is a calculation method developed to determine the likelihood of perturbations in enzyme expression and kinetics that causes instability of the steady state (Lee et al., 2014; Rivera et al., 2015). After pre-setting a reference steady state for the entire pathway, a total of 100 parameter sets were then generated and perturbed randomly from 0.1-fold to 10-fold for each enzyme. Results were reported as an indication of the robustness of the system, where $Y_{R,M}$ refers to the ratio of the 100 parameter sets that are robust at each point.

3.7.3.3 ^{13}C labelling experiment

Qualitative analysis of ^{13}C labelled acetate and formate was conducted by an Agilent Technologies 7890 gas chromatography along with a 5977B mass spectrometer. Samples were prepared by aliquoting the supernatant of the culture after centrifugation at 15000rpm for 3 minutes. 0.5µl of sample was injected into the GC. A DB-FFAP column (Agilent Technologies, 0.32 mm × 30 m × 0.25 µm) was utilized along with constant pressure of 7.0633 psi helium gas supply. The thermal cycle was carried out with an initial temperature of 40°C for 2 minutes, followed by a ramp rate of 10 °C/min to 60 °C and 100°C/min to 240 °C along with a final 2-minute hold.

3.7.3.4 Cell viability Test

The cell viability assay was done by using LIVE/DEAD BacLight Bacterial Viability Kit (Thermofisher Scientific, USA) following its protocol. The fluorescence of cells was then detected by a 2018 Attune NxT Flow Cytometer (Thermofisher Scientific, USA). The Blue laser (Excitation Wavelength 488nm) and BL1 filter (Emission filter 530±30nm) was selected for SYTO-9 detection, while the yellow laser (Excitation Wavelength 561nm) and YL2 filter (Emission filter 620±15nm) was used for propidium iodide detection.

3.7.3.5 Isolation of DPC complexes

The extraction protocol was modified from that reported by Qiu et al. (Qiu and Wang, 2009) and Barker et al. (Barker et al., 2005). 2 ml of *E. coli* (CFC526.41 and CFC680.24) was first pelleted by centrifugation at 5000 g for 5 minutes, and then resuspended in 100 µl of 10 mM MOPS buffer containing 2.0 mg/ml of lysozyme, with incubation of 30 minutes at 37°C. 500 µl of DNAzol reagent was then added and mixed for 5 minutes, followed by centrifugation at 12000 rpm for 10 minutes. The supernatant was then transferred into a new tube where 300 µl ice cold 100% ethanol was added for DNA precipitation. Samples were then stored at -80°C for at least 1

hour. Subsequently, after removal of supernatant by centrifugation at 12000 rpm for 5 minutes, the DNA pellet was re-dissolved in 190 µl of 8 mM NaOH. Subsequently, 10 µl of 1 M Tris-HCl, pH 7.4 was added while urea and SDS were added as well to a final concentration of 8 M & 2% w/v respectively for protein denaturing and disassociation of non-specific binding protein to DNA. The entire mixture was gently shaken at 37°C for 30 minutes. Protein was then salted out by adding equal volume of 5 M NaCl and subjected to gentle shaking at 37°C for 30 minutes. After centrifugation at 12000 rpm, 20 minutes, the supernatant was transferred to an Amicon Ultra-4 mL Centrifugal Filters with a 3 kDa cutoff (Millipore) and washed with 10 mM of Tris pH 7.4 thrice to a final dilution factor of 10000. When the volume was finally concentrated to 450 µl, 50 µl of 3 M potassium acetate and 1 ml of ice-cold 100% ethanol were added and stored once again at -80°C for 1 hour. After centrifugation at 12000 rpm, 20 minutes, 4°C, the DNA pellet was retained and washed with 1 ml 70% ethanol. The pellet was then dissolved with 10 mM Tris-HCl, typically 100 µl. The DNA was then quantified by 260 nm Nanodrop (Thermo).

3.7.3.6 Transmission electron microscopy

Purified DPC complex at a scale around 500 ng of DNA was mounted on an activated 300-mesh copper grid coated with carbon-stabilized formvar (Ted Pella) for 1 minute at room temperature. Following liquid removal by filter paper blotting, samples were stained with 2.5% uranyl acetate for 1 minute. After excess stain removal, the samples were air dried at room temperature. Transmission electron microscopy was performed with a Tecnai G2 Spirit Bio TWIN (FEI Co.), while images were recorded with a K3 Base IS CCD camera (Gatan) at a magnifying ratio of 2700x to 15000x.

3.7.3.7 Purification of DPC protein portion

Decrosslinking of the DNA was done by incubation at 70°C for 1 hour, followed by a DNAase I (NEB) and S1 nuclease (Thermo) treatment, 1 µl each. The small digested DNA fragment was then removed by using an Amicon Ultra-0.5 mL Centrifugal Filters with a 3 kDa cutoff and the final volume was reduced to 50 µl. Sample concentration was then estimated by 280 nm Nanodrop. SDS-PAGE was run for quick analysis with a 12% precast gel (Biorad), while staining was done with a Pierce silver staining kit (Thermo Scientific).

3.7.3.8 Protein Sample preparation for quantitative proteomics and LC-MS/MS analysis

Proteins were denatured by adding urea to a final concentration of 4 M, followed by reduction with 10 mM dithioerythritol at 37°C for 45 minutes, and cysteines alkylation with 25 mM iodoacetamide at room temperature in the dark for 1 hour. Protein samples were digested overnight at 37°C using LysC protease and trypsin at an enzyme-to-substrate ratio of 1:50 (w/w). After tryptic digestion, the peptides were desalted directly by C18 StageTip. Samples were performed on EASY-nLC^{TM} 1200 system connected to a Thermo Scientific Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany). Data analysis was performed using SEQUEST HT algorithm integrated in the Proteome Discoverer 2.4 (Thermo Finnigan). MS/MS scans were matched against an *E. coli* K12 database (UniProtKB/Swiss-Prot 2019_10 Release).

The data was analyzed by first normalizing the abundance of the internal standard DNase to the same value within different time points of the same sample. Then the normalized abundance of each sample was divided by their DNA concentration. Last, the heat map was plotted by listing out the top 100 hits of ranked by protein abundance and taking the common hits to visualize in a

log scale, with descending order based on average abundance of the final time point of individual samples.

3.7.3.9 DNA Next Generation Genome Sequencing

Genomic DNA was purified by Qiagen Puregene kit (Qiagen). All strains that were sequenced are summarized in Table S2. Samples that are collected throughout the adaptive evolution were sequenced by either Illumina Miseq or Illumina Hiseq Rapid (Illumina), with a 2 x 150 bp pair-end format. Samples that were in the middle of adaptive evolution were all ensured to have a coverage of at least 60 to differentiate sequencing error from SNPs. Data was then processed by Geneious 11 software (Geneious), by trimming with BBduk and then mapped to a reference by the software's native mapper. SNP variants were called by setting the criteria to a frequency of 25%.

The final stain was then sequenced by Pacbio Sequel (Pacific Biosciences) and Nanopore sequencing (Oxford Nanopore Technologies). For Pacbio Sequel, a 25kb SMRTbell library (Pacific Biosciences) was prepared and its quality was assessed by fragment analyzer (Agilent). The library was then run in a CLR 20-hr diffusion mode. The reads were assembled by HGAP4.

To perform DNA sequencing using Nanopore technology, an adapted protocol from Chang et al. was employed (Chang et. al, 2019). A KAPA Hyper Prep Kit (Kapa Biosystems) was used to repair fragment ends and then ligate barcode adaptors (Native Barcoding Expansion 1-12) to the genomic DNA (3~8 μg). The SQB and LB buffer in the Ligation Sequencing Kit (Oxford Nanopore Technologies) was then added to and mixed thoroughly with the barcoded genomic DNA library. Finally, the mixture was loaded to a Flow Cell (Oxford Nanopore Technologies, R9.4.1), and then sequenced by GridION devices for 24 hours or longer.

3.7.3.10 Digital PCR

CNVs were detected by droplet digital PCR (ddPCR) with the QX200TM ddPCR system (Bio-Rad). Genomic DNA was first extracted as done in the NGS experiment. 0.5 μg of DNA was then digested with HindIII for 1 hour. The PCR reaction was carried out by a "ddPCR Supermix for Probes" kit (Bio-rad) after loading 25pg of the digested DNA. Data was analyzed by QuantaSoft Analysis Pro Software.

3.7.3.11 Copy number variation Dynamics

SM1 was streaked out on a LB plate, where a single colony was picked and inoculated in LB. The LB culture was then passed to LB culture 3 more times, followed by a MM culture inoculation. The MM culture was then streaked out onto LB plates twice, where 7 colonies were then inoculated into LB once again. Data shown in this manuscript starts from this point. The strain was then passed into LB for 7 times. The 1st, 4th and 7th pass (annotated as "Pas1", "Pas4", "Pas7" culture were passed into MM culture to test methanol growth. The genomic DNA of the LB culture and the subsequent MM culture were extracted by the Qiagen Puregene kit, where the copy number were tested by digital PCR.

3.7.3.12 qRT-PCR analysis

E. coli total RNA was prepared using RNeasy mini kit (Qiagen) and reverse transcribed by QuantiNova reverse transcription kit (Qiagen). Detection of cDNA levels were performed using CFX Connect™ Real-Time PCR detection system (BioRad Laboratories). All samples were measured in triplicate in hard-shell 96-well PCR plate using QuantiNova SYBR green RT-PCR kit (Qiagen). The expression fold change was analyzed by ΔΔCt values normalized to *E. coli* 16S rRNA. The overexpressed heterologous genes were categorized in formaldehyde consuming and producing gene data set. The data sets were first tested by a Shapiro-Wilk test to test the data is

normally distributed, which was the case. A two-tailed F-test was then done to evaluate whether a t-test with equal or unequal variance should be used. Finally, the t-test was done to evaluate if the fold changes of formaldehyde consuming and producing genes are statistically different from each other.

3.7.3.13 RNA-seq analysis

E. coli total RNA was extracted by RNeasy mini kit (Qiagen). rRNA was prepped by Ribo-Zero (Bacteria) kit. Data was then processed by CLC genomics workbench 20. The TPM were calculated, and the following metabolic pathway includes the following gene when calculating the TPM distribution: TCA includes *aspA*, *fdrA*, *fdrB*, *fumA*, *fumB*, *fumC*, *gltA*, *icd*, *mdh*, *mqo*, *ppc*, *prpC*, *prpD*, *sdhA*, *sdhB*, *sdhC*, *sdhD*, *sucA*, *sucB*, *sucC*, *sucD*, *yahF* and *ybhJ;* EMP (Glycolysis) includes: *aceE*, *aceF*, *cra*, *eno*, *fbaA*, *fbaB*, *gpmA*, *gpmM, lpd*, *pfkB*, *pgj*, *pykA*, *pykF*, *tpiA* and *gapC;* ED includes*: pgi*, *zwf*, *pgl*, *edd* and *eda;* RuMP includes*: medh, hps, phi, tal, tkt, rpe, rpiA*. The TPM sets sorted by metabolic pathway were evaluated if they are statistically different from others by the methodology mentioned in the qRT-PCR section.

3.7.3.14 Reverting deleted genes and assessment of their phenotypic effects

Native operons of *pfkA*, *frmA*, *gapA* were cloned into BAC (bacteria artificial chromosome) with an AmpR selection marker and transformed back to SM1 strain. SM1 strains re-expressing *pfkA*, *frmA*, *gapA* or both *pfkA* and *gapA* were then re-inoculated into 400 mM methanol medium. Growth curves were recorded and compared with SM1 strain transformed with an empty BAC.

3.7.3.15 Methanol consumption and fermentation product analysis

Samples were prepared by aliquoting the supernatant of the culture after centrifugation at 15000rpm for 3 minutes and then filtered through a 0.22 filter (Milipore). Methanol concentration was determined by an Agilent Technologies 7890 gas chromatography with a flame-ionization-detector. Nitrogen gas with constant pressure of 19.082psi was flowed through a DB-624UI column (Agilent Technologies, 0.32 mm × 30 m × 0.25 μm) a thermal cycle consisting of the following stages: initial 45°C for 1 minute, ramp rate of 20 °C/min to 150 °C, and 45°C/min to 240 °C with a final 1-minute hold.

The fermentation products, namely acetate and formate, were measured by an Agilent 1290 UPLC using an Hi-plex H column (Agilent Technologies, 300x6.5 mm). A run was done with the mobile phase consisting 30 mM sulfuric acid with a flow rate of 0.6mL/min for 30 minutes.

3.11 References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Barker, S., Murray, D., Zheng, J., Li, L., and Weinfeld, M. (2005). A method for the isolation of covalent DNA–protein crosslinks suitable for proteomics analysis. Analytical Biochemistry *344*, 204–215.

Bassalo, M.C., Garst, A.D., Halweg-Edwards, A.L., Grau, W.C., Domaille, D.W., Mutalik, V.K., Arkin, A.P., and Gill, R.T. (2016). Rapid and Efficient One-Step Metabolic Pathway Integration in *E. coli*. ACS Synth. Biol. *5*, 561–568.

Belkhelfa, S., Roche, D., Dubois, I., Berger, A., Delmas, V.A., Cattolico, L., Perret, A., Labadie, K., Perdereau, A.C., Darii, E., et al. (2019). Continuous Culture Adaptation of *Methylobacterium extorquens* AM1 and TK 0001 to Very High Methanol Concentrations. Front Microbiol *10*, 1313.

Bennett, R.K., Gonzalez, J.E., Whitaker, W.B., Antoniewicz, M.R., and Papoutsakis, E.T. (2018a). Expression of heterologous non-oxidative pentose phosphate pathway from *Bacillus methanolicus* and phosphoglucose isomerase deletion improves methanol assimilation and metabolite production by a synthetic *Escherichia coli* methylotroph. Metabolic Engineering *45*, 75–85.

Bennett, R. K., Steinberg, L. M., Chen, W. and Papoutsakis, E. T. (2018b) Engineering the bioconversion of methane and methanol to fuels and chemicals in native and synthetic methylotrophs. Current Opinion in Biotechnology *50***,** 81–93.

Bertau, M., Offermanns, H., Plass, L., and Schmidt, F. (2014). Methanol: the basic chemical and energy feedstock of the future (Springer).

Bogorad, I.W., Chen, C.-T., Theisen, M.K., Wu, T.-Y., Schlenz, A.R., Lam, A.T., and Liao, J.C. (2014). Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. Proceedings of the National Academy of Sciences *111*, 15928–15933.

Brautaset, T., Jakobsen, Ø.M., Josefsen, K.D., Flickinger, M.C., and Ellingsen, T.E. (2007). *Bacillus methanolicus*: a candidate for industrial production of amino acids from methanol at 50 degrees C. Appl Microbiol Biotechnol *74*, 22–34.

Chen, C.-T., Chen, F.Y.-H., Bogorad, I.W., Wu, T.-Y., Zhang, R., Lee, A.S., and Liao, J.C. (2018). Synthetic methanol auxotrophy of *Escherichia coli* for methanol-dependent growth and production. Metabolic Engineering *49*, 257–266.

Chang, T.-S., Wang, T.-Y., Hsueh, T.-Y., Lee, Y.-W., Chuang, H.-M., Cai, W.-X., Wu, J.-Y., Chiang, C.-M., and Wu, Y.-W. (2019). A Genome-Centric Approach Reveals a Novel Glycosyltransferase from the GA A07 Strain of Bacillus thuringiensis Responsible for Catalyzing 15-O-Glycosylation of Ganoderic Acid A. Int J Mol Sci *20*.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9–14.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009). The Expanding World of Methylotrophic Metabolism. Annu. Rev. Microbiol. *63*, 477–499.

Conrado, R.J., and Gonzalez, R. (2014). Envisioning the bioconversion of methane to liquid fuels. Science, *84*, 621-623.

Cotton, C.A., Claassens, N.J., Benito-Vaquerizo, S., and Bar-Even, A. (2019). Renewable methanol and formate as microbial feedstocks. Current Opinion in Biotechnology *62*, 168–180.

De Meur, Q., Deutschbauer, A., Koch, M., Wattiez, R., and Leroy, B. (2018). Genetic Plasticity and Ethylmalonyl Coenzyme A Pathway during Acetate Assimilation in Rhodospirillum rubrum S1H under Photoheterotrophic Conditions. Applied and Environmental Microbiology *84*.

Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D., and Steiger, M.G. (2020). The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO2. Nature Biotechnology 1–16.

Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., et al. (2019). Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO2. Cell *179*, 1255–1263.e12.

Gonzalez, J.E., Bennett, R.K., Papoutsakis, E.T., and Antoniewicz, M.R. (2018). Methanol assimilation in *Escherichia coli* is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. Metabolic Engineering *45*, 67–74.

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015). Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. Applied and Environmental Microbiology *81*, 2506–2514.

Kim, S., Lindner, S.N., Aslan, S.X.U., Yishai, O., Wenk, S., Schann, K., and Bar-Even, A. (2020). Growth of *E. coli* on formate and methanol via the reductive glycine pathway. Nat Chem Biol *332*, 1–12.

Kotlarz, D., Garreau, H., and Buc, H. (1975). Regulation of the amount and of the activity of phosphofructokinases and pyruvate kinases in Escherichia coli. Biochim. Biophys. Acta *381*, 257–268.

Kuk, S.K., Singh, R.K., Nam, D.H., Singh, R., Lee, J.-K., and Park, C.B. (2017). Photoelectrochemical Reduction of Carbon Dioxide to Methanol through a Highly Efficient Enzyme Cascade. Angewandte Chemie *129*, 3885–3890.

Lee, Y., Lafontaine Rivera, J.G., and Liao, J.C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. Metabolic Engineering *25*, 63–71.

Lin, P.P., Jaeger, A.J., Wu, T.-Y., Xu, S.C., Lee, A.S., Gao, F., Chen, P.-W., and Liao, J.C. (2018). Construction and evolution of an *Escherichia coli* strain relying on nonoxidative glycolysis for sugar catabolism. Proc. Natl. Acad. Sci. U.S.a. *115*, 3538–3546.

Meyer, F., Keller, P., Hartl, J., Gröninger, O.G., Kiefer, P., and Vorholt, J.A. (2018). Methanol-essential growth of *Escherichia coli*. Nature Communications *9*, 1–10.

Moser, J.W., Prielhofer, R., Gerner, S.M., Graf, A.B., Wilson, I.B.H., Mattanovich, D., and Dragosits, M. (2017). Implications of evolutionary engineering for growth and recombinant protein production in methanol-based growth media in the yeast Pichia pastoris. Microbial Cell Factories *16*, 1–16.

Nayak, D.D., and Marx, C.J. (2014). Genetic and Phenotypic Comparison of Facultative Methylotrophy between Methylobacterium extorquens Strains PA1 and AM1. PLoS ONE *9*, e107887–10.

Noor, R. (2015). Mechanism to control the cell lysis and the cell survival strategy in stationary phase under heat stress. Springerplus *4*, 599.

Patel, S.K.S., Jeon, M.S., Gupta, R.K., Jeon, Y., Kalia, V.C., Kim, S.C., Cho, B.K., Kim, D.R., and Lee, J.-K. (2019). Hierarchical Macroporous Particles for Efficient Whole-Cell Immobilization: Application in Bioconversion of Greenhouse Gases to Methanol. ACS Appl Mater Interfaces *11*, 18968–18977.

Qiu, H., and Wang, Y. (2009a). Exploring DNA-Binding Proteins with In Vivo Chemical Cross-Linking and Mass Spectrometry. J. Proteome Res. *8*, 1983–1991.

Rivera, J.G.L., Lee, Y., and Liao, J.C. (2015). An entropy-like index of bifurcational robustness for metabolic systems. Integrative Biology *7*, 1–9.

Serwer, P. (1978). A technique for observing extended DNA in negatively stained specimens: observation of bacteriophage T7 capsid-DNA complexes. J. Ultrastruct. Res. *65*, 112–118.

Shlien, A., and Malkin, D. (2009). Copy number variations and cancer. Genome Med *1*, 62–69.

Smejkalová, H., Erb, T.J., and Fuchs, G. (2010). Methanol assimilation in Methylobacterium extorquens AM1: demonstration of all enzymes and their regulation. PLoS ONE *5*, e13001.

Stingele, J., and Jentsch, S. (2015). DNA-protein crosslink repair. Nat. Rev. Mol. Cell Biol. *16*, 455–460.

Tuyishime, P., Wang, Y., Fan, L., Zhang, Q., Li, Q., Zheng, P., Sun, J., and Ma, Y. (2018). Engineering Corynebacterium glutamicum for methanol-dependent growth and glutamate production. Metabolic Engineering *49*, 220–231.

Tyo, K.E.J., Ajikumar, P.K., and Stephanopoulos, G. (2009). Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nature Biotechnology *27*, 760–765.

Woolston, B.M., King, J.R., Reiter, M., Van Hove, B., and Stephanopoulos, G. (2018). Improving formaldehyde consumption drives methanol assimilation in engineered *E. coli*. Nature Communications *9*, 1–12.

Wu, Y., Wu, M., He, G., Zhang, X., Li, W., Gao, Y., Li, Z., Wang, Z., and Zhang, C. (2012). Glyceraldehyde-3-phosphate dehydrogenase: a universal internal control for Western blots in prokaryotic and eukaryotic cells. Analytical Biochemistry *423*, 15–22.

Wu, T.-Y., Liu, J.T.-J., and Damoiseaux, R. (2016). Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. Appl Microbiol Biotechnol *100*, 4969–4983.

Zhou, J., Lemos, B., Dopman, E.B., and Hartl, D.L. (2011). Copy-Number Variation: The Balance between Gene Dosage and Expression in Drosophila melanogaster. Genome Biology and Evolution *3*, 1014–1024.