nature portfolio

Peer Review File

Evolutionary engineering of methylotrophic E. coli enables fast growth on methanol

Corresponding Author: Professor James Liao

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
o General comments

- In this study, the author reported the development of a synthetic methylotrophic strain with a doubling time of 3.5 h in methanol. In the previous study by the same group, the E. coli strain SM1 was able to grow on methanol (10.1016/j.cell.2020.07.010) with a doubling time of 8 h. In this study, bacterial artificial chromosome (BAC) was used in the SM1 strain, which helped overcome the formaldehyde-induced toxicity using dynamic copy number variation (CNV), further enabling the adaptive evolution of the SM1 strain. This study is useful for the scientific community in three aspects. 1) The advantage of CNV in overcoming the bottlenecks of adaptive evolution 2) synthetic methylotroph which can grow better than native methylotrophs 3) provides meaningful insights regarding the metabolic rewiring of synthetic methylotrophy.

- Although the doubling time of SM6 has been improved considerably in this study, the maximum OD achieved was similar to that of the SM1 strain (10.1016/j.cell.2020.07.010). Moreover, the scale-up fermentation in methanol minimal medium has not been shown in this study, and MOPS media was used, which requires vitamin supplementation. Also, the role of CNV and BAC in metabolic rewiring is unanswered since previous studies from the same group have been conducted; therefore, further experimental data or discussion is required for better understanding.

- In comparison, a very recent publication by Michael et. al. (doi.org/10.1038/s41929-024-01137-0) used adaptive evolution to construct a synthetic methylotroph that can grow in the methanol-minimal medium in a fed-batch fermenter. Although the doubling time (4.3 h) of the strain constructed by Michael et. al. was lower than in this study, in fermentation a remarkable OD of 100.2 was achieved. In addition, the production of lactic acid, PHB, and etc. was demonstrated using a synthetic methylotrophy strain. Therefore, to consider this manuscript for publication in Nature Communications, at least the fermentative growth of synthetic methylotrophs needs to be shown. Also, I suggest the author include this reference (doi.org/10.1038/s41929-024-01137-0) for discussion and comparison.

- Moreover, the discussion part is poorly written without explanation and requires references. The mutations in genes, for example, OMPs, need to be characterized and explained properly for better understanding.

o Specific comments:

1. Figure 6b: TEM images were taken until OD 1.3. SM6 grew to OD 1.3 in less than 20 h, whereas SM1 took almost 30+ h. Why were no further time points compared? In the previous publication, it was mentioned that SM1 alleviated the DPC problem, but it seems higher with the passage of time in TEM images. Therefore, please explain Figure 6b in line 205.
2. Line 216: How does DPC interfere with growth? Add some explanation with references.
3. Line 219: The increase in ribosome abundance may be a key reason underlying the fast growth of SM6. Please explain with references.
4. Line 224: “These changes suggest that SM6 derives 2-carbon compounds not from pyruvate dehydrogenase (coded by aceEF) or pyruvate-formate lyase (coded by pflB) but from pyruvate oxidase (coded by poxB) and acetate utilization (coded by acs, glcB)”. Is the poxB route advantageous in terms of energy and carbon conservation?
5. Line 262: “Neither SM1 nor SM6 could grow in methanol, possibly because of electron imbalance under fermentative conditions”. What the author is trying to convey here, please rewrite the sentence.

6. Line 308: "…… central carbon metabolism pathways"? Which pathway is the author referring to? Also, please provide references. Also, discuss the roles of the ihfA and gadW mutations.
7. Line 316: "It is difficult to balance……" rewrite and provide a reference.
8. Line 131: "…. Presumably because of large protein burden". Please explain how and provide the reference.
9. Figure 2a What is the meaning of GFP, PH, and merge? Please provide some context in the figure caption
10. In Figure 2B to whom green line is referring? Also, provides the value of fluorescence intensity in lines 97-98.
11. Please provide some contexts that the DPC issue is directly related to formaldehyde toxicity.
12. Line 199: Why extracellular formaldehyde was compared only? In line 311, the author mentioned mutations in OMP. What kind of mutations were detected? Is there a possibility that formaldehyde transport will decrease due to these mutations? Please explain.
13. No references are provided in the methods section. Were all these assays used for the first time in this study? Please provide appropriate references in the materials and methods section.
14. Figure 1a: do not use the "cartoon" word in the caption. Replace with schematic, diagram, graphic, or any other suitable word.
15. Can SM6 strain grow with a doubling time of 3.5 h in 1200 mM methanol? Why was methanol decreased from 1200 mM to 400 mM? What was the final concentration of methanol to re-culture the SM6 strain?
16. Why was OD 2.0 selected only to compare fermentative products such as ethanol, acetate, and formate?

Reviewer #2

(Remarks to the Author)
Nieh et al. show based on their previous work the further evolution of the model bacterium E. coli to grow on potentially renewable substrate methanol. The combine rational introduction of pathway enzymes (including some in a region that leads to variable genome copy numbers) with serial cultivation. This in the end leads to a growth rate equalling a doubling time of 3.5 hours. The resulting strain (SM6) has a lot of mutations, probably as it got a hypermutator phenotype. Some mutations are discussed in more detail.
Despite that the result of the fast-growing strain is an impressive step, I still have a range of major and textual comments to improve this manuscript. There is still quite some sloppy things in the figures, text and also in evolutionary storyline.
Major comments
1. The author extensively discuss the ddp-BAC system that leads to variable copy numbers of inserted gene cargo. They first demonstrate that this system leads to variable GFP expression (to which the authors pay a lot of attention and figure, even though this is not the focus of the paper). Then the authors introduce the key assimilation genes (RHTTP genes) via this system to allow the cell to vary the expression levels of this system because of the suspected issue for the cells in balancing (toxic) formaldehyde assimilation. They introduce this system in strain SM1p, however based on figure 4 the growth rate of the resulting strain goes down and only recovers to the original growth rate of their previously published strain (SM1) after RecA is inactived, basically bringing the copy number of the RHTTP genes back to 1. So I disagree that this system has a major contribution. It seems to facilitate serial passaging (Figure 3c), but does not increase growth rate till RecA is mutate and not many other mutations happen. Then the authors introduce RecA again on a plasmid as well as an extra copy of MDH. Only then the growth rate increase , but also only again clearly when the RecA plasmid is lost. To me it seems introducing thie MDH made all the difference and I do not see any strong evidence that the varying copy number ddp-BAC system was key for the evolution. Also it is not clearly discussed where the MDH is inserted, from the supplementary tables I get the idea the MDH is also connected to the ddp-BAC system to allow for a different copy number, but that is not clearly explained. According to the supplementary strain table MDH was inserted into ddpA on the genome and eventually on pLY147 (the BAC), what is meant by eventually? Does this mean from that time point on the copy number system is not active anymore (ddpA is replaced). Important to not is that in the growth rates measure on from that point on the growth rate increased above the wild-type growth rate.
In the discussion (L292) the authors argue: " The presence of
292 mul8-copy RHTTP allowed the cells to accumulate beneficial mutationsons. Once other
293 alterna8ve pathways to alleviate DPC have emerged, the cell lowered the copy number
1. 294 by inac8va8ng recA." However it is not clear at all which 'beneficial mutations; (very little according to figure) emerged in this period that could explain this a better discussion of the mutations happening in this period should be included.
2. The authors claim a doubling time of 3.5 hours for strain SM6, but it's not clear on which data this is based. They refer to figure 4a, but there it seems it he growth rate of strain SM6 is 0.16 h-1 which is more than 4 hours. From other figures (6 and 8) it is not clear to me what the doubling time is.
3. The authors make a big claim in the title that the strains grows faster than native methylotrophs, but I would argue this is a bit bold, there are native PQQ-MDH dependent methylotrophs growing between 1 and 2 hours doubling time at mesophilic temperatures, and for NADH-dependent methylotrophy Bacillus methanolicus can also double faster at is optimal temperature. This should be more clearly discussed and I suggest to change the title is this comes across deceiving to me. Recently this trade-off between PQQ-MDH (supporting faster growth at mesophilic temperatures) and NADH-MDH potentially supporting high yields was discuss in a review by Kruesemann et al., 2023, Current Opinion in Biotechnology
4. In figure 7 panel C, it is stated that the reference is the wild-type, but what is the wild-type here, a native E. coli strain? How can there be a log fold change than for the Mdh, Hps, Phi, these are non-native, so should be an infinite fold-change! Also several labs are not clear to what enzyme they belong. E.g. the label next to E4P, does this belong to Tal or Tkt (one is missing a label for sure).
5. Due to the MutS mutation the SM6 strain is not stable genetically and this is an issue for downstream applications (as seen in figure 4 it's growth rate even drops again during further cultivation/evolution). This issue should be honestly discussed. The mutation could be restored to make a stable strain.
6. For a good strain for bioproduction not only growth rate but especially also yield is relevant. I urge the authors to also

make estimates on the yields, as with green methanol being still an expensive/limiting substrate this is another key parameter, that if not optimal yet, should be optimized.
7. Within a few days after submitting this work (also released as preprint) the lab of Julia Vorholt published an article in Nature Catalysis, which showed a similarly fast growing E. coli strain on methanol. This strain does not have the hyper mutator phenotype and has some interesting mutations in MDH, and the PEP/pyruvate node. They actually claim a higher upward flux through Pps, could this be a similar case in your work and that the extra copy of MDH was really key. I suggest the authors to also discuss their work in relation to that work and comparisons.
8.

Minor comments
1. Figure 1 show a comparison for the ddp-BAC system with and without RecA, however the genetic background of the strains with and without RecA is different (BW25113 and DH5alpha), preventing a clean comparison. It is known DH5alpha has several other key mutations, so this is not full proof the difference is because of RecA, even though later evolution hints to this. However, I would argue for a proper comparison this experiment should be done in BW25113 delta RecA (which is readily available in the KEIO collection).
2. The original organism and gene sequence of the introduced MDH are not discusse
3. The genome changes section (from line 177) starts by discussing the shrinking of the IS5 mediate tandem repeat, but it it's not clear what the role of the genes on it could be.
4. For most of the figures there is not clear mention how much methanol was used for the growth curves, this should be included (e.g. figure 6 and 8).
5. The authors discuss in line 304 global regulatory genes, but it is not clear to me why they not include RpoB and RpoC and this, which are both known to also be global regulatory mutations happening a lot in ALE. Also I would not say OmpC is a global regulator and the discussed reason that this mutation could be beneficial sounds very speculative to me.
6. Line 314 the authors discuss that limitation of SM6 is that it cannot do fermentation, but by the nature of methanol assimilation costing more ATP then fermentation can yield via acetate formation it can never be used for fermentation as far as I can see. So I do not see what further work the authors refer to but I find this a very confusing statement. Also in nature all RuMP organisms are strict respirers.
7. In figure 4 the authors mention before strain SM6 is obtained, 'sorting by FACs', but nowhere else in the manuscript is this step discussed.
8. There is several issues in how the proteomics data in figure 7 are presented. What is the wild-type reference (strain, growth condition not clear at all). Also in the methods these are not well described.
9. In Figure 7, panel B, it is not clear to me what the percentages refer to. Normally I am use d to seeing proteomics data with % in case of relative contributions to the proteome, but that is not what this seems to be. It seems if each protein here contributes roughly 3.5%, but this just a top 30 of upregulated proteins, that should be depicted differently. And I do wonder where the raw/full proteomics data are, they should be included supplementary.

10. There are multiple variants of the RuMP Cycle (e.g. TA/SBP/ED variants as discussed in Cotton et al., 2020), which variant the authors actually engineered
11. A plasmid table is included in the data, but not plasmid maps are included, they should for completeness also be included.
12. L77 native methylotroph should be native methylotrophs
13. L107 native methylotroph should be native methylotrophs
14. L210 Hps, Phi should be Hps and Phi
15. L215 impedes should be impede
16. L236 inserted genes is an unclear term (interreupted maybe) and it should say 'with a copy of the intact gene'
17. L241 'As methanol is… .. more so than glucose' is not a complete sentence
18. The strain construction section (L349), has very little detail (what regions were inserted were, at least reference to strain table and plasmid would be good). And how was electroporation done?
19. L317, genomic DNA is quantified with a plate reader? This is a bit vague, what reagent/wavelength is used?
20. L432 mentions growth under a high-aeration system, which data are these? This is not mentioned in the figures

Reviewer #3

(Remarks to the Author)
This manuscript describes the development of a methylotrophic E. coli strain (SM6) for the faster growth by evolutional mutation based on the development of a bacterial artificial chromosome (BAC) with dynamic copy number variation (CNV) to overcome the formaldehyde-induced DNA-protein cross linking (DPC).

The results are certainly of interest, and worth publication in Nature Communications. However, there are many questions to be clarified before acceptance in relation to the phenotypic mechanism on how the gene level mutation caused/contributed to the increase in the cell growth by regulating the central carbon metabolism.

1. In Figure 6a, the methanol consumption characteristics should also be shown to see the effect of methanol consumption rate on the cell growth.
2. From the result of Fig.6c, the roles of Tal, Tkt, and Rpi may not be critical/important for the consumption of H6P, or F6p via Phi ?
3. Although it is understandable that the increased activities of Hps and Phi contributed to the increased cell growth of SM6 by the higher consumption of methanol, it is not clear if these enzymes are under enzyme level regulation in response to the

formaldehyde concentration.
4. It is stated that SM6 changed its preferred substrate from sugars to methanol in relation to the result of Fig.8b. The substrate consumption characteristics should also be shown in Fig.8b. Note that the result of Fig.8b does not guarantee that the methanol is the preferred carbon source among other carbon sources. Since PTS may not have been mutated for the evolutional experiments using methanol, the catabolite repression by sugar such as glucose still exists, and the glucose may be the preferred carbon source even in the presence of methanol ? This can be confirmed by the experiment of using a mixture of glucose and methanol.
5. In Fig.8c, the specific production rates should be compared, instead of concentration.
6. I wonder why formate was produced by WT and SM6 under aerobic condition in Fig.8c.
7. The protein expression level as shown in Fig.7c does not reflect the metabolic fluxes in general, and thus the discussion on the result must be careful. Metabolic flux analysis should be considered.
8. It is not clear why the cell growth on glucose became lower after evolutional mutation (Fig.8b). Is this due to the modulation (up-regulation) of Cra, which represses the glycolytic pathway genes, while activating the gluconeogenic pathway genes.
9. In relation to Cra regulation, the intracellular metabolite of FBP should be measured and compared ?
10. The NADH balance should be checked based on the flux analysis.
11. I wonder why NADH produced by electron-rich substrate/methanol cannot contribute to ATP production through the respiratory chain under aerobic condition, but by bypassing the PDH pathway with Pox pathway and acetate consumption pathway (Acs) ?
12. In the ribosome synthesis from amino acids, nitrogen regulation may also play the important role, and therefore, please discuss about the nitrogen regulation via glutamate and glutamine.
13. Mased on the metabolic map, methanol is consumed to produce H6p and then to F6p, and then going down to the glycolytic pathway, and not the gluconeogenic carbon source, while the gluconeogenic pathway such as glyoxylate pathway etc. are activated instead of TCA cycle via GltA etc. This must be discussed in view of energy generation (catabolism) and biosynthesis (anabolism).

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In accordance with reviewers' comments, the authors have successfully addressed the comments and revised manuscript well. I'd like to recommend that the revised manuscript can be accepted as is for publication.

Reviewer #2

(Remarks to the Author)
Overall this work got significantly improved with additional experiments and rewriting of the work. I applaud the authors for creating a stable strain now! Als the writing of the manuscript substantially improved. Still I have some minor questions I would ask the authors to tackle:

1. It's good title is changed but I would change it a bit further to 'Evolutionary engineering' rather than 'Evolution engineering', I think that is grammatically correct. And I think 'growth with methanol' can better be replaced by 'growth on methanol'.
2. Even though several reviewers asked to compare/clearly refer to the rent paper with similar results recently published in nature Catalysis it's only briefly mentioned when discussing the mutations, I suggest to also mention it in the introduction/discussion.
3. Figure 8, pane e and f sow in the red lines the same strain and same conditions (SM7 on 400 mM methanol MOPS medium but show very different maximum OOD (1.2 and 3). So that is inconsistent how can this be explained? And also the glucose data look different. Also panel E does basically have to little sampling points to show proper growth curve I would not show this her like this .
4. L506, point missing at the end of the sentence
5. L544 'air flow' should be 'gas flow' if it's 100% O2
6. For the fed-batch experiment in Figure 8 in the bioreactor and falcon tubes (nice that these experiments are added) I miss the total amount of methanol added, this would be good to report, especially for the closed Falcon tube system this could give additional data to calculate late yield.
7. Figure 7 caption " 44 E. coli ribosomes" do you mean ribosomal genes?
8. Figure 7b I still not find clear enough, what do you mean by gene ratio on the x-axis, define this clear in caption
9. In reply to my previous comment you reported yields, but these cannot be true, 38 g/g methanol is unrealistically high yield. I think you made a calculation error. In figure S3 you show an OD increase of about 3 (*0.5 gCDW/OD = 1.5gCDW) for a consumption of about 0.1 M methanol. (0.1 M * 24 g/mol = 2.4 g methanol). So yield is something like 1.5/2.4 = 0.6 gCDW/g methanol (or 1.5/0.1 = 15 gCDW/mole methanol). This is actually close to the theoretical yield of the RuMP (see e.g. Cotton et al., Current opinion). So I suggest to highlight and also show a correct yield estimation in the manuscript text (with the remark its it not real CDW measured but estimated from OD600)
10. Title at start page supplementary files does not match with manuscript title

11. Good that plasmids maps are included now, but for reproducibility also plasmid sequences should be included, I did not spot them in supplementary

Reviewer #3

(Remarks to the Author)
The manuscript has been appropriately modified, and improved, and therefore, the MS can be now accepted.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I compliment the authors on their revised manuscript and all my comments have been properly addresses. There is one small typo that I spotted in revised sentence in a L189: ‘Falcone’ should be spelled ‘Falcon’

Response to Reviewer's comments

Reviewer #1 (Remarks to the Author):

o General comments

- In this study, the author reported the development of a synthetic methylotrophic strain with a doubling time of 3.5 h in methanol. In the previous study by the same group, the E. coli strain SM1 was able to grow on methanol (10.1016/j.cell.2020.07.010) with a doubling time of 8 h. In this study, bacterial artificial chromosome (BAC) was used in the SM1 strain, which helped overcome the formaldehyde-induced toxicity using dynamic copy number variation (CNV), further enabling the adaptive evolution of the SM1 strain. This study is useful for the scientific community in three aspects. 1) The advantage of CNV in overcoming the bottlenecks of adaptive evolution 2) synthetic methylotroph which can grow better than native methylotrophs 3) provides meaningful insights regarding the metabolic rewiring of synthetic methylotrophy.

The authors thank the reviewer for the positive comments.

- Although the doubling time of SM6 has been improved considerably in this study, the maximum OD achieved was similar to that of the SM1 strain (10.1016/j.cell.2020.07.010). Moreover, the scale-up fermentation in methanol minimal medium has not been shown in this study, and MOPS media was used, which requires vitamin supplementation.

Thank you for your suggestion. Following your suggestion, we tried different methods of cultivation and found out that the maximum biomass of SM8 can reach 20 g/L in a bioreactor using methanol modified-M9 medium supplemented with trace amounts of vitamins (8.19 µM biotin, 10.49 µM calcium pantothenate, 4.53 µM folic acid, 40.94 µM nicotinamide, 13.29 µM riboflavin, 14.82 µM thiamine hydrochloride, and 0.07 µM vitamin B12) (Line 325). Also, SM6 can grow on MOPS-based methanol medium without vitamin supplementation (data not shown).

Also, the role of CNV and BAC in metabolic rewiring is unanswered since previous studies from the same group have been conducted; therefore, further experimental data or discussion is required for better understanding.

We have added the following descriptions:

1) CNV in the parental strain SM1: Although in SM1 we only saw IS-mediated CNVs, they provide useful inspiration.

   Lines 82-90: "Previously, we reported a synthetic methylotroph[5], SM1, that grows on methanol as the sole carbon source with a doubling time of 8 hours. This strain features a chromosomal segment containing 5 copies of IS5-mediated tandem repeats, each spanning 70Kb. This result suggested that transposon-mediated copy number variation (CNV) of methanol assimilation genes may be a major breakthrough in the evolution of a synthetic methylotroph. However, increasing copy number through transposons is a random process, and cannot generate a broad range of copy numbers. While this strain has alleviated the DPC problem, its growth rate (T_d= 8.5 hr) is still far below that of native methylotrophs."

2) CNV in the strain BB1 co-evolved with SM1, which provides the inspiration for the *ddp*-induced CNV.

   Lines 106-113: "To address the DPC issue that hinders long-term evolution, we developed a dynamic CNV approach to tune the expression levels of formaldehyde consumption enzymes. This scheme is based on our previous discovery[5] of yet another tandem repeat of 7Kb region in the strain (BB1) (see Supplementary Fig. 1) which spontaneously emerged with the synthetic methylotrophic SM1 strain that can utilize methanol as the sole carbon source for growth (see Supplementary Fig. 2 for ^{13}C labeling data). Interestingly, this 7K tandem repeat is non-IS mediated, and contains the *ddp* operon."

3) The role of ddp-BAC to remove the bottleneck in evolution during passage 81-132.

   Lines 380-404: Since the descendent of SM1 after 81 passages (SM1p) did not improve growth much and still produced formaldehyde in high methanol concentrations (Fig. 3a&b), we over-expressed the formaldehyde consuming enzymes RHTTP using ddp-BAC, BAC, or pUC19. The ddp-BAC produces tandem repeats concatemers with a wide range of copy numbers, which provides a library of various RHTTP copy numbers for competition during evolution. Without *ddp*, the BAC system exists in a single copy in the cell. In contrast, plasmid pUC19 consistently exists in high copy (~500). We found that only ddp-BAC systems expressing RHTTP

can cause the SM1p to grow in high methanol concentrations (1000mM) consistently (Fig. 3c). This result suggests that the SM1p was able to find the acceptable copy numbers of RHTTP in the ddp-BAC library for growth in such a condition, and removed the bottleneck in evolution. The metabolic burden introduced by the concatemers of ddp-BAC was tentatively tolerated.

During this period, the cells was able to grow and evolve. Beneficial mutations that improved growth or reduced formaldehyde production using more effective ways occurred. For example, the 70Kb IS5-mediated tandem repeat region that emerged during the evolution of the parental strain SM1 shrank to 46Kb, which greatly reduced the protein burden and possibly allowed more balanced expression of the formaldehyde consuming enzymes. Other mutations that occurred during this time may directly or indirectly contribute to the balanced consumption of formaldehyde. After 50 passages of evolution, the ddp-BAC copy number was reduced by inactivating *recA* to reduce the metabolic burden. Presumably the cells had found a better way to balance the formaldehyde metabolism.

After other mutations accumulates, the copy number returned to a low level, demonstrating the dynamic tuning of copy numbers to fit the particular genetic background.

4) The role of ddp-BAC in passages 165-200

Lines 407-412: This phenomenon was demonstrated again at passages 165 by reinstalling *recA* using a multicopy plasmid. At this genetic background, the cells evolved to find the optimal copy number of MRHTTP on ddp-BAC to be 2, and purged the *recA* plasmid at passage 200 (Fig. 4). When the 2-copy ddp-BAC was removed from SM7, the strain grew slower, indicating the contribution of the ddp-BAC system.

- In comparison, a very recent publication by Michael et. al. (doi.org/10.1038/s41929-024-01137-0) used adaptive evolution to construct a synthetic methylotroph that can grow in the methanol-minimal medium in a fed-batch fermenter. Although the doubling time (4.3 h) of the strain constructed by Michael et. al. was lower than in this study, in fermentation a remarkable OD of 100.2 was achieved. In addition, the production of lactic acid, PHB, and etc. was

demonstrated using a synthetic methylotrophy strain. Therefore, to consider this manuscript for publication in Nature Communications, at least the fermentative growth of synthetic methylotrophs needs to be shown. Also, I suggest the author include this reference (doi.org/10.1038/s41929-024-01137-0) for discussion and comparison.

Following the reviewer suggestion, we conducted a fed-batch fermentation in a bioreactor and showed a maximum cell density achieved in both modified M9 and modified MOPS.

Line 324: In a bioreactor using two kinds of methanol minimal medium: MOPS-based and M9-based, SM8 can reach maximal dry cell weight of 19.9 and 12 g DCW L^{-1}, respectively (Fig. 8b & c).

The maximal biomass achieved in a bioreactor is comparable to that of the Vorholt's group (22 g/L). Further optimization may be required for performance in bioreactors, which is outside of the present scope.

- Moreover, the discussion part is poorly written without explanation and requires references. The mutations in genes, for example, OMPs, need to be characterized and explained properly for better understanding.

Thank you for the suggestion. The characterization of mutations is summarized in Fig. 8a and is discussed below.

Lines 283-317: There are several mutations on SM7 (Fig5 & Supplementary Table S2). We tested the effect of a few, including gene (*gatY, gtrS, quuD, rusA, ylcG*) and promoter interruption by insertion sequences (*gltA*), codon change in protein coding regions (cra, *gadW, ihfA, ompR*) and frameshift (*rhlB*). To probe the effect of these mutations on growth, we complemented these disrupted genes with a copy of the intact gene on BAC in SM7n and tested their effects on growth rate. Results (Fig. 8a) showed that the complementation by BAC-borne *gltA*, *rhlB* and *ihfA* genes significantly reduced the growth rate of SM7n. This result indicates that the *gltA* insertion by IS2 in the promoter region and *rhlB* frameshift were beneficial for growth. GltA is the first enzyme in the TCA cycle, which produces NADH. As methanol is a more reduced substrate than glucose, a strong TCA cycle would increase the burden of respiration. In the SM7 strain the *gltA* promoter was interrupted by IS2, but it did not completely abolish the gltA expression.

Proteome data suggested that GltA protein was still produced, albeit at about 10% level compared to the wild-type strain (Fig. 7c). Thus, it is possible that by properly knocking down *gltA*, the cell could balance NADH flux better. The same promoter mutation actually existed in the SM1 strain, however, re-introducing a wild-type gltA copy did not affect SM1 growth significantly5.

The rhlB gene encodes an ATP-dependent RNA helicase. In SM7, this gene contains a frameshift mutation and abolishes its function. It has been reported that overexpression of rhlB "increased fitness and buffered deleterious mutations" in E. coli mutator strains30. The inactivation of rhlB probably increased the effect of mutS mutation and accelerated evolution. However, its exact role in SM7 requires further investigation.

We also tested whether the mutations in global regulatory genes—*cra* (E190Q), ompR (R17C), *ihfA* (V78E), and *gadW* (frameshift)—played a critical role in the enhanced growth rate of SM7 (Fig. 5). The complementation of *cra* (E190Q), *ompR* (R17C), and gadW (frameshift) by their respective wild-type copy using BAC did not show any difference in growth, suggesting that these mutations did not contribute to the fast-growth phenotype. However, the complementation of *ihfA* (V78E) by its wild-type version using BAC significantly decreased the growth rate of SM7n. Thus, this mutation appears to be beneficial for the growth phenotype. The *ihfA* gene codes for a global transcriptional regulator affecting the expression of hundreds of genes, including those involved in amino acid and fatty acid metabolism, making its role difficult to elucidate at this point.

o Specific comments:

1. Figure 6b: TEM images were taken until OD 1.3. SM6 grew to OD 1.3 in less than 20 h, whereas SM1 took almost 30+ h. Why were no further time points compared? In the previous publication, it was mentioned that SM1 alleviated the DPC problem, but it seems higher with the passage of time in TEM images. Therefore, please explain Figure 6b in line 205.

Thank you for your comment. We selected the OD range where the results are clearly differentiable when DPC is present. In addition to use TEM, we also use lag time as an indication for the presence of DPC. In this case, the further evolved strain (now renamed as SM7) did not show a lag time.

2. Line 216: How does DPC interfere with growth? Add some explanation with references.

The following sentence is added in to the main text in lines 52-57 with appropriate references:

Lines 52-57: Formaldehyde is an essential intermediate in methanol utilization, but its accumulation can induce DNA-protein crosslinking (DPC), which forms aggregates between DNA and proteins that disrupt cellular functions. Severe DNA-protein crosslinking can ultimately cause cell death[16,17]. Native methanol-utilizing organisms have evolved sophisticated regulatory mechanisms to avoid DPC by balancing the formaldehyde production and consumption[18].

Klein VJ, Irla M, Gil López M, Brautaset T, Fernandes Brito L. Unravelling Formaldehyde Metabolism in Bacteria: Road towards Synthetic Methylotrophy. Microorganisms. 2022 Jan 20;10(2):220. doi: 10.3390/microorganisms10020220. PMID: 35208673; PMCID: PMC8879981.

Bazurto, Jannell V., et al. "EfgA is a conserved formaldehyde sensor that leads to bacterial growth arrest in response to elevated formaldehyde." PLoS biology 19.5 (2021): e3001208.

3. Line 219: The increase in ribosome abundance may be a key reason underlying the fast growth of SM6. Please explain with references.

Thank you for the comment. We had added the following.

Lines 269-271: It has been reported that ribosome amount is positively associated with the cell's growth rate due to an increase in the overall translation rate in cells[28,29]

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., & Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science, 330*(6007), 1099-1102.

Dai, X., Zhu, M., & Warren, M. (2018). Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nature Microbiology, 3*(5), 618-626

4. Line 224: "These changes suggest that SM6 derives 2-carbon compounds not from pyruvate dehydrogenase (coded by aceEF) or pyruvate-formate lyase (coded by pflB) but

from pyruvate oxidase (coded by poxB) and acetate utilization (coded by acs, glcB)". Is the poxB route advantageous in terms of energy and carbon conservation?

This is an interesting point. We added the following in Discussion:

Lines 446-454: From the proteomic analysis, we found that AceEF, PflB, GltA decreased while PoxB, Acs and GlcB increased. GltA expression decrease was caused by an IS insertion in its promoter region that reduces its expression. The PoxB route is actually less energy efficient than the AceEF route, as it produces acetate, which requires more ATP for reuptake. As methanol is richer in electron (and energy) than glucose, *E. coli* which typically grows in sugar-abundant environment has to find a way to deal with surplus electrons and energy. Thus, the PoxB route and the reduced TCA cycle was evolved to dump excess carbon to acetate. Fermentation in bioreactor also demonstrated that the cells produced acetate and consumed it later.

Reference: Abdel-Hamid, A. M., Attwood, M. M., & Guest, J. R. (2001). Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli. Microbiology, 147(6), 1483-1498.

5. Line 262: "Neither SM1 nor SM6 could grow in methanol, possibly because of electron imbalance under fermentative conditions". What the author is trying to convey here, please rewrite the sentence.

We initially meant: Neither SM1 nor SM6 could grow anaerobically in methanol, possibly due to electron imbalance in the absence of oxygen.

To avoid confusion and too much speculation, we deleted this line form in the revision.

6. Line 308: "…… central carbon metabolism pathways"? Which pathway is the author referring to? Also, please provide references. Also, discuss the roles of the ihfA and gadW mutations.

We deleted this ambiguous line and rewrote the section:

The effects of mutations were summarized in Fig. 8a, using BAC to restore the mutation. Specifically, *gadW* had no detectable effect, and *ihfA* exhibited a significant effect, which is discussed in the following:

Lines 312-317: However, the complementation of ihfA (V78E) by its wild-type version using BAC significantly decreased the growth rate of SM7n. Thus, this mutation appears

to be beneficial for the growth phenotype. The ihfA gene codes for a global transcriptional regulator affecting the expression of hundreds of genes, including those involved in amino acid and fatty acid metabolism, making its role difficult to elucidate at this point.

Lines 435-437: Previous studies have reported that rhlB and ihfA genes modulate the proteome of bacterial cells. Specifically, rhlB deletion modulates ribosome biogenesis35, and ihfA deletion elevates amino acid biosynthesis36. These genotypic changes align with our observations on proteomic changes

7. Line 316: "It is difficult to balance……" rewrite and provide a reference.

We deleted this line in the revision.

8. Line 131: "…. Presumably because of large protein burden". Please explain how and provide the reference.

We amended the text with references at Line 157-159: "presumably due to high metabolic burden. High-copy-number plasmids often inhibit growth by diverting cellular resources away from essential components."

It has been reported that the overexpression of recombinant protein leads to growth retardation in E. coli. The reason is that the overexpression of heterologous protein requires a large amount of resources, including ribosome allocation, ATP, and amino acids, which disrupt proteome allocation and diverts resources away from growth. The following references have been added to the main text.

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature, 436(7050), 588-592.

Rajacharya, G. H., Sharma, A., & Yazdani, S. S. (2024). Proteomics and metabolic burden analysis to understand the impact of recombinant protein production in E. coli. Scientific Reports, 14(1), 12271.

Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol. Bioeng. 35, 668–681 (1990).

.

9. Figure 2a What is the meaning of GFP, PH, and merge? Please provide some context in the figure caption

We apologize for the missing information. 'GFP' represents the view of green fluorescent protein fluorescence, 'PH' represents the view of phase contrast optical microscopy, and 'Merge' is the composite image combining both the GFP and the PH views. Such explanations are added into the figure captions of Fig.2a.

10. In Figure 2B to whom green line is referring? Also, provides the value of fluorescence intensity in lines 97-98.

The green line refers to the cutoff between positive and negative signals. This cutoff line is determined by the background signal of pure water. We added relevant information in the legend of Figure 2Bb Also, the fluorescence intensity was added in the main text at Line 122-123.

11. Please provide some contexts that the DPC issue is directly related to formaldehyde toxicity.

Formaldehyde can form covalent bonds between DNA and proteins, creating aggregates that disrupt cellular function and induce oxidative stress, leading to genotoxicity and proteotoxicity. Ultimately, cells will die from severe DNA-protein crosslinking (DPC). To increase the paper's clarity, we further explained DPC and formaldehyde toxicity in line 52 to 55, and added the following references to the main text.

Lines 52-55: "Formaldehyde is an essential intermediate in methanol utilization, but its accumulation can induce DNA-protein crosslinking (DPC), which forms aggregates between DNA and proteins that disrupt cellular functions. Severe DNA-protein crosslinking can ultimately cause cell death[16,17]"

Klein, V. J., Irla, M., Gil López, M., Brautaset, T., & Fernandes Brito, L. (2022). Unravelling formaldehyde metabolism in bacteria: road towards synthetic methylotrophy. *Microorganisms*, *10*(2), 220.

Bazurto J.V., Riazi S., Alton S.D., Deatherage D.E., Bruger E.L., Barrick J.E., Marx C.J. Global Transcriptional Response of Methylorubrum extorquens to Formaldehyde Stress Expands the Role of EfgA and Is Distinct from Antibiotic Translational Inhibition. Microorganisms. 2021;9:347. Doi: 10.3390/microorganisms9020347

12. Line 199: Why extracellular formaldehyde was compared only? In line 311, the author mentioned mutations in OMP. What kind of mutations were detected? Is there a possibility that formaldehyde transport will decrease due to these mutations? Please explain.

Since formaldehyde is a very small molecule, it can penetrate the cell membrane without a transporter. This is why it is so toxic to all kinds of cells. Additionally, the intracellular formaldehyde is too reactive and difficult to detect. Instead, we used TEM to image the formaldehyde toxicity, as the DPC complexes are extracted from the cells.

ompR mutation is codon change in the protein coding regions: R17C. This information is added to line 308.

We reinstalled the wild-type *ompR* gene in SM7 (renamed from SM6) and found out there is no growth difference (Fig. 8a). Accordingly, we have eliminated the speculation regarding *ompR*.

13. No references are provided in the methods section. Were all these assays used for the first time in this study? Please provide appropriate references in the materials and methods section.

Thank you for your comment. We added references in the entire section.

14. Figure 1a: do not use the "cartoon" word in the caption. Replace with schematic, diagram, graphic, or any other suitable word.

Noted with thanks and replaced.

15. Can SM6 strain grow with a doubling time of 3.5 h in 1200 mM methanol? Why was methanol decreased from 1200 mM to 400 mM? What was the final concentration of methanol to re-culture the SM6 strain?

The SM7 (renamed from SM6) strain grow in 1200 mM methanol with a doubling time of 5.9 hours (Data not shown). Using high concentration methanol is a strategy to impose an evolution pressure for the strain to solve the DPC problem. All the methanol concentration for re-culturing SM6 was 400 mM.

16. Why was OD 2.0 selected only to compare fermentative products such as ethanol, acetate, and formate?

In the revised reversion, we conducted a lab-scale bioreactor fermentation and reported the fermentation profile in Fig. 8d and Line 330.

Reviewer #2 (Remarks to the Author):

Nieh et al. show based on their previous work the further evolution of the model bacterium E. coli to grow on potentially renewable substrate methanol. The combine rational introduction of pathway enzymes (including some in a region that leads to variable genome copy numbers) with serial cultivation. This in the end leads to a growth rate equalling a doubling time of 3.5 hours. The resulting strain (SM6) has a lot of mutations, probably as it got a hypermutator phenotype. Some mutations are discussed in more detail.

Despite that the result of the fast-growing strain is an impressive step, I still have a range of major and textual comments to improve this manuscript. There is still quite some sloppy things in the figures, text and also in evolutionary storyline.

Major comments

1. The author extensively discuss the ddp-BAC system that leads to variable copy numbers of inserted gene cargo. They first demonstrate that this system leads to variable GFP expression (to which the authors pay a lot of attention and figure, even though this is not the focus of the paper). Then the authors introduce the key assimilation genes (RHTTP genes) via this system to allow the cell to vary the expression levels of this system because of the suspected issue for the cells in balancing (toxic) formaldehyde assimilation. They introduce this system in strain SM1p, however based on figure 4 the growth rate of the resulting strain goes down and only recovers to the original growth rate of their previously published strain (SM1) after RecA is inactived, basically bringing the copy number of the RHTTP genes back to 1.

> We would like to point out that the growth rate did not go down after the introduction of RHTTP, but rather is because of growing the strain in high methanol (1200mM) (Fig. 4a) Thus, the ddp-BAC was indeed beneficial for strain evolution, since the growth rate in 1200mM eventually improved to be the same as the original growth rate of 400mM. To avoid such confusions, we changed the growth rate curve at 1000mM and 1200mM with different colors to highlight this difference. (Fig. 4a)

So I disagree that this system has a major contribution. It seems to facilitate serial passaging (Figure 3c), but does not increase growth rate till RecA is mutate and not many other mutations happen. Then the authors introduce RecA again on a plasmid as well as an extra copy of MDH. Only then the growth rate increase, but also only again clearly when the RecA

plasmid is lost. To me it seems introducing the MDH made all the difference and I do not see any strong evidence that the varying copy number ddp-BAC system was key for the evolution.

This is indeed an intricate point, and we revised the manuscript extensively to explain the benefit of ddp-BAC. It is best summarized in discussion:

Starting from Line 370:

A BAC system with dynamic copy number tuning capability allowed the cell to by-pass the DPC problem

In this work, we used the ddp-BAC system that exhibits a dynamic CNV property to accomplish the "gain-of-activity" function, which is difficult to emerge through evolution. The presence of *ddp* sequence causes the whole BAC plasmid to generate tandem repeat concatemers with a wide range of copy numbers through a *recA*-dependent mechanism (Figs. 1&2). Such a system provides an expedient fix to bypass a bottleneck in evolution while the cell seeks for a better solution. In parallel to our effect in understanding the mechanistic details, we employed this system to solve the DPC problem in synthetic methylotrophic *E. coli*. The dynamic CNV is reminiscent of the copy number engineering technique[34] but with a different mechanistic basis.

Since the descendent of SM1 after 81 passages (SM1p) did not improve growth much and still produced formaldehyde in high methanol concentrations (Fig. 3a&b), we over-expressed the formaldehyde consuming enzymes RHTTP using ddp-BAC, BAC, or pUC19. The ddp-BAC produces tandem repeats concatemers with a wide range of copy numbers, which provides a library of various RHTTP copy numbers for competition during evolution. Without *ddp*, the BAC system exists in a single copy in the cell. In contrast, plasmid pUC19 consistently exists in high copy (~500). We found that only ddp-BAC systems expressing RHTTP can cause the SM1p to grow in high methanol concentrations (1000mM) consistently (Fig. 3c). This result suggests that the SM1p was able to find the acceptable copy numbers of RHTTP in the ddp-BAC library for growth in such a condition, and removed the bottleneck in evolution. The metabolic burden introduced by the concatemers of ddp-BAC was tentatively tolerated.

During this period, the cells was able to grow and evolve. Beneficial mutations that improved growth or reduced formaldehyde production using more effective ways occurred. For example, the 70Kb IS5-mediated tandem repeat region that emerged during the

evolution of the parental strain SM1 shrank to 46Kb, which greatly reduced the protein burden and possibly allowed more balanced expression of the formaldehyde consuming enzymes. Other mutations that occurred during this time may directly or indirectly contribute to the balanced consumption of formaldehyde. After 50 passages of evolution, the ddp-BAC copy number was reduced by inactivating *recA* to reduce the metabolic burden. Presumably the cells had found a better way to balance the formaldehyde metabolism.

After other mutations accumulates, the copy number returned to a low level, demonstrating the dynamic tuning of copy numbers to fit the particular genetic background. This phenomenon was demonstrated again at passages 165 by reinstalling *recA* using a multicopy plasmid. At this genetic background, the cells evolved to find the optimal copy number of MRHTTP on ddp-BAC to be 2, and purged the *recA* plasmid at passage 200 (Fig. 4). When the 2-copy ddp-BAC was removed from SM7, the strain grew slower, indicating the contribution of the ddp-BAC system.

Also it is not clearly discussed where the MDH is inserted, from the supplementary tables I get the idea the MDH is also connected to the ddp-BAC system to allow for a different copy number, but that is not clearly explained. According to the supplementary strain table MDH was inserted into ddpA on the genome and eventually on pLY147 (the BAC), what is meant by eventually? Does this mean from that time point on the copy number system is not active anymore (ddpA is replaced). Important to not is that in the growth rates measure on from that point on the growth rate increased above the wild-type growth rate.

We revised the text at line 176: "We then focused on increasing methanol consumption by introducing additional copies of *mdh* (an improved *mdh* from Cupriavidus necator N-1 through directed evolution26) into the *ddpA* site on the chromosome and the ddp-BAC. ". We confirmed that even with an inactivated *ddpA*, the system can still generate tandem repeat concatemers, as shown in Fig. 4.

1. In the discussion (L292) the authors argue: " The presence of
292 mul8-copy RHTTP allowed the cells to accumulate beneficial mutationsons. Once other
293 alterna8ve pathways to alleviate DPC have emerged, the cell lowered the copy number
294 by inac8va8ng recA." However, it is not clear at all which 'beneficial mutations; (very little according to figure) emerged in this period that could explain this a better discussion

of the mutations happening in this period should be included.

Thank you for your suggestion. We summarized the beneficial changes in Lines 394-455 and Fig. 8b. Briefly these include

1) Increased expression of Hps and Psi in the late phase.
2) Shrinkage of the 70Kb IS-mediated tandem repeat to 46Kb leads to an increase of ribosome expression that supports better growth.
3) Increased ribosomal abundance and amino acid pathway enzyme expression.
4) *rhlB* (frameshift), *ihfA*(V78E) and *gltA* promoter insertion *by IS2*
5) Metabolic requiring to use the PoxB pathway for dumping excess energy and reduction of GltA expression to avoid excess NADH production in the electron-rich methanol medium.

2. The authors claim a doubling time of 3.5 hours for strain SM6, but it's not clear on which data this is based. They refer to figure 4a, but there it seems it he growth rate of strain SM6 is 0.16 h-1 which is more than 4 hours. From other figures (6 and 8) it is not clear to me what the doubling time is.

The growth curve of SM7 (renamed for clarity) in 400 mM methanol is displayed in new Figure 3d, as well as other growth curves in Fig. 8b. And the specific growth rate of SM7 in Figure 4a reaches 0.2 h^{-1}.

3. The authors make a big claim in the title that the strains grows faster than native methylotrophs, but I would argue this is a bit bold, there are native PQQ-MDH dependent methylotrophs growing between 1 and 2 hours doubling time at mesophilic temperatures, and for NADH-dependent methylotrophy Bacillus methanolicus can also double faster at is optimal temperature. This should be more clearly discussed and I suggest to change the title is this comes across deceiving to me. Recently this trade-off between PQQ-MDH (supporting faster growth at mesophilic temperatures) and NADH-MDH potentially supporting high yields was discuss in a review by Kruesemann et al., 2023, Current Opinion in Biotechnology.

We changed the title to "Evolution engineering of methylotrophic E. coli enables fast growth with methanol" following the reviewer's comment.

4. In figure 7 panel C, it is stated that the reference is the wild-type, but what is the wild-type here, a native E. coli strain? How can there be a log fold change than for the Mdh, Hps, Phi, these are non-native, so should be an infinite fold-change! Also several labs are not clear to what enzyme they belong. E.g. the label next to E4P, does this belong to Tal or Tkt (one is missing a label for sure).

We apologize for the confusion. Yes, the WT is a native E. coli strain growing in glucose. We amended the text in Line273-274 "Comparing the proteomes of SM7 and the wild-type parental BW25113 grown in 1% glucose,"

The non-native enzyme protein values in WT were imputed by the proteomics data analysis. In the revised version, these ratios have been changed to SM7/SM1 ratios for better clarity.

Thank you for pointing out the missing label in Fig. 7c. We also added lines to link the enzyme to the boxes for better clarity.

5. Due to the MutS mutation the SM6 strain is not stable genetically and this is an issue for downstream applications (as seen in figure 4 it's growth rate even drops again during further cultivation/evolution). This issue should be honestly discussed. The mutation could be restored to make a stable strain.

Thank you for your comment. Following your suggestion, we restored our final strain and established a SM8 strain that is genotypically and phenotypically stable. This strain also exhibits a doubling time of 3.5 hrs. See Line 241-245: "To stabilize the strain, the wild-type mutS was restored in SM7 using the Lambda-Red system21, resulting in strain SM8, which is genetically stable with a mutation rate similar to the wild-type strain (Supplementary Fig. 5). After sub-culturing for 15 passages in methanol, the growth rate of SM8 remained around 3.5 hours and no more mutations were observed (Fig. 4 & 5)."

6. For a good strain for bioproduction not only growth rate but especially also yield is relevant. I urge the authors to also make estimates on the yields, as with green methanol being still an expensive/limiting substrate this is another key parameter, that if not optimal yet, should be optimized.

In response to the request for yield estimates, we calculated the yields based on data

from Supplementary Figure 3. We employed the widely accepted conversion factor, where an OD_{600} of 1.0 is equivalent to approximately 0.5 g CDW/L of E. coli culture, based on our experimental result. The biomass yield for strain SM1 was calculated to be 30.9 (g biomass/g methanol), while strain SM6 exhibited a yield of 38.8 (g biomass/g methanol). That indicates SM7 is more efficient at converting methanol into the biomass compared to SM1.

Ref:

Simensen V, Schulz C, Karlsen E, Bråtelund S, Burgos I, Thorfinnsdottir LB, García-Calvo L, Bruheim P, Almaas E. Experimental determination of Escherichia coli biomass composition for constraint-based metabolic modeling. PLoS One. 2022 Jan 27;17(1):e0262450. doi: 10.1371/journal.pone.0262450. PMID: 35085271; PMCID: PMC8794083.

7. Within a few days after submitting this work (also released as preprint) the lab of Julia Vorholt published an article in Nature Catalysis, which showed a similarly fast growing E. coli strain on methanol. This strain does not have the hyper mutator phenotype and has some interesting mutations in MDH, and the PEP/pyruvate node. They actually claim a higher upward flux through Pps, could this be a similar case in your work and that the extra copy of MDH was really key. I suggest the authors to also discuss their work in relation to that work and comparisons.

Thank you for the comment. The Vorholt strain is actually a hypermutant strain caused by a *dnaQ* mutation (their previous work back in 2022), with more than 1000 mutations. Following your suggestion, we added text for the comparison of their strain from line 439-444.

Minor comments

1. Figure 1 show a comparison for the ddp-BAC system with and without RecA, however the genetic background of the strains with and without RecA is different (BW25113 and

DH5alpha), preventing a clean comparison. It is known DH5alpha has several other key mutations, so this is not full proof the difference is because of RecA, even though later evolution hints to this. However, I would argue for a proper comparison this experiment should be done in BW25113 delta RecA (which is readily available in the KEIO collection).

Edited and replaced with BW25113 delta recA.

2. The original organism and gene sequence of the introduced MDH are not discussed.

We amended the text in line 176: We then focused on increasing methanol consumption by introducing additional copies of *mdh* (an improved *mdh* from Cupriavidus necator N-1 through directed evolution) into the ddpA site on the chromosome and the ddp-BAC.

3. The genome changes section (from line 177) starts by discussing the shrinking of the IS5 mediate tandem repeat, but it it's not clear what the role of the genes on it could be.

The large segment of tandem repeats increased the metabolic burden for protein expression and disrupted the proteome balance. Thus, the shorter segment of 46Kb of tandem repeats that still retain the beneficial genes (*gsh* and RuMP pathway genes) was selected. The reduced metabolic burden presumably contributed to the higher ribosome amounts that is beneficial to growth. The entire amended text is in Line 215-232.

4. For most of the figures there is not clear mention how much methanol was used for the growth curves, this should be included (e.g. figure 6 and 8).

Amended with thanks!

5. The authors discuss in line 304 global regulatory genes, but it is not clear to me why they not include RpoB and RpoC and this, which are both known to also be global regulatory mutations happening a lot in ALE. Also I would not say OmpC is a global regulator and the discussed reason that this mutation could be beneficial sounds very speculative to me.

We did not characterize the *rpo* genes due to several reasons: 1) As the reviewer said, it is a common mutation in laboratory evolution, thus we thought it would not be specific to methanol growth. 2). It is relatively difficult to characterize it since its native promoter

is distanced from the gene up to 4kb, with around 4 genes in the middle. 3.) The effect may be combinatorial, as rpoBC is coupled together to be expressed.

We deleted any discussion related to OmpC. However, we did not find any place that OmpC was referred to as a global regulator. The only place we mentioned OmpC in the original version was "The EnvZ/OmpR pair is a two-component regulatory system that is responsible for osmotic response, as it regulates the expression of outer membrane proteins (OMPs) like OmpF and OmpC." We deleted these sentences.

6. Line 314 the authors discuss that limitation of SM6 is that it cannot do fermentation, but by the nature of methanol assimilation costing more ATP then fermentation can yield via acetate formation it can never be used for fermentation as far as I can see. So I do not see what further work the authors refer to but I find this a very confusing statement. Also in nature all RuMP organisms are strict respirers.

We eliminated this statement to avoid confusion.

7. In figure 4 the authors mention before strain SM6 is obtained, 'sorting by FACs', but nowhere else in the manuscript is this step discussed.

We fixed this in Lines 181-184. Quote: "At passage 259, we isolated a culture by fluorescence-activated cell sorting (FACS), which had a doubling time of less than 4 hours and named it SM6. From this culture, we continued to evolve for 28 passages and isolated a strain (SM7) with an even better growth rate, having a doubling time of 3.5 hours (growth rate = 0.2 hr^{-1}) "

8. There is several issues in how the proteomics data in figure 7 are presented. What is the wild-type reference (strain, growth condition not clear at all). Also in the methods these are not well described.

Thank you for pointing out this. The WT reference is shown in Line 273-275: "Comparing the proteomes of SM7 and the wild-type parental BW25113 grown in 1% glucose, we found that the RuMP cycle enzymes were increased as expected." We also amended this info in the figure 7c legend. We re-analyzed the proteomics data with gene set

enrichment analysis, and re-wrote the entire section of proteomics for better clarity.

9. In Figure 7, panel B, it is not clear to me what the percentages refer to. Normally I am use d to seeing proteomics data with % in case of relative contributions to the proteome, but that is not what this seems to be. It seems if each protein here contributes roughly 3.5%, but this just a top 30 of upregulated proteins, that should be depicted differently. And I do wonder where the raw/full proteomics data are, they should be included supplementary.

As answered in 8., we re-analyzed the result with GSEA analysis with Fig 7b. We also replotted the top 30 entries with protein abundance ratio in Suppl. Figure 6. Also, the raw proteomics data was requested by the editor shortly after submission, with PRIDE accession ID: PXD051688.

10. There are multiple variants of the RuMP Cycle (e.g. TA/SBP/ED variants as discussed in Cotton et al., 2020), which variant the authors actually engineered

We just simply used the most traditional RuMP cycle. Since we did not overexpress SBP, there wouldn't exists SNP modes in our system. With the MHPTTRR present, the cycle would be able to work in both EMP (TA) & ED mode, thus indifferentiable.

11. A plasmid table is included in the data, but not plasmid maps are included, they should for completeness also be included.

Following the reviewer suggestion, we added plasmids map in the appendix.

12. L77 native methylotroph should be native methylotrophs

Fixed, thanks.

13. L107 native methylotroph should be native methylotrophs

Fixed, thanks.

14. L210 Hps, Phi should be Hps and Phi

Fixed, thanks.

15. L215 impedes should be impede

Fixed, thanks.

16. L236 inserted genes is an unclear term (interreupted maybe) and it should say 'with a copy of the intact gene'

Fixed, thanks.

17. L241 'As methanol is… .. more so than glucose' is not a complete sentence

Sentence deleted, thanks

18. The strain construction section (L349), has very little detail (what regions were inserted were, at least reference to strain table and plasmid would be good). And how was electroporation done?

We revised the entire section with more details. Line 498 Quote:" The electroporation method was performed as follows: 1 ml of bacterial culture was washed twice with 1 ml of ice-cold 10% glycerol. The bacterial pellet was then resuspended in 100 µl of 10% glycerol and mixed with 50-100 ng of plasmid DNA for electroporation at 1.8kV. Finally, the cells after electroporation were recovered for 30 minutes and plated on LB agar plates containing the appropriate antibiotics."

19. L317, genomic DNA is quantified with a plate reader? This is a bit vague, what reagent/wavelength is used?

We added the following text at line 570: The concentration of genomic DNA (gDNA) was determined using the BioTek Synergy H1 Plate Reader (Agilent Technologies) by measuring absorbance at 260 nm, 280 nm, and 230 nm. The concentration and purity of gDNA were calculated using the ratios of A260/A280 and A260/A230.

20. L432 mentions growth under a high-aeration system, which data are these? This is not mentioned in the figures.

We have removed this confusing statement. We revised with the following text at line 647:

In the RTS-8 Plus system, we cultivated bacteria in a volume of 20 mL in a 50 mL Falcon tube with an air-permeable membrane, spinning at 2700 rpm. Although we did not directly measure the dissolved oxygen level in the cultures running with the Biosan

system, the aeration level is very likely to be higher than that of common bacterial cell culture tubes.

Reviewer #3 (Remarks to the Author):

This manuscript describes the development of a methylotrophic E. coli strain (SM6) for the faster growth by evolutional mutation based on the development of a bacterial artificial chromosome (BAC) with dynamic copy number variation (CNV) to overcome the formaldehyde-induced DNA-protein cross linking (DPC).

The results are certainly of interest, and worth publication in Nature Communications. However, there are many questions to be clarified before acceptance in relation to the phenotypic mechanism on how the gene level mutation caused/contributed to the increase in the cell growth by regulating the central carbon metabolism.

1. In Figure 6a, the methanol consumption characteristics should also be shown to see the effect of methanol consumption rate on the cell growth.

We included the methanol consumption curve in Supplementary figure 3b. The methanol consumption is less in SM7 compared to SM1, implying SM7 released less formaldehyde to extracellular environment during the growth. This effect also resulted in a higher biomass yield from methanol for SM7.

2. From the result of Fig.6c, the roles of Tal, Tkt, and Rpi may not be critical/important for the consumption of H6P, or F6p via Phi ?

Agreed. Since Tal, Tkt and Rpi are reversible enzymes, they may not be as important as the irreversible Hps enzyme.

3. Although it is understandable that the increased activities of Hps and Phi contributed to the increased cell growth of SM6 by the higher consumption of methanol, it is not clear if these enzymes are under enzyme level regulation in response to the formaldehyde concentration.

Hps and Phi showed increased activity in OD3 (stationary phase) compared to OD1, most

likely due to complex effect resulting from global regulations, copy number changes of the IS-mediated tandem repeat region, and ribosome changes.

Enzyme level regulation is much more difficult to evolve, and we did not observe any changes in hps and phi sequences.

4. It is stated that SM6 changed its preferred substrate from sugars to methanol in relation to the result of Fig.8b. The substrate consumption characteristics should also be shown in Fig.8b. Note that the result of Fig.8b does not guarantee that the methanol is the preferred carbon source among other carbon sources. Since PTS may not have been mutated for the evolutional experiments using methanol, the catabolite repression by sugar such as glucose still exists, and the glucose may be the preferred carbon source even in the presence of methanol? This can be confirmed by the experiment of using a mixture of glucose and methanol.

Thank you very much for the great question. We tested and found out there is no diauxic growth as the SM7 (renamed from SM6) strain consumes glucose and methanol simultaneously. Thus, it would not be appropriate to state methanol being the "preferred substrate". We revised the text in line 335-342

Quote: Interestingly, SM7 (and SM8) grows faster in methanol than in glucose minimal media (Fig. 8e). However, SM7 did not exhibit diauxic growth in a medium containing both glucose and methanol (Fig 8e & Supplementary Fig. 6). The lack of diauxic growth control between glucose and methanol was expected, as such a mechanism must be evolved or designed specifically. In the presence of glucose in methanol medium, SM7 consumes glucose and methanol simultaneously for growth demonstrating a better growth rate with a doubling time of 2.9 hr (Fig. 8f). This result suggest that the strain can be used in a mixed substrate fermentation for optimal performance.

5. In Fig.8c, the specific production rates should be compared, instead of concentration.

We appreciate the reviewer's suggestion, but still opt to use the titer of the fermentation products. At this stage of characterization, fermentation products are used only as a rough idea of how the strain performs. We do not aim to characterize its product formation rate, which is not the purpose of this work. Additionally, as the growth rate differs significantly between SM1 and SM7, the specific production rates would not be

too meaningful.

6. I wonder why formate was produced by WT and SM6 under aerobic condition in Fig.8c.

Formate production by wild-type E. coli has been observed in glucose-containing media. This occurs primarily due to oxygen limitation, which can arise from insufficient mixing or as cultures enter the stationary phase. Under these conditions, E. coli switches to anaerobic metabolism, producing fermentation products including formate through pyruvate-formate lyase. Additionally, high concentrations glucose can lead to overflow metabolism, rapidly consuming oxygen and causing subsequent oxygen depletion, which also results in formate production.

Reference: High cell density media for Escherichia coli are generally designed for aerobic cultivations – consequences for large-scale bioprocesses and shake flask cultures Jaakko Soini, Kaisa Ukkonen and Peter Neubauer

7. The protein expression level as shown in Fig. 7c does not reflect the metabolic fluxes in general, and thus the discussion on the result must be careful. Metabolic flux analysis should be considered.

We do agree that proteomics may not totally reflect metabolic fluxes, and indicated in line 278 that “further flux measurements are required to confirm”.

8. It is not clear why the cell growth on glucose became lower after evolutional mutation (Fig.8b). Is this due to the modulation (up-regulation) of Cra, which represses the glycolytic pathway genes, while activating the gluconeogenic pathway genes.

It is indeed possible that cra changed the glucose metabolism. Also, the pfkA KO and gapA KO engineered back in SM1 mostly likely also contributed to the slow growth.

9. In relation to Cra regulation, the intracellular metabolite of FBP should be measured and compared?

Since we found out that restoring cra did not affect methanol growth, we did not focus on further characterization regarding FBP concentration.

10. The NADH balance should be checked based on the flux analysis.

Flux analysis would be beneficial to understand our system, but is out of the scope for this paper. Under aerobic conditions, NADH balance cannot be analyzed without detailed CO_2 and O_2 measurements.

11. I wonder why NADH produced by electron-rich substrate/methanol cannot contribute to ATP production through the respiratory chain under aerobic condition, but by bypassing the PDH pathway with Pox pathway and acetate consumption pathway (Acs) ?

E. coli has been evolved in nature in sugar-rich environment, thus its metabolism is optimized for sugar metabolism. *E. coli* has never been exposed to methanol, which is richer in electron (and energy). It is possible that methanol utilizing E. coli has to find a way to dump excess ATP or NADH. Thus, it chose to use a less efficient way (PoxB and Acs) to do so, and reduce the activity in TCA cycle. The former is supported by the bioreactor product formation under highly aerated condition. It still produces acetate, but is consumed later. The latter is supported by fixing the IS-insertion in gltA promoter and showed that the reduced gltA expression actually contributed to the growth of E. coli in methanol.

We added the following text in line 291:

GltA is the first enzyme in the TCA cycle, which produces NADH. As methanol is a more reduced substrate than glucose, a strong TCA cycle would increase the burden of respiration. In the SM7 strain the gltA promoter was interrupted by IS2, but it did not completely abolish the gltA expression. Proteome data suggested that GltA protein was still produced, albeit at about 10% level compared to the wild-type strain (Fig. 7c). Thus, it is possible that by properly knocking down gltA, the cell could balance NADH flux better. The same promoter mutation actually existed in the SM1 strain, however, re-introducing a wild-type gltA copy did not affect SM1 growth significantly5.

12. In the ribosome synthesis from amino acids, nitrogen regulation may also play the important role, and therefore, please discuss about the nitrogen regulation via glutamate and glutamine.

We agree with the reviewer that nitrogen regulation may also be important in SM strains. However, according to our additional gene enrichment analysis from proteomics, there is currently no evidence that nitrogen regulation is significantly altered in SM strains grown in methanol.

13. Based on the metabolic map, methanol is consumed to produce H6p and then to F6p, and then going down to the glycolytic pathway, and not the gluconeogenic carbon source, while the gluconeogenic pathway such as glyoxylate pathway etc. are activated instead of TCA cycle via GltA etc. This must be discussed in view of energy generation (catabolism) and biosynthesis (anabolism).

As mentioned in the response to Comment 11, the evolved SM6, 7, and 8 strains chose to use a less efficient way to dump excess ATP and electrons when grown in the substrate richer in electron than glucose. As shown in Fig. 8d, the strain first produced acetate, even though under highly aerobic conditions, and consumes acetate almost completely after 40 hrs. The glyoxylate pathway is essential in acetate utilization. It was selected during evolution such that the over efficiency for methanol utilization is higher.

We added the following in Discussion:

Lines 440-455: From the proteomic analysis, we found that AceEF, PflB, GltA decreased while PoxB, Acs and GlcB increased. GltA expression decrease was caused by an IS insertion in its promoter region that reduced its expression. The PoxB route is actually less energy efficient than the AceEF route, as it produces acetate as a carbon drain, which requires more ATP for reuptake. As methanol is richer in electron (and energy) than glucose, *E. coli* which typically grows in sugar-abundant environment has to find a way to deal with surplus electrons and energy. Thus, the PoxB route and the reduced TCA cycle was evolved to dump excess carbon to acetate. Fermentation in bioreactor also demonstrated that the cells produced acetate and consumed it later. The glyoxylate shunt enzymes were essential for acetate consumption, and their expression levels were increased for acetate reuptake.

Response to reviewer's comments

Reviewer #1 (Remarks to the Author):

In accordance with reviewers' comments, the authors have successfully addressed the comments and revised manuscript well. I'd like to recommend that the revised manuscript can be accepted as is for publication.

Thank you for the positive comments.

Reviewer #2 (Remarks to the Author):

Overall this work got significantly improved with additional experiments and rewriting of the work. I applaud the authors for creating a stable strain now! Als the writing of the manuscript substantially improved. Still I have some minor questions I would ask the authors to tackle:

1. It's good title is changed but I would change it a bit further to 'Evolutionary engineering' rather than 'Evolution engineering', I think that is grammatically correct. And I think 'growth with methanol' can better be replaced by 'growth on methanol'.

Thank you for your suggestion. We made the changes accordingly.

2. Even though several reviewers asked to compare/clearly refer to the rent paper with similar results recently published in nature Catalysis it's only briefly mentioned when discussing the mutations, I suggest to also mention it in the introduction/discussion.

We added the following text in the introduction and discussion sections:

Introduction, Line 38-41: "However, current synthetic methylotrophic strains evolved from non-C1-utilizing microbes still grow much slower than typical industrial strains[5,7,9,14], except a recent publication which appeared after submission of this work[13]."

Discussion, Line 457-468:

"The final stabilized SM8 strain grows with a doubling time of around 3.5 hr under optimal conditions, and exhibits no lag phase in re-growth. This growth rate is faster than that of native model methylotrophs, including *Methylorubrum extorquens* AM1 (Td ~ 4 hr) and *Bacillus methanolicus* at 37°C (Td ~ 5 hr), as well as the synthetic methylotroph (Td ~ 4.3 hr) reported by *Reiter* et al[13]. The maximal biomass that SM8 achieved is 19.9 g/L, which is comparable to the previously reported strain[13]. The SM8 strain contains about 50 mutations compared to its parental strain SM1 which already had 21 mutations. The strain reported by *Reiter* et al [8,13] contains more than 1000 mutations, which may be due to a *dnaQ* mutation that caused hypermutation. We observed that hypermutation led to a decrease in the growth rate of SM7 after successive passages. By restoring the wild-type *mutS* gene in SM7 to create SM8, SM8 maintained a doubling time of around 3.5 hours."

3. Figure 8, pane e and f sow in the red lines the same strain and same conditions (SM7 on 400 mM methanol MOPS medium but show very different maximum OOD (1.2 and 3). So that is inconsistent how can this be explained?

The low OD (OD 1.2) resulted from incubation of the testing strains in cell culture tubes, and was measured using the GENESYS™ 30 Visible Spectrophotometer which saturates at around OD 1.2. Additionally, we did not appropriately dilute the samples measured in the previous Figure 8e, leading to detection saturation. We have deleted this figure. A high OD (OD 3.0) is achieved through cultivation in the 50 mL Falcon tube system (current Fig. 8e)

And also the glucose data look different. Also panel E does basically have to little sampling points to show proper growth curve I would not show this her like this .

The glucose data appear different because previous version of Figure 8e was done with 1% glucose (the description in the legend incorrectly stated 0.1% glucose). 1% glucose is commonly used to support the fast growth of wild-type *E. coli*, while 0.1% glucose is generally used for testing diauxic growth. To avoid confusion and provide a more complete result, we demonstrated growth on 1% glucose in a new Figure 8e and revised the text as follows:

Line 337-340: “Interestingly, SM7 grows faster in methanol than in 1% glucose minimal medium, a concentration that is sufficient to support fast growth of wild-type E. coli in the MOPS medium (Fig. 8e). However, SM7 did not exhibit diauxic growth in a medium containing both 0.1% glucose and 400 mM methanol (Fig. 8e & Supplementary Fig. 7).”

The method for media preparation was also revised:

Line 549-550: “The MOPS-based glucose minimal medium was composed of MOPS EZ buffer, 50 mg/mL chloramphenicol, trace vitamin mix, 1 mM IPTG, and either 1% or 0.1% glucose.”

4. L506, point missing at the end of the sentence

Fixed, thank you

5. L544 ‘air flow’ should be ‘gas flow’ if it’s 100% O2

Fixed, thank you

6. For the fed-batch experiment in Figure 8 in the bioreactor and falcon tubes (nice that these experiments are added) I miss the total amount of methanol added, this would be good to report, especially for the closed Falcon tube system this could give additional data to calculate late yield.

For the bioreactor system, the amount of methanol actually added is difficult to determine, because of significant evaporation.

For the Falcone tube culture, we measured the dry weight of the biomass and the methanol consumed in the Falcone tube culture, and obtained the biomass yield of 0.39 g biomass/g methanol. (Line 188-190)

7. Figure 7 caption “ 44 E. coli ribosomes” do you mean ribosomal genes?

It should be 44 ribosomal subunit proteins. We revised it in the legend of Figure 7a.

8. Figure 7b I still not find clear enough, what do you mean by gene ratio on the x-axis, define this clear in caption

To make it clear, we revised the legend of Figure 7b as follows:

"The "gene ratio" on the x-axis represents the proportion of genes from a given gene set that are found to be differentially expressed or enriched in the analyzed dataset. A higher gene ratio indicates that a larger proportion of the gene set is involved in the indicated biological pathway."

9. In reply to my previous comment you reported yields, but these cannot be true, 38 g/g methanol is unrealistically high yield. I think you made a calculation error. In figure S3 you show an OD increase of about 3 (*0.5 gCDW/OD = 1.5gCDW) for a consumption of about 0.1 M methanol. (0.1 M * 24 g/mol = 2.4 g methanol). So yield is something like 1.5/2.4 = 0.6 gCDW/g methanol (or 1.5/0.1 = 15 gCDW/mole methanol). This is actually close to the theoretical yield of the RuMP (see e.g. Cotton et al., Current opinion).

The calculation you showed was almost correct, except methanol MW is 32 g/mol. Using the correct MW, the yield should be 0.388 g biomass/g methanol.

We made a mistake in typing the yield unit (it should be % rather than g/g). The correct value should be 0.388 (38.8%) g biomass/g methanol.

So I suggest to highlight and also show a correct yield estimation in the manuscript text (with the remark its it not real CDW measured but estimated from OD600)

We actually measured the real CDW, not estimated. The yield number is now reported in Line 188-190: We measured the dry weight of the biomass and the methanol consumed in the Falcone tube culture, and obtained the biomass yield of 0.39 g biomass/g methanol.

10. Title at start page supplementary files does not match with manuscript title.

Thank you for the reminder. We have put the correct title on the cover of supplementary files.

11. Good that plasmids maps are included now, but for reproducibility also plasmid sequences should be included, I did not spot them in supplementary.

Following your suggestion, we added the plasmid sequences in revised supplementary.

Reviewer #3 (Remarks to the Author):

The manuscript has been appropriately modified, and improved, and therefore, the MS can be now accepted.

Thank you for the positive comments.

Response to reviewer's comments

Reviewer #2 (Remarks to the Author):

I compliment the authors on their revised manuscript and all my comments have been properly addresses. There is one small typo that I spotted in revised sentence in a L189: 'Falcone' should be spelled 'Falcon'

Thank you for the positive comments, and the typo was revised.