Mechanism of PQQ and CoQ10 overproduction in Methylobacterium as revealed by comparative genomic analysis between mutant and wild-type strains

Changle Zhao  
Shihezi University

Yinping Wan  
Shihezi University

Xiaojie Cao  
Shihezi University

Huili Zhang (✉️ zangchch02@126.com)  
Shihezi University  https://orcid.org/0000-0002-2772-3618

Xin Bao  
Xinjiang Fufeng Biotechnology Co., Ltd.

Research

Keywords: pyrroloquinoline quinone, carotenoid, coenzyme Q10, Methylobacterium, comparative genomic analysis

DOI: https://doi.org/10.21203/rs.3.rs-36145/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

The microbial synthesis of pyrroloquinoline quinone (PQQ) and Coenzyme Q10 (CoQ10) remains the most promising industrial production route. *Methylobacterium* has been used to generate PQQ and other value-added chemicals from cheap carbon feedstocks. However, the low PQQ and CoQ10 production capacity of the *Methylobacterium* strains is a major limitation. The regulation mechanism for PQQ and CoQ10 biosynthesis in this strain has also not been fully elucidated.

Results

*Methylobacterium* sp. CLZ strain was isolated from soil contaminated with chemical wastewater, which can simultaneously produce PQQ, CoQ10, and carotenoids by using cheap methanol as carbon source. We investigated a mutant strain NI91, which increased the PQQ and CoQ10 yield by 72.44% and 59.80%, respectively. Whole-genome sequencing of NI91 and wild-type strain CLZ revealed that both contain a 5.28 Mb chromosome. The comparative genomic analysis and validation study revealed that a significant increase in biomass and PQQ production was associated with the base mutations in the methanol dehydrogenase (MDH) synthesis genes, *mxaD* and *mxaJ*. The significant increase in CoQ10 production may be associated with the base mutations in *dxs* gene, a key gene in the MEP/DOXP pathway.

Conclusions

A PQQ producing strain that simultaneously produces CoQ10 and carotenoids was selected and after ANI analysis, named as *Methylobacterium* sp. CLZ. After random mutagenesis of this strain, we obtained NI91 strain, which showed increased production of PQQ and CoQ10. Based on comparative genomic analysis of the whole genome of mutant strain NI91 and wild-type strain CLZ, a total of 270 SNPs and InDels events were detected, which provided a reference for subsequent research. The mutations in *mxaD*, *mxaJ* and *dxs* genes may be related to the high yield of PQQ and CoQ10. These findings will enhance our understanding of the PQQ and CoQ10 over-production mechanism in *Methylobacterium* sp. NI91 at the genomic level. It will also provide useful clues for strain engineering in order to improve the PQQ and CoQ10 production.

Background

*Methylobacterium* derives energy by the degradation of toxic substances such as formaldehyde and methanol. Besides, it converts methanol and formaldehyde into protein, Coenzyme Q, vitamin B12, PHB, various amino acids, carotenoids, PQQ, and other secondary metabolites (1). Methanol utilizing bacteria were first discovered in the 1900s (2).
PQQ is an oxidoreductase Coenzyme that was discovered after pyridine nucleotides (NAD+, NADP+) and riboflavin (FMN, FAD) (3). The functional role of PQQ includes anti-oxidation, scavenging free radicals, and thus it has a crucial role in food, medicine, and health (4–6). CoQ10 plays a vital role in the electron transfer chain, and it is involved in multiple biological processes such as anti-oxidation, immunity enhancement, and scavenging of free radicals (7). Although preliminary information on the biosynthetic pathways of PQQ (8) and CoQ10 (9, 10) is available, the global genetic basis for PQQ and CoQ10 biosynthesis has not been elucidated so far.

Comparative genomics analysis based on the whole genome sequence has been employed to investigate genome variations between wild-type strains in both natural and artificial environments (11) and comparative genomic analysis could unravel the underlying mechanism of metabolite overproduction. A comparative genomic and transcriptomic analyses of a high-producing strain E3 with the wild-type strains NRRL23338 might detect the mutations, and the changes in transcription and translation machinery that could be responsible for the erythromycin overproduction (12). Zeng WZ et al. performed a similar comparative genomic analysis of a series of *Yarrowia lipolytica* WSH-Z06 mutants. The outcome of this study indicated that different α-ketoglutarate production capacity of these mutants might be associated with different mechanisms (13). Even though comparative genomic analysis provides new insight into the mechanism of secondary metabolite overproduction, these hypotheses demand further validation through functional experiments.

Since *Methylobacterium* was discovered for a short time, its complex metabolic pathways have not been fully explored yet. Additionally, the whole genome sequences of only a handful of PQQ synthesizing *Methylobacterium* strains are available. Thus, it is challenging to study genomics and metabolic pathways of this bacterium. In this study, we sequenced the whole genome of PQQ and CoQ10 overproducing strains NI91, and wild-type strain CLZ. Comparative genomic analysis of mutant strains with different PQQ and CoQ10 production capacity may decipher the associated molecular mechanisms (14), and elucidate the mutation-sites associated with the PQQ and CoQ10 overproduction. We also determined the expression of stably expressed *Methylobacterium* gene *gapA*, which was used as an internal reference gene, and served as a reference for the determination of *Methylobacterium* gene expression.

**Results**

**Genomic features of wild type strain CLZ**

We acquired 3,887,833,448 bp raw reads by using PromethION, and 1,582,330,800 bp raw reads by using MGISEQ-2000 from wild-type strain CLZ for genome-assisted correction. The main features of the genome sequence of wild-type strain CLZ are schematically represented in Fig. 1. The assembled genome consists of 5,409,250 bp long contig. The CLZ genome sequence was checked to confirm if it includes a plasmid sequence. However, the analysis revealed that it did not contain the plasmid sequence. The GC content of the CLZ genome was found to be 68.5%. Besides, the wild-type strain CLZ was found to
contain a total of 5025 predicted coding sequences, including 4949 protein-coding sequences (CDSs), 50 tRNA genes, 15 rRNA genes, and five copies of 16S-23S-5S rRNA operons.

Among the 4949 predicted protein-coding sequences, 2769, 2136, 2495, 4851, 3774, 1287 proteins were annotated by the COG, KEGG, GO, RefSeq, Pfam, and TIGRFAMs databases, respectively. As per the statistical analysis, a total of 4854 protein sequences, which accounts for 98.08% of the total sequences, were annotated in one of the above 6 databases.

**16S rDNA sequence and ANI analyses for the characterization of isolated bacterial strain**

The 16S rDNA gene is a widely used and most reliable genetic marker for the study of bacterial taxonomy. Figure 2 depicts a phylogenetic tree constructed from the 16S rDNA of the CLZ strain. CLZ and *Methylorubrum aminovorans* JCM 8240<sup>T</sup> were clustered on the same branch. The bootstrap value was found to be 99%, which indicated that CLZ is closely related to *Methylorubrum aminovorans* JCM 8240<sup>T</sup>. The 16S rDNA led identification of species has certain limitations. Thus, we performed ANI calculations for the entire genome sequence of CLZ. The Average Nucleotide Identity (ANI) (15) based on entire genomes provides another appropriate gauge for bacterial species delineation. Table 1 shows the result of ANI analysis. The highest ANI between CLZ and *Methylorubrum populi* BJ001 was 88.608%. Since the ANI of CLZ and all *Methylobacterium* were far below the threshold of 95%(16), we decided to name CLZ as *Methylobacterium* sp. CLZ.
Table 1
The ANI results.

| ANI      | Coverage | Bacterial strain               | Genome Assembly          |
|----------|----------|--------------------------------|--------------------------|
| 88.61%   | 75.4697% | *Methylorubrum populi* BJ001   | GCA_000019945.1          |
| 88.12%   | 75.5661% | *Methylorubrum extorquens* CM4 | GCA_000021845.1          |
| 88.11%   | 74.4657% | *Methylorubrum extorquens* AM1 | GCA_000083545.1          |
| 88.04%   | 51.9527% | *Methylorubrum extorquens*     | GCA_900234795.1          |
| 84.62%   | 32.8288% | *Methylorubrum salsuginis*     | GCA_900114375.1          |
| 83.55%   | 29.7255% | *Methylobacterium dankookense* | GCA_902141855.1          |
| 82.90%   | 26.5309% | *Methylobacterium oxalidis*    | GCA_007992195.1          |
| 82.43%   | 25.8210% | *Methylobacterium haplocladii* | GCA_007992175.1          |
| 82.47%   | 25.8210% | *Methylobacterium soli*        | GCA_008806385.1          |
| 82.71%   | 25.7201% | *Methylobacterium radiotolerans* JCM 2831 | GCA_000019725.1          |

ANI values were sorted from highest to lowest, and the top ten ANI values were listed in this table. *Methylorubrum* and *Methylobacterium* are used alternatively for the same species.

The comparative genomic analysis identified mutations in the mutant strain NI91

We acquired 3,463,849,211 bp raw reads by using PromethION, and 1,208,600,700 bp raw reads by using MGISEQ-2000 from genome sequencing of the NI91 strain for genome-assisted correction. The genome of NI91 consisted of one contig with a size of 5,409,262 bp and 68.5% GC content. As per the collinearity analysis of the CLZ and NI91 strains, the size and gene sequence of the two genomes were highly conserved without deletion, duplication, inversion, and translocation of the large fragment (Fig. 3).

*Methylobacterium* sp. NI91 is a mutant strain obtained by random mutagenesis of *Methylobacterium* sp. CLZ. Thus, SNPs and InDels analyses of its genome are highly significant in exploring the mechanism of secondary metabolite synthesis. We found a total of 270 SNPs and InDels events in the genome of NI91, a mutant strain of *Methylobacterium* sp. as compared to the CLZ, a wild-type strain *Methylobacterium* sp. Out of the 270 mutations, 182 mutations occurred in the exon region, of which 102 were missense...
mutations, 20 were frameshift mutations, 2 terminators were obtained, and 58 were synonymous mutations.

Furthermore, we annotated a total of 24 genes/proteins with missense mutations, frameshift mutations, and terminator occurrences by using the COG database. 103 proteins were annotated, as shown in Fig. 4. The top four groups (each group > 10 genes) with predicted functions included the gene involved in replication, recombination and repair (L), inorganic ion transport and metabolism (P), and signal transduction mechanisms (T). Additionally, in the NI91 strain, two terminators were present, which resulted in the deletion of the ABC transporter ATP-binding protein and porin subfamily, which are involved in the cell's transmembrane transport. It is worth noting that these mutations do not affect the metabolism and compound synthesis, but it affects the material transport and regulation, which in turn alters the physiological outcomes.

Dynamic changes between wild-type strain CLZ and mutant strain NI91

The continuous fermentation in a 3-liter fermenter with wild-type strain CLZ for 144 h resulted in 330.14 mg/L (DCW), 11.21 mg/L (PQQ yield), 10.92 mg/g (CoQ10 yield) (Fig. 5a). In the case of NI91 mutant strain, it was the 349.70 mg/L (DCW), 19.33 mg/L (PQQ yield), 17.45 mg/g (CoQ10 yield) of mutant strain NI91 (Fig. 5b). The yields of DCW, PQQ, and CoQ10 of mutant strain NI91 were increased by 5.92%, 72.44%, and 59.80%, respectively, as compared to the wild-type strain CLZ. The yield of carotenoids in mutant strain NI91 and wild-type strain CLZ was 3 mg/g with no significant change. We observed that the biomass of the medium did not increase, but the PQQ and CoQ10 production still increased, which might be due to mutations in the genes involved in PQQ and CoQ10 regulation.

Related genes between wild-type strain CLZ and mutant strain NI91

We investigated the candidate genes involved in the PQQ and CoQ10 synthesis by amplifying 600–900 bp regions adjacent to mutation sites by PCR. The gene fragments were sequenced by the chain termination method. The validation outcomes indicated that no base mutation was present at 1824306 bp, which was 79 bp upstream of the mxaJ gene of the NI91 genome as compared to the CLZ genome, but the G base at 1824372 bp was deleted, which was 13 bp upstream of the mxaJ gene. The base mutations in the NI91 genome as compared to the CLZ genome were validated.

In the CLZ strain, the protein sequences encoded by the two DNA sequences of 4624834 bp-4625283 bp (dxs1) and 4625223 bp-4626809 bp (dxs2) are annotated as DXS. The dxs gene of the NI91 strain has 3 mutations as compared to CLZ strain (Table 2). Since the C base was inserted at 4625176 bp, the two dxs genes in the NI91 strain were translated into a complete DXS (Fig. 6). Additionally, there were mutations at 4625957 bp and 4626013 bp, but the mutation at 4625957 bp did not change the amino acid sequence of DXS. The three-dimensional structure of DXS was altered greatly
due to these three mutations in the \textit{dxs} gene (Fig. 6). It affected the DXS activity to a significant extent. Primers were designed as per the \textit{dxs}^2 gene sequence of the CLZ strain without a frameshift mutation. The expression level of the \textit{dxs} gene in the CLZ and NI91 strain in the lag phase was found to be similar, but its expression in the logarithmic phase and the stable phase of the NI91 strain was comparatively less as compared to the CLZ strain (Fig. 8).

Table 2 demonstrates the mutations in the \textit{mxaD} and \textit{mxaJ} genes of the NI91 strain. The \textit{mxaD} gene causes the $\beta$-sheet of the translated protein to be extended (Fig. 7). However, a deleted base at -13 bp upstream of the \textit{mxaJ} gene led to the higher expression of the \textit{mxaJ} gene in the mutant strain NI91 as compared to the wild-type strain CLZ, in the lag phase, log phase, and stable phase (Fig. 8).

| Gene | Mutation site (bp) | Allelic base | Allelic protein | Enzyme |
|------|-------------------|--------------|----------------|--------|
| \textit{mxaD} | 4434961 | G391A | G131S | MDH |
| \textit{mxaJ} | 1824372 | C-13* | Upstream | MDH |
| \textit{dxs} | 4625176 | *344C | Ser116Frameshift | DXS |
| \textit{dxs} | 4625957 | C735T | Thr245Thr | DXS |
| \textit{dxs} | 4626013 | C791T | Thr264Ile | DXS |

**Discussion**

An exhaustive metabolic investigation of \textit{Methylobacterium} will be highly significant as this bacterium can utilize one-carbon compounds that are inexhaustible in nature as a carbon source for energy production. Research investigations conducted over the last few decades have shown that \textit{Methylobacterium} has a distinguished metabolic system; thus, it can be used as a metabolic model to understand biological metabolism and evolution (17). As \textit{Methylobacterium} is recently discovered, limited metabolic and genomic data is available on this bacterium. Thus, in this study the wild-type strains and the mutant strains with significantly increased yields of PQQ and CoQ10 that were randomly induced were subjected to whole-genome sequencing. Comparative genomics analysis of the mutant strain and the wild-type strain revealed five mutation sites that may be related to the significant increase in the production of PQQ and CoQ10, which provides a reference for the molecular breeding of PQQ and CoQ10 producing bacteria.

In this study, a pink strain was screened from the soil at the sewage outlet of a chemical plant. The wild-type strain CLZ was mutated for 11 generations using UV, NTG, EMS, and UV-LiCl, and a mutant strain NI91 was obtained. The wild-type, CLZ, and mutant strain NI91 were sequenced using PromethION and
MGISEQ-2000. The wild-type, CLZ, and mutant strain, NI91, genome sizes were found to be 5,409,250 bp and 5,409,262 bp, respectively, with one contig.

Till now, comprehensive investigations on the *Methylobacterium* gene expression level have not been attempted and, so it is crucial to select an internal reference gene that can be stably expressed in this bacterium. Six commonly used internal reference genes selected for qRT-PCR: 16S rRNA (16S ribosomal RNA), GAPDH (Glyceraldehyde-3-phosphate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase), recA (Recombinase A, gene encoding recA protein involved in DNA repair), DnaN (DNA-directed DNA polymerase activity), rpoB (DNA-directed RNA polymerase subunit beta, a coding gene for RNA polymerase beta), and proteinS5 (30S ribosomal protein). Premier 6 was used to design primers for sequencing (Supplemental Table S2). As depicted in Fig. 5a, the CT values of the *gapA*, *RecA*, and *rpoB* genes were between 27–32, and their expression levels were relatively high and stable. The CT value of the gene 16S gene was between 13–18, and its expression level was high, and stability was poor; thus, it was not used as a reference gene for *Methylobacterium*. The CT values of *dnaN* gene and protein S5 were between 31–40 with low expression levels and poor stability; thus, it was not used as an internal reference gene for *Methylobacterium*. Furthermore, we analyzed the CT values of the six candidate internal reference genes, at three stages of growth for the CLZ and NI91 strain by using the NormFinder software. The stability value was sorted as *gapA* (0.037) < *rpoB* (0.044) < protein S5 (0.062) < *RecA* (0.068) < *dnaN* (0.070) < 16S (0.123). Lower the stability value, more stable was the reference gene expression. Therefore, the gene expression level and expression stability indicated that *gapA* gene was most suitable as a reference gene for *Methylobacterium*.

In prokaryotic cells, isopentenyl diphosphate (IPP) is produced by the non-mevalonate pathway (MEP/DOXP), which is a common precursor of the CoQ10 side-chain and terpenoids such as carotenoids (18–20). The *dxs* gene, which encodes 1-deoxy-D-xylulose-5-phosphate synthase (DXS) in the MEP/DOXP pathway, is the first gene in the MEP / DOXP pathway. The *dxs* gene of the NI91 strain has 3 mutations as compared to CLZ strain. The three-dimensional structure of DXS was greatly altered due to the frameshift mutations in the *dxs* gene, it might affect the DXS activity to a significant extent. Besides, the over-expression of the *dxs* gene might facilitate the side chain synthesis of CoQ10 and carotenoids.

The methanol dehydrogenase in methylotrophic bacteria is MXAF-type MDH, which is a quinone protein with PQQ as a prosthetic group (21). It serves as the vital enzyme for the methylotrophic bacteria, which facilitates methanol utilization to generate energy for bacterial growth (22). It consists of 5 gene clusters (*mxa*, *mxb*, *pqqABCDE*, *pqqFG*, and *mxc*). In the MXAF-type methyl dehydrogenase, the methanol oxidation system participates in the catalysis and regulation, and the gene cluster *mxa* is responsible for the synthesis, assembly, and stability of MDH (23, 24). The up-regulation of the methanol dehydrogenase gene enhances PQQ production (25). Thus, the PQQ production of mutant strain NI91 was found to be higher than the wild-type strain CLZ, which can be attributed to the base mutation upstream of the *mxaJ* gene. The *mxaD* gene causes the β-sheet of the translated protein to be extended. The *mxaD* gene mutation alters the catalytic activity of the corresponding enzyme, which affects the rate of MDH assembly as well as synthesis.
Conclusion

We screened a pink strain from the soil at the sewage outlet of a chemical plant. 16S rDNA and ANI analysis were performed, and this strain was characterized as *Methylobacterium* sp. CLZ. The wild-type strain CLZ was mutated for 11 generations using UV, NTG, EMS, and UV-LiCl, and a mutant strain NI91 was obtained, which showed 5.92%, 72.44%, 59.80% increased yield of DCW, PQQ, and, CoQ10 yields, respectively. The wild-type, CLZ, and mutant strain NI91, were sequenced using PromethION and MGISEQ-2000 and their genome sizes were found to be 5,409,250 bp and 5,409,262 bp, respectively, with one contig. As compared to the wild-type strain of *Methylobacterium* sp. CLZ, 270 SNPs, and InDels events were detected in the genome of the NI91, mutant strain of *Methylobacterium* sp. Out of the 270 mutations, 182 occurred in the exon region resulting in 102 missense mutations, 20 frameshift mutations, 2 terminators, and 58 synonymous mutations. The deletion of upstream bases of the *mxaJ* gene and the missense mutation of the *mxaD* gene can be associated with an increase of PQQ production. The code shift mutation of *DXS* gene can be associated with an increase in CoQ10 production. The outcome of this study will serve as a reference for subsequent molecular breeding. Also, based on the yield variation of the target product, the comparative genomic analysis of mutant and wild-type strain will be useful for navigating through the distinct mutation sites for molecular breeding.

Materials And Methods

Bacterial Strains

The wild-type strain: We isolated a pink strain from the soil of sewage outlets of six chemical plants located at Shandong, China. The isolated strain was identified as a Gram-negative bacillus. It utilizes methanol as a carbon source and simultaneously produces PQQ, CoQ10, and carotenoids. The strain was named as CLZ (GenBank accession no: CP047608).

The Mutant strain: The wild-type strain CLZ was mutated by UV, NTG, EMS, and LiCl-UV for 11 consecutive generations to obtain a mutant strain with significantly improved biomass and enhanced capacity of PQQ and CoQ10 production. The strain was named as NI91 (GenBank accession no: CP047607).

Bacterial culture

Seed medium: Methanol (10 g/L), yeast powder (0.5 g/L), MgSO$_4$$\cdot$7H$_2$O (1.5 g/L), (NH$_4$)$_2$SO$_4$ (3 g/L), KH$_2$PO$_4$ (1.4 g/L), Na$_2$HPO$_4$ (3 g/L), pH 6.8.

Fermentation medium: Yeast powder (1 g/L), methanol (30.0 g/L), MgSO$_4$$\cdot$7H$_2$O (1.5 g/L), (NH$_4$)$_2$SO$_4$ (3 g/L), KH$_2$PO$_4$ (1.4 g/L), Na$_2$HPO$_4$ (3 g/L), iron citrate (30 mg/L), CaCl$_2$$\cdot$2H$_2$O (30 mg/L), MnCl$_2$$\cdot$4H$_2$O (0.5 mg/L), ZnSO$_4$$\cdot$7H$_2$O (5 mg/L), CuSO$_4$$\cdot$5H$_2$O (0.5 mg/L), CoCl$_2$$\cdot$6H$_2$O (3.2 mg/L), pH 7.
A single colony was inoculated into 100 mL of seed medium in a 500 mL flask. It was incubated for 36–40 h in a rotary shaker at 200 RPM and 30 °C till the OD$_{600}$ reached 1.5-3.0 (logarithmic phase). 100 mL of this seed culture was mixed with 1 L fermentation medium and transferred to a 3 L fermenter and incubated at 34 °C for 144 h at 120 L/h and 500 RPM.

**Genome extraction, sequencing, and assembly**

Genomic DNA was extracted using the QIAGEN® kit (Qiagen, Valencia, CA) as per the manufacturer's instructions. The indexed 1D genomic DNA libraries and DNA sequence were determined by PromethION (Oxford Nanopore Technology sequencer). Subsequently, we used the sequencing quality value (mean_qscore_template) ≥ 7 and sequence length of ≥ 1000 bp to eliminate low quality reads. The filtered reads were assembled into the genome using software Canu, and the parameters were set to default. Simultaneously, indexed paired-end genomic DNA libraries were sequenced on the MGISEQ-2000 platform (paired-end, 150 bp read length, MGI Tech Co., Ltd.). Furthermore, genome assembly results were corrected using Pilon software, and parameters were set to default. In the corrected genome, the origin of the nucleic acid sequence was moved to the replication initiation site of the genome using Circlator software, and thus the final genome sequence was obtained. The genome sequence was aligned with the plasmid database using BLASR, and the length of the aligned sequence was calculated. If the alignment length was more than 20% of the sequence length and less than 1 Mb size, the sequence was considered to be a plasmid sequence.

**16S rDNA and average nucleotide identity (ANI) analysis**

The 16S rDNA (GenBank: MN795902) sequences were fetched, and homology analysis of the 16S rDNA sequence of the CLZ strain was carried out by using the EzBioCloud database. The 16S rDNA sequence of the 9 model strains with the highest similarity and the 16S rDNA sequence of the CLZ strain were used for cluster analysis. The phylogenetic tree analysis was performed by using the Molecular Evolutionary Genetics Analysis (MEGA) software (26) version 6.06, based on the neighbor-joining method to identify the genera and strains.

Average nucleotide identity (ANI) test was used to compare the submitted genome sequence against the GenBank genome sequences of the strains. Whole-genome ANI analysis was performed using the Prokaryotic Genome Annotation Pipeline (PGAP) (27).

**Genome annotation and analysis**

CDS is a protein-coding sequence of the genome. Prodigal software was used to predict the coding genes and retain the complete CDS with the following parameters: p None -g 11. The main databases used for gene/protein annotation were the Cluster of Rfam (28), RefSeq (29), Pfam (30), TIGRFAMs (31), Gene Ontology (GO) (32), Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) (33), Clusters of Orthologous Groups of proteins (COG) (34).

**Genomic collinearity, SNPs, and InDels analyses**
The genomic sequences of the wild-type strain CLZ and the mutant strain NI91 were compared using the MUMmer (35) software version 3.23, and the analysis results were plotted using the Gnuplot (36) software version 5.2.

The MUMmer software was used to detect the SNPs and InDels loci using the genomic sequence of the wild-type strain CLZ as the reference genome, and the genomic sequence of the mutant strain NI91 as the target genome. The CLZ's genomic files and annotation files were used to build the index. The impact of SNP and InDel sites on the genes were analyzed using the ANNOVAR (37) software version 20191024. It detected the mutation position in the genes, along with synonymous replacement, nonsynonymous replacement, frameshift mutation, and so on.

**PCR and qRT-PCR**

PCR analysis was used to identify mutation sites in the genome. Primers were designed against the gene sections that were 500 bp upstream and downstream of mutation sites using the Primer software version 6.0 (Supplemental Table S1). The PCR products were sequenced using the chain termination method (Shanghai Biotech Bioengineering Co., Ltd.) to verify SNP and InDel loci.

qRT-PCR was used to measure the expression levels of genes of interest. Primers were designed using the Primer premier (38) software version 6.25 (Supplemental Table S2). The bacterial strains, CLZ and NI91, were cultured till they reached: lag phase (16 h), log phase (36 h), and stable phase (60 h). Total RNA was extracted using the RNA prep pure Cell/Bacteria Kit (TIANGEN, Beijing, China) as per the manufacturer's instructions. cDNA was synthesized using the Prime Script RT reagent Kit with gDNA Eraser (Vazyme, Nanjing, China). PCR was performed in a 20 µL volume employing LightCycler 480 SYBR Green I Master (Roche, Shanghai, China) following the manufacturer's instructions on a LightCycler 480 II (Roche, Basel, Switzerland).

**Analytical methods**

Bacterial cultures were diluted, and the OD$_{600}$ value was used to measure the biomass. The optical density of cultures was measured with a spectrophotometer at 600 nm after appropriate dilution and converted to dry cell weight (DCW) as per the predetermined calibration line: DCW (g·L$^{-1}$) = 0.0235 * OD$_{600}$ + 0.2128, R$^2$ = 0.9957.

For the extraction of PQQ, the fermentation broth was centrifuged at 16000 g for 5 min at 4 °C and supernatant was separated. PQQ concentration in the fermentation medium was analyzed as described previously (39).

The isolation of carotenoids was performed as per the method described previously (40) but with slight modifications. The fermentation broth was centrifuged to separate the supernatant, bacterial cells were washed twice with sterile water and dissolved in acetone-methanol solution (7/2, v/v), and sonicated for 45 min, and shaken at room temperature in the dark for 60 min. The cells were repeatedly extracted twice
to obtain the colorless cells, which were later mixed. The determination of carotenoids was performed as per the method described previously (41).

The isolation and determination of CoQ10 were performed as per the method described previously (42); however, the extraction method of CoQ10 was similar to that of the carotenoids.

Homology modeling was applied to predict three-dimensional structures of the proteins of interest using Phyre (43) version 2.0.

**Abbreviations**

PQQ  
pyrroloquinoline quinone  
CoQ10  
Coenzyme Q10  
CLZ  
*Methylobacterium* sp. CLZ  
NI91  
*Methylobacterium* sp. NI91  
COG  
Clusters of Orthologous Groups

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

All data generated or analysed during this study are included in this published article.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the National Natural Science Foundation of China (grant number 21366028) and China Postdoctoral Science Foundation (grant number 2017M613301XB).
Authors' contributions

HL Z and X B conceived and designed the research. CL Z and XJ C and YP W conducted the experiments. HL Z and CL Z contributed new reagents or analytical tools. CL Z analyzed the data. HL Z and CL Z wrote the manuscript. All the authors read and approved the manuscript.

Acknowledgements

Not applicable.

References

1. Reeve JN. 1983. The biochemistry of methylotrophs: by C. Anthony, Academic Press, 1982. £24.00/$49.50 (xv + 431 pages) ISBN 0 120 58820 X. Trends in Biotechnology 1:31–32.
2. Anthony C. The biochemistry of methylotrophs. London: Academic Press; 1982.
3. Robert S, Moritz M, Sandra P. The pyrroloquinoline quinone biosynthesis pathway revisited: A structural approach. BMC Biochem. 2008;9:1–11.
4. Felton LM, Anthony C. 2005. Biochemistry: Role of PQQ as a mammalian enzyme cofactor? 433:10–11.
5. Misra HS, Khairnar NP, Barik A, Priyadarsini Kl, Mohan H, Apte SK. Pyrroloquinoline-quinone: a reactive oxygen species scavenger in bacteria. Febs Letters. 2004;578:0–30.
6. Misra HS, Khairnar NP, Rajpurohit YS. Pyrroloquinoline-quinone and its versatile roles in biological processes. J Biosci. 2012;37:313–25.
7. Tian Y, Yue T, Yuan Y, Soma PK, Lo YM. Improvement of cultivation medium for enhanced production of coenzyme Q10 by photosynthetic Rhodospirillum rubrum. Biochem Eng J. 2010;51:160–6.
8. RoseFigura JM, Imsand EM, Bonnot F, Sjoelander K, Klinman JP, Shen Y-Q. Distribution and Properties of the Genes Encoding the Biosynthesis of the Bacterial Cofactor, Pyrroloquinoline Quinone. Biochemistry. 2012;51:2265–75.
9. Hajime Y, Yukinobu K, Keiko O, Kazumi A. 1998. Production of ubiquinone-10 using bacteria. The Journal of general and applied microbiology 44.
10. Young IG, Stroobant P, Macdonald CG, Gibson F. Pathway for ubiquinone biosynthesis in Escherichia coli K-12: gene-enzyme relationships and intermediates. J Bacteriol. 1973;114:42–52.
11. Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MAS. Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics. 2008;9:524–4.
12. Li YY, Chang X, Yu WB, Li H, Ye ZQ, Yu H, Liu BH, Zhang Y, Zhang SL, Ye BC, Li YX. Systems perspectives on erythromycin biosynthesis by comparative genomic and transcriptomic analyses of S. erythraea E3 and NRRL23338 strains. BMC Genom. 2013;14:523.
13. Zeng W, Fang F, Liu S, Du G, Chen J, Zhou J. Comparative genomics analysis of a series of Yarrowia lipolytica WSH-Z06 mutants with varied capacity for alpha-ketoglutarate production. J Biotechnol.
14. Serpova EV, Naumova ES, Naumov GI. Genome variability of the yeast Yarrowia lipolytica. Microbiology. 2010;79:229–36.

15. Ciufò S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol. 2018;68:2386–92.

16. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.

17. Chistoserdova L, Chen SW, Lapidus A, Lidstrom ME. Methylotrophy in Methylobacterium extorquens AM1 from a genomic point of view. J Bacteriol. 2003;185:2980–7.

18. Cluis CP, Pinel D, Martin VJ. The Production of Coenzyme Q10 in Microorganisms. Subcell Biochem. 2012;64:303–26.

19. Kawamukai M. Biosynthesis, Bioproduction and Novel Roles of Ubiquinone. J Biosci Bioeng. 2002;94:511–7.

20. Barkovich R, Liao JC. Metabolic engineering of isoprenoids. Metab Eng. 2001;3:27–39.

21. Zheng YJ, Xia Z-x, Chen Z-w, Mathews FS, Bruice TC. Catalytic mechanism of quinoprotein methanol dehydrogenase: A theoretical and x-ray crystallographic investigation. Proc Natl Acad Sci USA. 2001;98:432–4.

22. Gäde G, Head EJH. A rapid method for the purification of octopine dehydrogenase for the determination of cell metabolites. Experientia. 1979;35:304–5.

23. Zhang M, Lidstrom ME. Promoters and transcripts for genes involved in methanol oxidation in Methylobacterium extorquens AM1. Microbiology. 2003;149:1033–40.

24. Anthony C. Methanol dehydrogenase, a PQQ-containing quinoprotein dehydrogenase. Subcell Biochem. 2000;35:73–117.

25. Anthony C. The quinoprotein dehydrogenases for methanol and glucose. Archives of Biochemistry Biophysics. 2004;428:0–9.

26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.

27. Tatusova T, DiCuccio M, Badretdin A, Chetverinin N, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24.

28. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2018;46:335–42.

29. O'Leary NA, Wright MW, Brister JR, Ciufò S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetverinin N, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li
W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O’Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shked A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44:D733-45.

30. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-85.

31. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res 41:D387-95.

32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unication of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

33. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42:D199–205.

34. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43:D261-9.

35. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

36. Williams T, Kelley C. 2010. GNUPLOT: An Interactive Plotting Program. User’s Manual.

37. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

38. Singh VK, Mangalam AK, Dwivedi S, Naik S. Primer premier: program for design of degenerate primers from a protein sequence. Biotechniques. 1998;24:318–9.

39. Yongfang Z, Ning X, Yingshan W, Xiao H. Determination of a new prothetic group PQQ using noenzymatic system. Journal of Wuhan University(Natural Science Edition). 1995;41:777–80.

40. Sun Z, Tian B, Xu Z. Identification and functional analysis of a phytoene desaturase gene from the extremely radioresistant bacterium Deinococcus radiodurans. Microbiology. 2007;153:1642–52.

41. Zheng H, Liu H, Tao Z, Wang S, Sun Z, Liu W, Li Y. Total carotenoid differences in scallop tissues of Chlamys nobilis (Bivalve: Pectinidae) with regard to gender and shell colour. Food Chem. 2010;122:1164–7.

42. Lunetta S, Roman M. Determination of coenzyme Q10 content in raw materials and dietary supplements by high-performance liquid chromatography-UV: collaborative study. J AOAC Int. 2008;91:702–8.

43. Kelley LA, Mezulis S, Yates CM, Wass MN, Stemberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845–58.
Figure 1

A circular map of the CLZ nuclear genome. From outside to inside, coding gene (sense-strand), coding gene (negative sense-strand), tRNA and rRNA, gene island (green), GC content, GC-skew (sense-strand), GC-skew (negative sense-strand), sequencing depth. The replication origin (oriC) of the CLZ genome was determined using the Circlator software (parameter: fix start).
Figure 2

Phylogenetic tree of CLZ based on 16S rDNA sequences. Methytorubrum and Methylobacterium are used alternatively for the same species.
Figure 3

Global alignment of the genome.
Figure 4

Functional categorization of nonsynonymous genes as annotated by the COG database.
Figure 5

Fermentation characteristics of wild strain CLZ and mutant strain NI91. The fermentation result of the (a) wild-type strain CLZ, and (b) mutant strain NI91.
Figure 6

Prediction of tertiary structures of DXS. (a) Base insertion in the dxs gene of the NI91 strain caused a frameshift mutation, three-dimensional structure of DXS in the wild-type strain CLZ (b, c), and the mutant strain NI91 (d), differences in their structure are highlighted in green.
Figure 7

Prediction of tertiary structures of mxaD. Structure of the protein encoded by the mxaD gene of the wild-type strain CLZ (a) and the mutant strain NI91 (b) and the green region highlights the mutation.
Figure 8

Expression levels of the dxs and mxaD genes in different samples.
Figure 9

Expression of internal reference genes. The median cycle threshold values for each gene in different samples and the values at the top of the histogram were obtained by NormFinder analysis of CT values of candidate internal reference genes at different culture periods of CLZ and NI91 samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterial.pdf