



Pseudogene product YqiG is important for *pflB* expression and biohydrogen production in *Escherichia coli* BW25113

Muhammad Azman Zakaria¹ · Mohd Zulkhairi Mohd Yusoff^{1,2} · Mohd Rafein Zakaria^{1,2} · Mohd Ali Hassan¹ · Thomas K. Wood³ · Toshinari Maeda⁴

Received: 8 June 2018 / Accepted: 24 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Pseudogenes in the *Escherichia coli* genome are assumed to be non-functional. In this study, Keio collection BW25113Δ*yqiG* and YqiG-producing strain (BW25113/pCA24N-YqiG) were used to evaluate the importance of pseudogene *yqiG* in hydrogen metabolism. Our results show pseudogene protein YqiG was identified as an essential protein in the production of biohydrogen from glucose. The mutant *yqiG* decreased biohydrogen production from 37 μmol mg⁻¹ protein to 6 μmol mg⁻¹ protein compared to the wild-type strain, and glucose consumption was reduced by 80%. Through transcriptional analysis, we found that the *yqiG* mutation represses *pflB* transcription tenfold; *pflB* encodes pyruvate-formate lyase, one of the key enzymes in the anaerobic metabolism of *E. coli*. Moreover, production of YqiG stimulated glycolysis and increased biohydrogen productivity 1.5-fold compared to that of the wild-type strain. Thus, YqiG is important for the central glycolysis reaction and is able to influence hydrogen metabolism activity in *E. coli*.

Keywords *Escherichia coli* · Biohydrogen · Pseudogene · Pyruvate-formate lyase (PflB) · YqiG

Introduction

Hydrogen is an environmental friendly, ecologically clean and renewable energy source, and many studies have been performed using *Escherichia coli* for metabolic engineering

and protein engineering for enhancing hydrogen production (Maeda et al. 2008; Sanchez-Torres et al. 2009). However, there are still some uncharacterized proteins related to biological hydrogen (biohydrogen) production by *E. coli* (Mohd Yusoff et al. 2012, 2013). In addition, there is potential for substantial enhancements for biohydrogen production from glucose or glycerol (Maeda et al. 2018). *E. coli* remains attractive for engineering to increase biohydrogen due to its accessibility for gene manipulation and modification for specific pathways to produce fermentation products (Seol et al. 2014; Nakashima et al. 2014). In addition, a single gene knock-out strain library known as the Keio collection is available (Baba et al. 2006), and these mutants have been used in metabolic engineering studies to investigate the specific genes of interest (Mohd Yusoff et al. 2012).

There are more than 4500 genes in the *E. coli* genome and 178 of the genes are categorised as pseudogenes (Zhou and Rudd 2013). Pseudogenes are prevalent (found between 1 and 5% of genomes) and are considered non-functional (Balakirev and Ayala 2003). Throughout evolution, their function has faded due to interruption of transcription and translation. Hence, pseudogenes are considered as genomic junk due to mutations such as inappropriate stop codons, repetitive elements, frame

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1461-2>) contains supplementary material, which is available to authorized users.

✉ Mohd Zulkhairi Mohd Yusoff
mzulkhairi@upm.edu.my

¹ Department of Bioprocess Technology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

² Laboratory of Biopolymer and Derivatives, Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

³ Department of Chemical Engineering and Biochemistry and Molecular Biology, Pennsylvania State University, 161 Fenske Laboratory, University Park, PA 16802, USA

⁴ Department of Biological Functions Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan

shifts, missense mutations, and lack of transcription elements (Balakirev and Ayala 2003). Moreover, a study of 64 prokaryote genomes including 7000 pseudogene candidates found at least 1–5% arise from large protein families which include the PPE families, ABC transporters, and cytochrome P450, and some pseudogenes have been found in pathogenic bacteria such as in *Mycobacterium leprae*; the conclusion was that the presence of pseudogenes is rare and assumed to be minimal (Liu et al. 2004).

According to Rouchka and Cha (2009), pseudogenes are classified into processed and unprocessed pseudogenes. Processed pseudogenes are from retro transposable events of reverse transcription of mRNA followed by retrotransposition into the genome. This type of pseudogene usually lacks intron regions and consists of flanking repeated regions and poly-A tracts at the carboxyl 3' end (Pink et al. 2011). 20% of processed pseudogenes are 90% similar to mRNA of beta-actin in humans (Bjarnadottir and Jonsson 2005). Meanwhile, improper duplication leads to the production of unprocessed pseudogenes. The unprocessed pseudogenes normally contain introns and regulatory sequences including stop codons, which halt their expression (Pink et al. 2011). Natural transposition events relocate insertion-sequences (IS) (Zhou and Rudd 2013) and this movement can interrupt gene regulation. Some of the IS elements stimulate the formation of hybrid promoters and activate the expression of neighbouring genes in *E. coli* (Barker et al. 2004), while some IS elements can inactivate the expression of genes (Griffiths et al. 2012; Barker et al. 2004).

yqiG is categorised as a pseudogene in *E. coli* genome due to its internal IS elements, *insC9* and *insD9* (Zhou and Rudd 2013) and is one of four *E. coli* pseudogenes that has a key role in hydrogen production in *E. coli* (Mohd Yusoff et al. 2013). However, a detailed understanding of how the pseudogene products influence biohydrogen production is lacking (Maeda et al. 2018). In this study, YqiG was produced to determine how it increases biohydrogen production. We found that YqiG stimulates *pflB* transcription to increase glycolysis and biohydrogen production.

Materials and methods

Bacteria strains and maintenance

The strains used in this study are described in Table 1. Wild-type strains of *E. coli* BW25113 and the mutant *yqiG*, JW5507 (BW25113Δ*yqiG*::kan) were obtained from the National Institute of Genetics. The overexpressed and complemented strains (BW25113/pCA243N-YqiG and BW25113Δ*yqiG*/pCA24N-YqiG) were constructed by introducing plasmid pCA24N harbouring *yqiG* obtained from the ASKA library (Kitagawa et al. 2005) which was also provided by the National Institute of Genetics. All strains were routinely streaked onto Luria–Bertani (LB) agar plates with the addition of 100 μg mL⁻¹ of kanamycin for mutant *yqiG* as described previously (Mohd Yusoff et al. 2013). Meanwhile, 30 μg mL⁻¹ chloramphenicol was added to strains that carry the ASKA plasmid. A single colony was inoculated into LB medium supplemented with an appropriate antibiotic and grown overnight in an incubator shaker at 37 °C, which was then agitated at 120 rpm. The overnight culture was used as an inoculum for the biohydrogen production.

Plasmid isolation, competent cell formation, and electroporation

The plasmid pCA24N was isolated from ASKA JW5507 strain (Kitagawa et al. 2005) and purified using the QIAprep Spin Miniprep kit (QIAGEN, Valencia, CA). Competent cells were prepared as described by Datsenko and Wanner (2000). Five microlitres of purified plasmid was gently mixed with 50 μL of competent cells before electroporation took place using a Bio-Rad Gene Pulser II unit at 12.5 kV cm⁻¹ (Bio-Rad, Hercules, CA). The electroporated cells were immediately re-suspended into 1 mL of LB broth and incubated at 37 °C for 1 h in a dry-bath incubator. The culture was then spread onto LB agar containing 30 μg mL⁻¹ chloramphenicol and further incubated at 37 °C overnight. A single colony from the overnight plate was purified by streaking it onto LB plate with an appropriate antibiotic.

Table 1 Strains and plasmids used in the present work

Strains and plasmids	Genotype/relevant characteristics	References
BW25113	F ⁻ Δ(araD-araB)567ΔlacZ4787 (::rrnB-3)λ ⁻ rph-1 Δ(rhaD-rhaB) 568 hsdR514; parental strain for the Keio collection	Baba et al. (2006)
BW25113 <i>yqiG</i>	BW25113Δ <i>yqiG</i> :: Km ^R	Baba et al. (2006)
BW25113/pCA24N-YqiG	Overexpressed YqiG phenotype in wild-type	This study
BW25113Δ <i>yqiG</i> /pCA24N-YqiG	Complementation YqiG phenotype in mutant <i>yqiG</i> via pCA24N	This study
ASKA JW5507	Harbouring YqiG Cm ^R	Kitagawa et al. (2005)

Km^R kanamycin resistance, Cm^R chloramphenicol resistance

E. coli growth and biohydrogen production

The fermentation medium was prepared using 5.0 g yeast extract, 5.0 g tryptone, 7.0 g K_2HPO_4 , 5.5 g KH_2PO_4 , 0.5 g L-cysteine-HCl· H_2O , 1.0 g $(NH_4)_2SO_4$, 0.25 g $MgSO_4 \cdot 7H_2O$, 0.021 g $CaCl_2 \cdot 2H_2O$, 0.029 g, $Co(NO_3)_2 \cdot 6H_2O$, 0.039 g, $Fe(NH_4)_2SO_4 \cdot 6H_2O$, 2.0 mg nicotinic acid, 0.172 mg Na_2SeO_3 , 0.02 mg $NiCl_2$ and 10 mL of trace element solution containing 0.5 g $MnCl_2 \cdot 4H_2O$, 0.1 g H_3BO_3 , 0.01 g $AlK(SO_4) \cdot 2H_2O$, 0.001 g $CuCl_2 \cdot 2H_2O$, and 0.5 g Na_2EDTA (Rachman et al. 1997) in 1 L with glucose (111 mM) or formate (100 mM) as the carbon source. The fermentation medium had an initial neutral pH (7.0 ± 0.2). One millilitre of overnight culture was inoculated into 9 mL of fermentation medium in 34 mL crimp-top serum vials. The mixture was sparged with nitrogen for 5 min prior to incubation (Maeda et al. 2008). The fermentation for biohydrogen quantification was conducted at 37 °C with three independent experiments. Cell growth was determined based on the optical density measured at 600 nm using a UV/Vis spectrophotometer (JASCO V-530).

Strain verification

A single colony of mutant *yqiG* was transferred into 20 μ L sterilised water and boiled for 5 min before the solution was used as a template for polymerase chain reaction (PCR). PCR verification with kanamycin-specific primers k1, k2 and locus-specific primers was conducted as described elsewhere (Datsenko and Wanner 2000). All primers used for

strain verification are listed in Table 2. The PCR verification was performed to confirm the absence of the *yqiG* gene and the presence of kanamycin gene in the *yqiG* mutant.

RNA extraction and quantitative real-time reverse transcription PCR

Cells were collected during the log phase of fermentation for biohydrogen production. RNA^{later} solution was added into a 1-mL sample at a ratio of 1:1 in screw cap tubes. The mixture was centrifuged at $11,963 \times g$ for 2 min with the supernatant discarded. The cell pellets obtained were immersed in 100 mL of ethanol with dry ice about 10 s and the cells were kept at -70 °C prior to RNA isolation. RNA was isolated using the RNeasy kit (QIAGEN, Valencia, CA) as previously described (Mohd Yusoff et al. 2012). A quantitative real-time reverse transcription PCR (qRT-PCR) was performed to evaluate transcription of the targeted genes (Table 2). Three replicates samples ($50 \text{ ng } \mu\text{L}^{-1}$) were analysed using *rrsG* as a reference for normalisation of the qRT-PCR data (Hagblom et al. 2002).. Expression levels were analysed using the delta–delta method (Pfaffl 2001).

Analytical procedures

Organic acid and glucose determination

At the end of each fermentation, samples were centrifuged and the supernatant was filtered using a syringe filter ($0.45 \mu\text{m}$) (Sartorius, Germany). The filtered samples were

Table 2 Primers used for transcriptional analysis and strain verification

Targeted gene	Sequences	Relevant characteristics
<i>rrsG</i>	f-5'-TATTGCACAATGGGCGCAAG-3' r-5'-ACTTAACAAACCGCTGCGT-3'	Housekeeping gene for expression data (Mohd Yusoff et al. 2012)
<i>fdhF</i>	f-5'-GGATTTCTACGGTGCGACTTAC-3' r-5'-GGTACTCGTCGGTGAGTTTGTC-3'	Catalyse $HCOO^-$ and water to hydrogen (Maeda et al. 2008)
<i>fhlA</i>	f-5'-GTGTATTGCAGGAACAGGAGTTTG-3' r-5'-GAATACGTTTCAGGCGGTAATAGAG-3'	Transcriptional activator for FHL complex (Baba et al. 2006)
<i>hycE</i>	f-5'-ATCAGCTGACTGTCACCGTAAAG-3' r-5'-GTAATCCAACACTTAGTGCCCTTC-3'	Large subunit of hydrogenase 3 (Sanchez-Torres et al. 2009)
<i>hycA</i>	f-5'-CTACTGCAATTCGCTGGTTTCAG-3' r-5'-CGACGTAATACTCGATGGTGTG-3'	Repressor gene for FHL complex (Bagramyan and Trchounian 2003)
<i>hybC</i>	f-5'-AGCAGGTTTATAAGGTTGATACCG-3' r-5'-AGGTATTCATCGGAATGAGAAGTG-3'	Large subunit of hydrogenase 2 (Forzi and Sawers 2007; Menon et al. 1994)
<i>hyaB</i>	f-5'-GGCGTTCGTTGAACGTATCT-3' r-5'-CAAGCTGATAGAAGTGCACCAG-3'	Large subunit of hydrogenase 1 (Forzi and Sawers 2007)
<i>pflB</i>	f-5'-CGGTATCGACTACCTGATGAAAG-3' r-5'-CAGAGATGTCGTAGCCGTATTTC-3'	Conversion of pyruvate to formate (this study)
<i>yqiG</i>	f-5'-TATGGAATCCCCTTTCGTTTC-3' r-5'-CCTAGCATGGGCATAGTATTACT-3'	Confirm deletion of <i>yqiG</i> gene (this study)
k1	5'-CAGTCATAGCCGAATAGCCT-3'	Confirmation of the presence of kanamycin gene (Datsenko and Wanner 2000)
k2	5'-CGGTGCCCTGAATGAAGTGC-3'	

analysed using high-performance liquid chromatography (HPLC); Shimadzu LC-10AD with a conductivity detector CDD-6A equipped with a Shim-packed SCR-102H (8.0 mm LD×300 mm L)×2 column (Mohd Yasin et al. 2013). Glucose concentrations were determined using the Glucose C2 Kit (Wako, Osaka). A 20- μ L filtered sample was added into 2.5 mL buffer solution. Then, the mixtures were incubated at 37 °C for 5 min. The optical density of the cells was measured at specific absorbance of 505 nm. Glucose standard curve was plotted to determine the glucose concentration in the samples.

Biohydrogen analysis

The biohydrogen produced in the headspace was measured using a 6890 N (Agilent Technologies, Glastonbury, CT) gas chromatograph equipped with an 80–100 mesh Porapak Q column (Suppelco, Bellefonte, PA) and a thermal conductivity detector. Nitrogen gas was used as the carrier at a flow rate of 20 mL min⁻¹. Column, injector, and detector temperature were maintained at 70 °C, 100 °C, and 200 °C, respectively (Mohd Yasin et al. 2013).

Results and discussion

Mutant *yqiG* reduced biohydrogen production

In a preliminary study, biohydrogen was assayed for mutant *yqiG* of the Keio collection (BW25113 $\Delta yqiG$), the YqiG-producing strain (BW25113/pCA24N-YqiG), the complemented strain (BW25113 $\Delta yqiG$ /pCA24N-YqiG), and wild-type (BW25113) using glucose as a carbon source. Biohydrogen production was detected after 2 h for the wild-type (maximum normalized productivity of 38 μ mol H₂ mg⁻¹ protein at 7 h), whereas mutant *yqiG* showed the lowest biohydrogen generation (1.9 μ mol H₂ mg⁻¹ protein) (Supplementary Fig. S1). The result confirms biohydrogen production ability was diminished by the deficiency of YqiG phenotype (Mohd Yusoff et al. 2013). However, biohydrogen production from the strain producing YqiG was lower at the beginning but eventually became higher after 7 h (Supplementary Fig. S1). The productivity reached 60 μ mol H₂ mg⁻¹ protein. The lower biohydrogen observed at 3 h fermentation indicates that YqiG plays an important role in the beginning of glycolysis process in biohydrogen metabolism or at least plays a role in catalysing the conversion of glucose into essential by-products for biohydrogen synthesis.

Higher amounts of absolute biohydrogen were obtained for the strains producing YqiG (Fig. 1), including BW25113 $\Delta yqiG$; hence, the *yqiG* mutation could be complemented. Also, higher biohydrogen production was observed for the wild-type strain producing YqiG compared to the wild-type

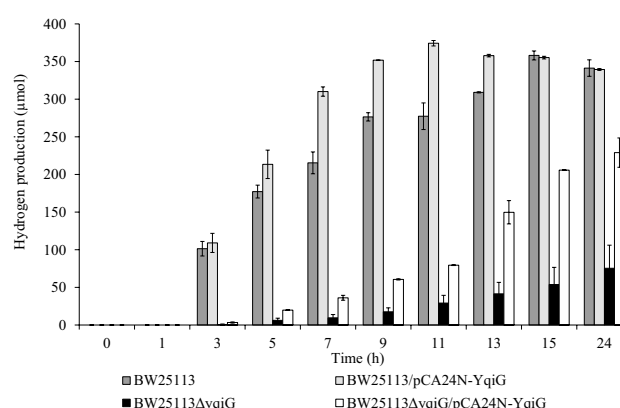


Fig. 1 Biohydrogen production profile in complex glucose medium. BW25113 (dark grey column), and BW25113/pCA24N-YqiG (light grey column), BW25113 $\Delta yqiG$ (black column) and BW25113 $\Delta yqiG$ /pCA24N-YqiG (plain column)

strain: 34 μ mol h⁻¹ vs. 25 μ mol h⁻¹, respectively. Hence, YqiG increases biohydrogen productivity.

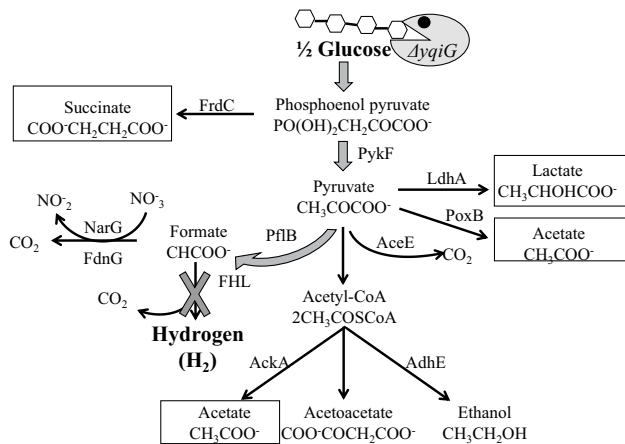
Overexpression of YqiG enhanced biohydrogen production during glycolysis

Biohydrogen production from mutant *yqiG* was further explored using glucose and formate as carbon sources. Table 3 shows a summary of the biohydrogen productivity, growth rate, and substrate consumption for BW25113, BW25113 $\Delta yqiG$, BW25113/pCA24N-YqiG, and BW25113 $\Delta yqiG$ /pCA24N-YqiG. Different trends were obtained in formate medium compared to glucose; for example, the complementation strain showed slightly lower biohydrogen productivity compared to the wild-type strain (Table 3). In addition, for the *yqiG* mutant, 219.6 μ mol H₂ mg⁻¹ protein was observed from fermentation with formate compared to 6 μ mol H₂ mg⁻¹ with glucose. This shows that the mutant has the ability to consume formate better than glucose. Formate is the main intermediate for the formate hydrogen lyase in hydrogen metabolism (Bagramyan and Trchounian 2003). This result suggests that the mutant has a deficiency in converting glucose into formate. To investigate this further, we assayed glucose consumption and found the highest glucose consumption was in BW25113/pCA24N-YqiG followed by BW25113 $\Delta yqiG$ /pCA24N-YqiG (Table 3). Corroborating this result, the lowest glucose consumption was observed with BW25113 $\Delta yqiG$. The glucose was consumed by the *yqiG* mutant mainly for cell growth and cell maintenance since its specific growth was about 0.79 h⁻¹ compared to the wild-type strain (0.85 h⁻¹) (Table 3).

The *yqiG* mutation also does not affect growth with formate (Table 3). Thus, the absence of *yqiG* was not involved in the growth-related phenotype. In addition, the growth of all strains was the same in formate, than 1.0 h⁻¹,

Table 3 Biohydrogen productivity, glucose consumption, and specific growth rate during anaerobic fermentations with glucose and formate

Strains	Glucose			Formate	
	Productivity ($\mu\text{mol H}_2 \text{ mg}^{-1} \text{ protein}$)	Specific growth rate (h^{-1})	Glucose reduction (%)	Productivity ($\mu\text{mol H}_2 \text{ mg}^{-1} \text{ protein}$)	Specific growth rate (h^{-1})
BW25113	37 ± 2	0.85 ± 0.03	36	250.8 ± 0.3	0.2 ± 0
BW25113 $\Delta yqiG$	6 ± 1	0.79 ± 0.03	21	219.1 ± 0.2	0.1 ± 0
BW25113/pCA24N-YqiG	56 ± 3	0.80 ± 0.05	44	279.1 ± 0.6	0.14 ± 0.04
BW25113 $\Delta yqiG$ /pCA24N-YqiG	46 ± 12	0.69 ± 0.02	38	211.5 ± 11	0.12 ± 0.06

**Fig. 2** General hydrogen metabolism and by-products generated during biohydrogen production from glucose in *E. coli*. Mutant *yqiG* showed biohydrogen deficiency when glucose was used as a carbon source

which was in line with a previous study (Maeda et al. 2008) reporting the reduced growth of *E. coli* in formate medium, especially in the presence of IPTG. Together, these results show YqiG plays an important role at the early stage of glycolysis and perhaps influences the synthesis of components of biohydrogen production such as maturation proteins and hydrogenase. Thus, a transcriptional analysis was performed to determine the expression level of specific genes during biohydrogen metabolism with the *yqiG* mutant.

Table 4 Distribution of organic acids after 24 h fermentation with glucose

	Organic acid concentration (mM)			
	Succinic acid	Lactic acid	Acetic acid	Formic acid
BW25113	ND	27.6 ± 0.2	14.8 ± 0.3	ND
BW25113 $\Delta yqiG$	1.5 ± 0	22 ± 3	7.8 ± 0.6	13.1 ± 0.9
BW25113/pCA24N	ND	32 ± 2	13 ± 2	ND
BW25113/pCA24N-YqiG	ND	43 ± 1	21.0 ± 0.1	ND
BW25113 $\Delta yqiG$ /pCA24N-YqiG	2.8 ± 0.1	32 ± 1	16.4 ± 1.6	ND

ND not detected from HPLC analysis

By-products and hydrogen metabolism

Figure 2 shows the possible pathway of glucose fermentation and biohydrogen production in *E. coli*. Acidic pH is one of crucial factors in biohydrogen yield (Ghimire et al. 2018); a high concentration of lactic acid in the fermentation medium reduces the culture pH, thus affecting biohydrogen production. Bacterial growth is suppressed since undissociated forms of acid will influence the cell internal pH (Jonsson et al. 2013). In our work, four organic acids were detected in the fermentation broth during biohydrogen production, namely succinic acid, lactic acid, acetic acid, and formic acid (Table 4). Lactic acid dominated all strains (20–43 mM), while the *yqiG* strain produced the lowest amount of lactic acid. Lactate dehydrogenase encoded by *ldhA* is responsible for the conversion of pyruvate to lactic acid in glycolysis (Fig. 2). Critically, the strain producing YqiG (BW25113/pCA24N-YqiG) produced the highest amount of lactic acid (43 mM). It appears that YqiG promotes the utilisation of glucose for generating succinate and lactate through fumarate reductase (*frdC*) and *ldhA* activity, respectively (Tran et al. 2014). These reactions occur prior to the conversion of pyruvate into formate.

Formic acid was not detected for the wild-type and other strains since all the formate generated was converted into biohydrogen through the pyruvate formate lyase system (PFL). However, small amounts of formic acid (13 mM) were detected for the *yqiG* mutant at the end of the fermentation. The small amount of formate was probably insufficient to be converted into biohydrogen through pyruvate-formate

lyse encoded by *pflB* (Hallenbeck and Ghosh 2009). This is supported in that with high formate concentrations (100 mM) a maximum of 219 $\mu\text{mol H}_2 \text{ mg}^{-1}$ protein was produced from mutant *yqiG* (Table 3). In formate metabolism under anaerobic conditions, PFL was induced and cleaved pyruvate to produce formate and acetyl-CoA (Sinha et al. 2015). On the other hand, synthesis of PFL was coordinated by the synthesis of formate transport protein encoded by *focA* (Beyer et al. 2013). A deletion of *focA* has resulted in accumulation of intracellular formate (Fan et al. 2009). Thus, it can be assumed that there are synergistic effects between the deficiency of YqiG and glycolysis flux in hydrogen metabolism, which may influence the decrease of biohydrogen production in the mutant *yqiG*.

Transcriptional analysis during biohydrogen production

Several genes are involved in biohydrogen production from glucose in *E. coli*. Biohydrogen is mainly produced from formate by formate hydrogen lyase (FHL) complex. The FHL complex is composed of hydrogenase 3 (*hycE* encodes large subunits) (Bagramyan and Trchounian 2003), formate dehydrogenase encoded by *fdhF* (Sanchez-Torres et al. 2009), and maturation proteins coded by *hyp* genes (Sinha et al. 2015). The FHL converts formate into 2H^+ , 2e^- and carbon dioxide. On the other hand, HycA is the repressor of the FHL complex (Bagramyan and Trchounian 2003) that can be activated by FhlA (Sanchez-Torres et al. 2009). Most of the potential genes involved in hydrogen metabolism were used as targeted genes in the transcriptional analysis study.

The RNA was isolated during the exponential phase of biohydrogen production (after 4–5 h), and the transcriptional analysis demonstrated the highest difference in regulation occurred for the *pflB* gene for the mutant *yqiG* compared to wild-type (10-fold repression, Fig. 3). The down regulation indicates YqiG stimulates *pflB* transcription either directly or indirectly. Moreover, production of YqiG restored *pflB* expression (Fig. 3). In addition, the mutant *yqiG* had down-regulation of *hyaB* and *hybC* that encode for the large subunits of hydrogenase 1 and hydrogenase 2, respectively (Forzi and Sawers 2007). HyaB and HybC actively participate in the utilisation of biohydrogen via an energy-conservation pathway (Noguchi et al. 2010). Thus, the down regulation of these genes in mutant *yqiG* can be neglected.

The repression of *pflB* in the mutant *yqiG* suggested that the YqiG has an interaction in formate metabolism. It appears that the role of *pflB* was interrupted during glycolysis, whereas the function of *pflB* is to receive high flux levels in the glycolytic pathway and convert it into formate for biohydrogen production through FHL complex system. Inactivation of intermediate steps in glycolysis may divert the reactions in the central metabolic network towards other

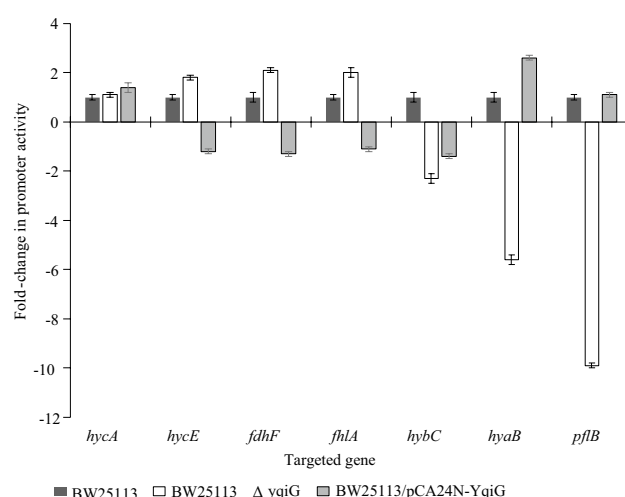


Fig. 3 Fold change in promoter activity during transcriptional analysis of targeted genes (*hycA*, *hycE*, *fdhF*, *fhlA*, *hybC*, *hyaB*, and *pflB*) using quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR)

reactions leading to the accumulation of other by-products (Meza et al. 2012). In addition, the phosphotransferase system is involved in transporting many sugars (e.g., glucose) through phosphoenolpyruvate (PEP). The phosphoryl group on PEP is transferred to imported sugars via several proteins. PEP is catalysed by pyruvate kinase (PykF) generating pyruvate and then formate via *pflB* (Muñoz and Ponce 2003). PEP is the key central metabolism intermediate involved in glucose transportation in several biosynthetic pathways and in allosteric regulation of glycolytic enzymes (Campos-Bermudez et al. 2010). Hence, without YqiG, high flux levels in the glycolytic pathway cannot be sustained, reducing the conversion efficiency of pyruvate into formate and affecting biohydrogen generation.

Protein purification of YqiG was performed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to confirm *yqiG* is converted into functional protein. The analysis was done using *yqiG*/pCA24N-YqiG with 1 mM IPTG. However, there was no targeted protein observed after purification (Supplementary Fig. S2). This suggested that YqiG is in outer membrane of *E. coli* due to the difficulty in protein separation (Kitagawa et al. 2005). Through protein sequencing analysis, the secondary structure of hydropathy, amphipathicity, and N-terminal targeting sequence was investigated. Most of the outer membrane proteins have β -barrel three-dimensional structures. Hence, a β -barrel finder (BFF) was created to predict the presence of β -barrel proteins in the *E. coli* genome. As a result, the YqiG protein was predicted to belong to a fimbrial export usher protein family and considered as an outer membrane usher protein. YqiG is also predicted to be involved in the export and assembly of a fimbrial subunits across the outer

membrane. This subunit is related to pilus formation for cell adhesion (Nuccio and Bäumler 2007). Due to lack of YqiG function in the *yqiG* mutant, perhaps some of the transportation activities between inner and outer cells in *E. coli* metabolism were disrupted leading to the reduction in biohydrogen production.

Indeed, YqiG is not solely responsible for the conversion of formate into biohydrogen. An impaired relationship between formate reduction and down regulation in *pflB* shows that YqiG might not absolutely control PFL regulation, but instead has an essential role in the central glycolysis. Interrelation of protein activities should be considered to understand the influence of uncharacterised proteins such as YqiG in hydrogen metabolism. Putting together the connection between fimbrial usher proteins and glycolytic pathways may answer how YqiG functions in *E. coli*. As a working model, the *yqiG* gene product might be related to the formation of pilus in the outer membrane, which may then affect biohydrogen productivity since some of hydrogenase enzymes are membrane associated (Pandelia et al. 2010). Indirectly, YqiG influenced the conversion of phosphoenolpyruvate to pyruvate and conversion of pyruvate to formate, which was an important precursor for biohydrogen. Thus, further investigation is required to identify how mutant YqiG increases biohydrogen production from glucose. A global transactional analysis using microarray or next generation of pyrosequencing including MiSeq analysis would be an important study for a detailed explanation on the role of pseudogene in hydrogen metabolism.

Conclusions

Inactivation of YqiG reduces biohydrogen in *E. coli*, and this defect may be complemented. Based on transcriptional analysis, we show here the *yqiG* mutation represses *pflB* transcription; PflB is involved in the conversion of pyruvate into formate. Moreover, a producing YqiG increases biohydrogen productivity by 1.5-fold over wild-type. It can be concluded that YqiG is important in glycolysis; thus the product of *yqiG* pseudogene is an essential proteins during hydrogen metabolism in *E. coli*.

Acknowledgements We express our gratitude to the National Institute of Genetics (Japan) for providing the Keio mutant and ASKA strains. The authors would like to thank the Ministry of Higher Education, Malaysia for providing Fundamental Research Grant Scheme (02-01-14-1401FR), the School of Graduate Studies, Universiti Putra Malaysia and Japan Student Services Organisation for the scholarship to Muhammad Azman Zakaria during this study.

Author contributions Study conception and design: MZMY and TM. Experiment, analysis and interpretation of data: MAZ and MZMY. Draft and revise the manuscript: MZMY, TM and MRZ. Coordinate the experiments and critical revision: MAH and TKW.

Compliance with ethical standards

Conflict of interest The authors declare that there have no conflict of interest in relation to this article.

References

- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006–2008
- Bagramyan K, Trchounian A (2003) Structural and functional features of formate hydrogen lyase, an enzyme of mixed-acid fermentation from *Escherichia coli*. *Biochemistry* 68:1159–1170
- Balakirev ES, Ayala FJ (2003) Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet* 37:123–151
- Barker CS, Pruss BM, Matsumura P (2004) Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the *flhD* operon. *J Bacteriol* 186:7529–7537
- Beyer L, Doberenz C, Falke D, Hunger D, Suppmann B, Sawers RG (2013) Coordination of FocA and pyruvate formate-lyase synthesis in *Escherichia coli* demonstrates preferential translocation of formate over other mixed-acid fermentation products. *J Bacteriol* 195:1428–1435
- Bjarnadottir H, Jonsson JJ (2005) A rapid real-time qRT-PCR assay for ovine beta-actin mRNA. *J Biotechnol* 117:173–182
- Campos-Bermudez VA, Bologna FP, Andreo CS, Drincovich MF (2010) Functional dissection of *Escherichia coli* phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation. *FEBS J* 277:1957–1966
- Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645
- Fan Z, Yuan L, Chatterjee R (2009) Increased hydrogen production by genetic engineering of *Escherichia coli*. *PLoS One* 4:32–44
- Forzi L, Sawers RG (2007) Maturation of [NiFe]-hydrogenases in *Escherichia coli*. *BioMetals* 20:565–578
- Ghimire A, Trably E, Frunzo L, Pirozzi F, Lens PNL, Esposito G, Cazier EA, Escudé R (2018) Effect of total solids content on biohydrogen production and lactic acid accumulation during dark fermentation of organic waste biomass. *Bioresour Technol* 248:180–186
- Griffiths AJF, Wessler SR, Varroll SB, Doebley J (2012) Introduction to genetic analysis, 10th edn. W. H. Freeman and Company, New York
- Hagblom MM, Apetroaie C, Andersson MA, Salkinoja-Salonen MS (2002) Quantitative analysis of cereulide, the emetic toxin of *Bacillus cereus*, produced under various conditions. *Appl Environ Microbiol* 68:2479–2483
- Hallenbeck PC, Ghosh D (2009) Advances in fermentative biohydrogen production: the way forward? *Trends Biotechnol* 27:287–297
- Jonsson LJ, Alriksson B, Nilvebrant NO (2013) Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol Biofuels* 6:16
- Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res* 12:291–299
- Liu Y, Harrison PM, Kunin V, Gerstein M (2004) Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol* 5(9):R64

- Maeda T, Sanchez-Torres V, Wood TK (2008) Metabolic engineering to enhance bacterial hydrogen production. *Microb Biotechnol* 1:30–39
- Maeda T, Tran KT, Yamasaki R, Wood TK (2018) Current state and perspectives in hydrogen production by *Escherichia coli*: roles of hydrogenases in glucose or glycerol metabolism. *Appl Microbiol Biotechnol* 102:2041–2050
- Menon NK, Chatelus CY, Dervartanian M, Wendt JC, Shanmugam KT, Peck HD, Przybyla AE (1994) Cloning, sequencing, and mutational analysis of the *hyb* operon encoding *Escherichia coli* hydrogenase 2. *J Bacteriol* 176:4416–4423
- Meza E, Becker J, Bolivar F, Gosset G, Wittmann C (2012) Consequences of phosphoenolpyruvate:sugar phosphotransferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microb Cell Fact* 11:127–140
- Mohd Yasin NH, Fukuzaki M, Maeda T, Miyazaki T, Che Maail CMH, Ariffin H, Wood TK (2013) Biohydrogen production from oil palm frond juice and sewage sludge by a metabolically engineered *Escherichia coli* strain. *Int J Hydrog Energy* 38:10277–10283
- Mohd Yusoff MZ, Maeda T, Sanchez-Torres V, Ogawa HI, Shirai Y, Hassan MA, Wood TK (2012) Uncharacterized *Escherichia coli* proteins YdjA and YhjY are related to biohydrogen production. *Int J Hydrog Energy* 37:17778–17787
- Mohd Yusoff MZ, Hashiguchi Y, Maeda T, Wood TK (2013) Four products from *Escherichia coli* pseudogenes increase hydrogen production. *Biochem Biophys Res Commun* 439:576–579
- Muñoz M, Ponce E (2003) Pyruvate kinase: current status of regulatory and functional properties. *Comp Biochem Physiol B Biochem Mol Biol* 135:197–218
- Nakashima N, Akita H, Hoshino T (2014) Establishment of a novel gene expression method, BICES (biomass-inducible chromosome-based expression system), and its application to the production of 2,3-butanediol and acetoin. *Metab Eng* 25:204–214
- Noguchi K, Riggins DP, Eldahan KC, Kitko RD, Slonczewski JL (2010) Hydrogenase-3 contributes to anaerobic acid resistance of *Escherichia coli*. *PLoS ONE* 5:e10132
- Nuccio S-P, Bäuml AJ (2007) Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev* 71:551–575
- Pandelia ME, Fourmond V, Tron-Infossi P, Lojou E, Bertrand P, Leger C, Giudici-Orticoni MT, Lubitz W (2010) Membrane-bound hydrogenase I from the hyperthermophilic bacterium *Aquifex aeolicus*: enzyme activation, redox intermediates and oxygen tolerance. *J Am Chem Soc* 132:6991–7004
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:2002–2007
- Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Francisco Carter DR (2011) Pseudogenes: pseudo-functional or key regulators in health disease? *RNA* 17:792–798
- Rachman MA, Furutani Y, Nakashimada Y, Kakizono T, Nishio N (1997) Enhanced hydrogen production in altered mixed acid fermentation of glucose by *Enterobacter aerogenes*. *J Ferment Bioeng* 83:358–363
- Rouchka EC, Cha IE (2009) Current trends in pseudogene detection and characterization. *Curr Bioinform* 4:112–119
- Sanchez-Torres V, Maeda T, Wood TK (2009) Protein engineering of the transcriptional activator FhlA to enhance hydrogen production in *Escherichia coli*. *Appl Environ Microbiol* 75:5639–5646
- Seol E, Ainala SK, Sekar BS, Park S (2014) Metabolic engineering of *Escherichia coli* strains for co-production of hydrogen and ethanol from glucose. *Int J Hydrog Energy* 39:19323–19330
- Sinha P, Roy S, Das D (2015) Role of formate hydrogen lyase complex in hydrogen production in facultative anaerobes. *Int J Hydrog Energy* 40:8806–8815
- Tran KT, Maeda T, Wood TK (2014) Metabolic engineering of *Escherichia coli* to enhance hydrogen production from glycerol. *Appl Microbiol Biotechnol* 98:4757–4770
- Zhou J, Rudd KE (2013) EcoGene 3.0. *Nucleic Acids Res* 41:613–624