

UC Merced

UC Merced Previously Published Works

Title

Complete genome sequence of Enterobacter sp. IIT-BT 08: A potential microbial strain for high rate hydrogen production

Permalink

<https://escholarship.org/uc/item/6t85t4vw>

Journal

Environmental Microbiome, 9(2)

ISSN

1944-3277

Authors

Khanna, Namita
Ghosh, Ananta Kumar
Huntemann, Marcel
[et al.](#)

Publication Date

2013-09-01

DOI

10.4056/sigs.4348035

Peer reviewed

Complete genome sequence of *Enterobacter sp.* IIT-BT 08: A potential microbial strain for high rate hydrogen production

Namita Khanna¹, Ananta Kumar Ghosh¹, Marcel Huntemann², Shweta Deshpande², James Han², Amy Chen², Nikos Kypides², Kostas Mavrommatis², Ernest Szeto², Victor Markowitz², Natalia Ivanova², Ioanna Pagani², Amrita Pati², Sam Pitluck², Matt Nolan², Tanja Woyke², Hazuki Teshima³, Olga Chertkov³, Hajnalka Daligault³, Karen Davenport³, Wei Gu³, Christine Munk³, Xiaojing Zhang³, David Bruce³, Chris Detter³, Yan Xu³, Beverly Quintana³, Krista Reitenga³, Yulia Kunde³, Lance Green³, Tracy Erkkila³, Cliff Han³, Evelyne-Marie Brambilla⁴, Elke Lang⁴, Hans-Peter Klenk⁴ Lynne Goodwin³, Patrick Chain³, Debabrata Das^{1*}

¹Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India

²Department of Energy Joint Genome Institute, Walnut Creek CA USA

³Lawrence Livermore National Laboratory, Livermore CA USA

⁴Leibniz-Institute—DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

Keywords: *Enterobacter sp.* IIT-BT 08, genome sequence, facultative anaerobe, biohydrogen

*Correspondence: Debabrata Das (ddas.iitkgp@gmail.com)

Enterobacter sp. IIT-BT 08 belongs to Phylum: *Proteobacteria*, Class: *Gammaproteobacteria*, Order: *Enterobacteriales*, Family: *Enterobacteriaceae*. The organism was isolated from the leaves of a local plant near the Kharagpur railway station, Kharagpur, West Bengal, India. It has been extensively studied for fermentative hydrogen production because of its high hydrogen yield. For further enhancement of hydrogen production by strain development, complete genome sequence analysis was carried out. Sequence analysis revealed that the genome was linear, 4.67 Mbp long and had a GC content of 56.01%. The genome properties encode 4,393 protein-coding and 179 RNA genes. Additionally, a putative pathway of hydrogen production was suggested based on the presence of formate hydrogen lyase complex and other related genes identified in the genome. Thus, in the present study we describe the specific properties of the organism and the generation, annotation and analysis of its genome sequence as well as discuss the putative pathway of hydrogen production by this organism.

Introduction

Hydrogen has great promise in contributing substantially to the renewable energy demands of the future. It is considered a dream fuel by virtue of the fact that it is renewable, does not evolve green house gases, has the highest energy content per unit mass of any known fuel (143 GJ t⁻¹), is easily converted to electricity by fuel cells and upon combustion, gives water as the only byproduct [1]. Moreover, hydrogen is the third most abundant element on Earth. However, finding simple, inexpensive ways to extract hydrogen and produce it in a pure gaseous form is a crucial step toward making the "hydrogen economy" a reality. Considering this, hydrogen production using microbes is thought to be a promising technique to produce economical, abundant hydrogen without utilizing fossil fuels. Many microbial species have been reported for hydrogen production [2].

Among them, *Enterobacter sp.* IIT-BT 08 (MTCC 5373, DSM 24603) was reported as a high rate hydrogen producer [3]. It is a Gram negative, facultative anaerobe that can grow and produce hydrogen from a wide range of simple sugars and complex polysaccharides [4]. In the past decade, the group at the Bioprocess Engineering Laboratory at IIT Kharagpur, India, has extensively worked on this organism using various fermentative approaches and established it as one of the highest yielding hydrogen producers [5]. The novelty of the organism lies in the amount of hydrogen (2.2 mol H₂ mol⁻¹ glucose) it can produce at ambient temperature (37 °C) and atmospheric pressure as compared to other closely related species reported in literature. Besides, high rate of continuous hydrogen production has been reported using immobilized *Enterobacter sp.* IIT-BT 08

and waste as substrate using 20 L and 800 L reactors [5]. Therefore, whole genome sequencing of this potential strain was considered to determine the genes responsible for the high rate hydrogen production. In this report we present a summary of the properties and features of *Enterobacter sp.* IIT-BT 08 genome and also suggest a putative pathway for hydrogen production.

Classification and features

E. sp. IIT-BT 08 was isolated from the leaves of a local plant near the Kharagpur railway station, Kharagpur, West Bengal, India [4]. The bacterium is a Gram negative, small, motile, catalase positive rod [4,6,7] belonging to the family *Enterobacteriaceae* (Table 1). To characterize the strain, a set of standard tests were carried out according to Bergey's Manual and the results showed that the strain belongs to *Enterobacter* species. 16S rRNA sequencing by Microbial Type Culture Collection (MTCC), Chandigarh further confirmed the strain identity. The genetic complexity of the organism is illustrated in the phylogenetic tree of the 16S RNA region (Figure 1). Initially the strain was classified as *Enterobacter sp.* IIT-BT 08, however, whole genome sequencing of the strain revealed sequence variation in the six 16S rRNA copies of the strain. We presume that this may have been the source of difficulty in the initial mis-identification of the strain. Currently, without a complete set of type strain genome sequences available for a more detailed taxonomic identification, the name of the strain has been changed to *Enterobacter sp.* IIT-BT 08.

Genome project history

Genome sequencing information

Enterobacter sp. IIT-BT 08 is a promising hydrogen producer and can utilize waste as substrate for hydrogen production [4]. Therefore, it was considered essential to sequence the whole genome of the organism to determine the genes that contributed towards hydrogen production. Besides, complete genome information was also critical to facilitate studies on genetic engineering of the organism for further enhancement of its hydrogen production potential. Therefore, the group applied for the Community Sequencing Program-2010 (CSP-2010) offered by DoE-JGI.

One of the DOE missions is to address the critical question of depleting energy reserves by creating a new generation of biological research enabled

by the genome revolution. This organism therefore appeared relevant to this mission and was selected for sequencing. The genome sequence was completed on May 21, 2012. Quality assurance was done by the DSMZ (Braunschweig, DE), finishing and annotation was completed at Joint Genome Institute. A summary of the project information is shown in Table 2, which also presents the project information and its association with MIGS version 2.0 compliance [8].

Growth conditions and DNA isolation

For genomic DNA isolation, *Enterobacter sp.* was cultivated overnight in nutrient broth at 37 °C and 200 rpm in a gyratory incubator shaker. DNA isolation was carried out by Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) institute. For DNA isolation, the strain was grown in DSMZ medium 381 (Luria-Bertani Medium) at 37°C. DNA was isolated from 1-1.5 g of cell paste using Jetflex Genomic DNA Purification Kit (Genomed_600100) following the manufacturer's recommendations for Gram-positive bacteria (which were more efficient than the conditions recommended for Gram-negative cells). The identity of the DNA was confirmed via 16S rRNA gene sequencing and the quality was analyzed following the recommendations of the sequencing center (JGI), including pulse-field gel electrophoresis.

Genome sequencing and assembly

The draft genome of *Enterobacter sp.* IIT-BT 08 was generated at the DOE Joint Genome Institute (JGI) using Illumina data [22]. For this genome, JGI constructed and sequenced an Illumina short-insert paired-end library with an average insert size of 231 +/- 59 bp which generated 24,130,984 reads and an Illumina long-insert paired-end library with an average insert size of 8,267 +/- 2,204 bp which generated 13,553,468 reads totaling 5,653 Mbp of Illumina data. (unpublished, Feng Chen). All general aspects of library construction and sequencing performed at the JGI can be found at the JGI website. The initial draft assembly contained 21 contigs in 3 scaffold(s). The initial draft data was assembled with Allpaths, version 39750, and the consensus was computationally shredded into 10 Kbp overlapping fake reads (shreds). The Illumina draft data was also assembled with Velvet, version 1.1.05 [23], and the consensus sequences were computationally shredded into 1.5 Kbp overlapping fake reads (shreds). The Illumina draft data was assembled again with Velvet using the shreds from the

first Velvet assembly to guide the next assembly. The consensus from the second Velvet assembly was shredded into 1.5 Kbp overlapping fake reads. The fake reads from the Allpaths assembly and both Velvet assemblies and a subset of the Illumina CLIP paired-end reads were assembled using parallel phrap, version 4.24 (High Performance Software, LLC). Possible mis-assemblies were corrected with manual editing in Consed [24-26]. Gap closure was accomplished using repeat resolution software (Wei Gu, unpublished), and sequencing of bridging PCR fragments with Sanger and/or PacBio (unpublished, Cliff Han) technologies. For

improved high quality draft and noncontiguous finished projects, one round of manual/wet lab finishing may have been completed. Primer walks, shatter libraries, and/or subsequent PCR reads may also be included for a finished project. A total of 0 additional sequencing reactions, 6 PCR PacBio consensus sequences, and 0 shatter libraries were completed to close gaps and to raise the quality of the final sequence. The total estimated size of the genome is 4.7 Mb and the final assembly is based on 5,653 Mbp of Illumina draft data, which provides an average 1,203× coverage of the genome.

Table 1. Classification and general features of *Enterobacter sp.* IIT-BT 08 according to the MIGS recommendation [8]

| MIGS ID | Property | Term | Evidence code |
|----------|----------------------------|---|----------------|
| | | Domain <i>Bacteria</i> | TAS [9] |
| | | Phylum <i>Proteobacteria</i> | TAS [10] |
| | | Class <i>Gammaproteobacteria</i> | TAS [11-13] |
| | Current classification | Order <i>Enterobacteriales</i> | TAS [14] |
| | | Family <i>Enterobacteriaceae</i> | TAS [15-17] |
| | | Genus <i>Enterobacter</i> | TAS [15,17-20] |
| | | Species <i>Enterobacter sp.</i> | [15,18,20] |
| | | Type strain IIT-BT 08 | |
| | Gram stain | Negative | NAS |
| | Cell shape | Rod shaped | IDA |
| | Motility | motile via peritrichous flagella | IDA |
| | Sporulation | Non-sporulating | IDA |
| | Temperature range | 25 – 40 °C | TAS [4] |
| | Optimum temperature | 37 °C | TAS [4] |
| | Carbon source | carbohydrates | TAS [4] |
| | Energy metabolism | Chemoorganotrophic | IDA |
| | Terminal electron receptor | Oxygen | IDA |
| MIGS-6 | Habitat | Plant leaves | TAS [4] |
| MIGS-6.3 | Salinity | Not studied | |
| MIGS-22 | Oxygen | Facultative anaerobe; grows well under oxic and anoxic conditions | TAS [4] |
| MIGS-15 | Biotic relationship | Free-living | IDA |
| MIGS-14 | Pathogenicity | None | |
| MIGS-4 | Geographic location | Kharagpur in the district of West Midnapur, West Bengal, India. | TAS [4] |
| MIGS-5 | Sample collection time | July, 1999 | |
| MIGS-4.1 | Latitude | 22 02' 30" | |
| MIGS-4.2 | Longitude | 87 11' 0" | |
| MIGS-4.3 | Depth | NA | |
| MIGS-4.4 | Altitude | NA | |

a) MIGS: The minimum information about a genome sequence

b) Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [21].

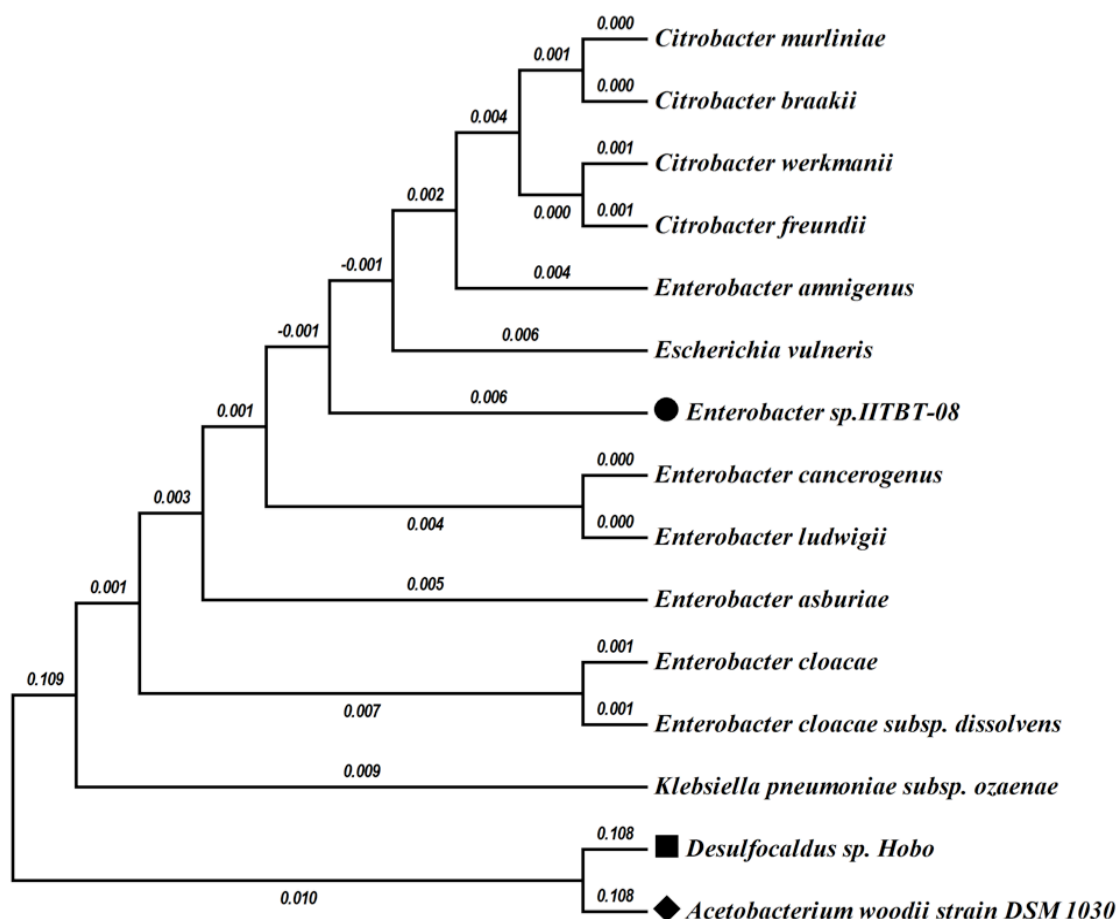


Figure 1. Phylogenetic tree high-lighting the position of “*Enterobacter sp.* IIT-BT 08 (●)” relative to other type and non-type strains within the *Enterobacteriaceae*. Strains shown are those within the *Enterobacteriaceae* having corresponding NCBI genome project ids. The tree was constructed using Mega4 software. The tree based on Jukes–Cantor distance was constructed using neighbor-joining algorithm with 1,000 bootstrapping. *Acetobacterium woodii* strain DSM 1030 (◆) and *Desulfocaldus sp.* (■) was considered as the out group. The scale bar represents 0.1 substitutions per nucleotide position. Numbers at the nodes are the bootstrap values.

Table 2. Genome sequencing project information

| MIGS ID | Property | Term |
|-----------|-------------------------|--|
| MIGS-31 | Finishing quality | High-quality draft |
| MIGS-28 | Libraries used | IGHT and IGGH |
| MIGS-29 | Sequencing platforms | Illumina, 454 |
| MIGS-31.2 | Fold coverage | 50× |
| MIGS-30 | Assemblers | Velvet v. 1.1.05, ALLPATHS v. 39750, Phrap v. 4.24 |
| MIGS-32 | Gene calling method | Gene Prodigal |
| | Genome Database release | September 6 th , 2012 |
| | NCBI ID | 1070842 |
| | Genbank Date of Release | Not determined |
| | GOLD ID | Gi12106 |
| | Project relevance | Biohydrogen production |

Genome annotation

Genes were identified using Prodigal [27] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [28]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation were performed within the Integrated Microbial Genomes Expert Review (IMG-ER) platform [29].

Genome properties

The genome of *E. sp.* IIT-BT 08 consists of one linear chromosome of 4,672,040 bp (Figure 2). The average G+C content for the genome is 56.01% (Table 3). There are 78 tRNA genes and 6 rRNA operons each consisting of a 16S, 23S, and 5S rRNA gene. There are 4,393 predicted protein-coding regions and 43 pseudogenes in the genome. A total of 3,881 protein-coding genes (85.64%) have been assigned a predicted function while the rest have been designated as hypothetical proteins (Table 4). The numbers of genes assigned to each COG functional category are listed

in Table 4. About 2% of the annotated genes were not assigned to COGs and have an unknown function.

Biohydrogen production pathway

The complete genome sequencing of the organism helps provide a preliminary idea of the genes involved in the hydrogen production pathway. The genome revealed the presence of formate hydrogen lyase (EntIIITBT8_2511) and its maturation operons HycH (EntIIITBT8_2678), NiFe hydrogenase III small and large subunit (EntIIITBT8_2679, EntIIITBT8_2681), their maturation operons and the FeS cluster containing hydrogenase components 1 and 2 (EntIIITBT8_0331, EntIIITBT8_2684). A complete list of all the genes predicted to be involved in the hydrogen production pathway is listed in Table 5. The whole genome information of the organism suggests that hydrogen production in *Enterobacter sp.* IIT-BT 08 is carried out through the formate hydrogen lyase (FHL) complex which consists of formate dehydrogenase (FDH-H), hydrogenase (Hyd-3) and the electron transfer mediators [30].

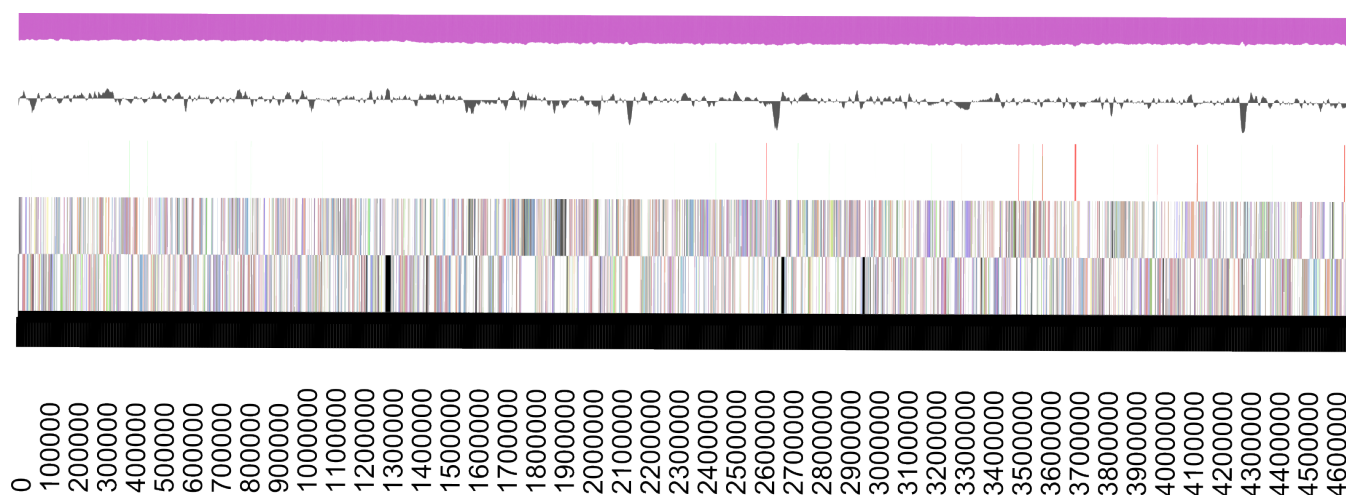


Figure 2. Graphical linear map of the genome. From left to right: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Nucleotide content and gene count levels of the genome

| Attribute | Value | % of total |
|----------------------------------|-----------|------------|
| Genome size (bp) | 4,672,040 | 100.00 |
| DNA coding region (bp) | 4220,082 | 90.33 |
| DNA G+C content (bp) | 2,616,273 | 56.01 |
| Total genes | 4,532 | 100.00 |
| RNA genes | 179 | 3.95 |
| Protein-coding genes | 4,393 | 96.05 |
| Genes in paralog clusters | 1,665 | 36.74 |
| Genes assigned to COGs | 3,780 | 83.41 |
| Genes assigned Pfam domains | 3,949 | 87.14 |
| Genes assigned TIGRfam domains | 1,715 | 37.84 |
| Genes with signal peptides | 1,603 | 35.37 |
| Genes with transmembrane helices | 1,101 | 24.29 |
| Pseudo Genes ^a | 43 | 0.95 |

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

b) Pseudogenes may also be counted as protein coding or RNA genes, so is not additive under total gene count.

Table 4. Number of genes associated with the general COG functional categories

| Code | Value | %age | Description |
|------|-------|------|--|
| J | 191 | 4.4 | Translation |
| A | 1 | 0.0 | RNA processing and modification |
| K | 372 | 8.5 | Transcription |
| L | 153 | 3.5 | Replication, recombination and repair |
| B | 0 | 0.0 | Chromatin structure and dynamics |
| D | 32 | 0.7 | Cell cycle control, mitosis and meiosis |
| Y | 0 | 0.0 | Nuclear structure |
| V | 56 | 1.3 | Defense mechanisms |
| T | 220 | 5.1 | Signal transduction mechanisms |
| M | 249 | 5.7 | Cell wall/membrane biogenesis |
| N | 148 | 3.4 | Cell motility |
| Z | 0 | 0.0 | Cytoskeleton |
| W | 0 | 0.0 | Extracellular structures |
| U | 155 | 3.6 | Intracellular trafficking and secretion |
| O | 146 | 3.4 | Posttranslational modification, protein turnover, chaperones |
| C | 244 | 5.6 | Energy production and conversion |
| G | 396 | 9.1 | Carbohydrate transport and metabolism |
| E | 402 | 9.2 | Amino acid transport and metabolism |
| F | 83 | 1.9 | Nucleotide transport and metabolism |
| H | 165 | 3.8 | Coenzyme transport and metabolism |
| I | 105 | 2.4 | Lipid transport and metabolism |
| P | 246 | 5.7 | Inorganic ion transport and metabolism |
| Q | 80 | 1.8 | Secondary metabolites biosynthesis, transport and catabolism |
| R | 446 | 10.2 | General function prediction only |
| S | 375 | 8.6 | Function unknown |
| - | 88 | 2.0 | Not in COGs |

^aThe total is based on the total number of protein coding genes in the entire annotated genome

Table 5. Preliminary genes involved in the hydrogen production pathway according to the MIGS recommendations [8]

| Protein product | Gene locus | Molecular mass (kDa) | Description |
|-----------------------|--|----------------------|---|
| FdhD | EntlITBT8DRAFT_3548 | 29.9 | formate dehydrogenase family accessory protein (protease) |
| FdhE | EntlITBT8DRAFT_3551 | 89.4 | formate dehydrogenase, alpha subunit, proteobacterial-type [EC:1.2.1.2] |
| FdhF | EntlITBT8DRAFT_1016 | 79.3 | formate dehydrogenase, alpha subunit, archaeal-type (EC:1.2.1.2) |
| HycA | EntlITBT8DRAFT_2685 | 17.5 | Transcriptional repressor of hyc and hyp operons |
| HycD | EntlITBT8DRAFT_2682 | 33.0 | Formate hydrogenlyase subunit 4 |
| HycE | EntlITBT8DRAFT_2681 | 64.8 | Ni, Fe-hydrogenase III large subunit |
| HycG | EntlITBT8DRAFT_2679 | 28.0 | Ni,Fe-hydrogenase III small subunit |
| HycH | EntlITBT8DRAFT_2678 | 15.2 | Formate hydrogenlyase maturation protein |
| Hycl | EntlITBT8DRAFT_2677 | 16.2 | hydrogenase maturation protease [EC:3.4.23.51] |
| ? | EntlITBT8DRAFT_2511 EntlITBT8DRAFT_2680 | 23.0 | Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase subunit (chain I) |
| ? | EntlITBT8DRAFT_2683 | 63.5 | Formate hydrogenlyase subunit 3/Multisubunit Na ⁺ /H ⁺ antiporter, MnhD subunit |
| HycB?? | EntlITBT8DRAFT_2684 | 21.7 | Fe-S-cluster-containing hydrogenase components 2 |
| HypA | EntlITBT8DRAFT_2686 | 13.1 | hydrogenase nickel insertion protein |
| HypB | EntlITBT8DRAFT_2687 | 31.1 | hydrogenase accessory protein |
| HypC/HupF | EntlITBT8DRAFT_2688 | 9.65 | hydrogenase assembly chaperone HypC/HupF |
| HypD | EntlITBT8DRAFT_2689 | 41.1 | hydrogenase expression/formation protein HypD |
| HypE | EntlITBT8DRAFT_2690 | 35.3 | hydrogenase expression/formation protein HypE |
| HypF | EntlITBT8DRAFT_2672 | 80.5 | [NiFe] hydrogenase maturation protein HypF |
| HoxN/HupN/NixA family | EntlITBT8DRAFT_2671 | 36.7 | high-affinity nickel-transporter |

However, in the future the hypothetical pathway must be verified with wet lab experiments. Based on the previous reported literature it may be that formate dehydrogenase and hydrogenase 3 together form a membrane protein complex that is responsible for hydrogen production in facultative anaerobes [30-32]. Rossmann et al. suggested that in facultative anaerobes hydrogen production was determined by the concentration of formate in the cell, which in turn determined the formation of the FHL complex [32]. A putative model (Figure 3) has been suggested based on the biochemistry of the reactions involved in the pathway [34]. Formate dehydrogenase is suggested to catalyze the oxidation

of formate into carbon dioxide. The electrons released in the process are transferred to Hyd3 encoded by hycABCDEFGH to generate molecular hydrogen under anaerobic conditions [33]. The model suggests a plausible scheme of electron transfer from FdhF to the catalytic subunit of hycE via hycBCFG subunits. Among these, hycB and hycF have been determined to be [4Fe-4S] ferredoxin type electron transfer proteins [35]. On the other hand, hycE and hycG shares homology with NADH ubiquinone oxidoreductase (NUO) subunits of the mitochondria and chloroplast [35]. In the model, hycC and hycD have been suggested to act as transmembrane proteins.

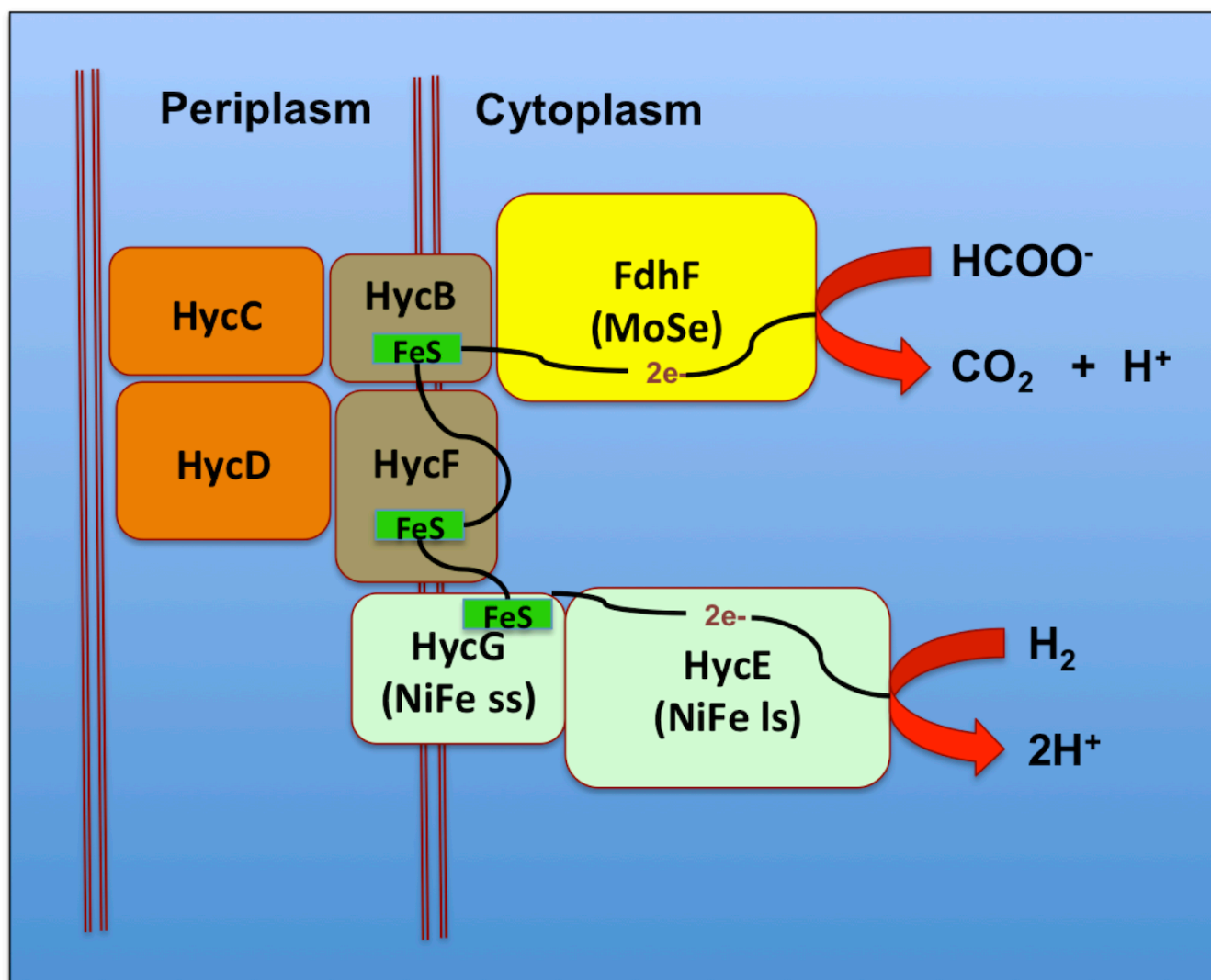


Figure 3. Putative mechanism of hydrogen production by *Enterobacter sp.* IIT-BT 08 based on the genes identified in the genome. Figure is adapted from [35].

Electron acceptors, like oxygen or nitrate, generally inhibit the expression of the FHL complex, whereas its biosynthesis is controlled by the concentration of formate in the cell [32]. Further, it has been suggested that the micro elements selenium and molybdenum are involved at the active site of FDH-H, while nickel is a component of the Hyd-3 active site [30,36]. Accordingly, it has been suggested that the FHL complex can be induced by regulating the presence of formate and metal ions in slightly acidic pH under anaerobic conditions.

Transcription of the FHL complex is under the control of several genes, including *fhlA*, which codes for the FHL activator protein FHLA, a tetramer that binds to the upstream region of the DNA encoding the FHL complex and promotes

the transcription of the FHL complex [34,37]. Moreover, *hycA* codes for the FHL repressor protein that binds to FHLA or to the FHLA-formate complex. Since *fhlA* and *hycA* control the transcription of the FHL complex, it is theoretically possible to control the specific FHL activity and the specific hydrogen production rate by manipulating these genes or their genetic controls [38].

Conclusion

The genome of *Enterobacter sp.* IIT-BT 08 was sequenced and annotated by the DOE Joint Genome Institute. The genomic properties of the organism were analyzed using various IMG tools, and, based on the genome sequence, a putative pathway of hydrogen production based on formate hydrogen lyase complex was discussed.

Acknowledgement

The work was conducted by the U.S. Department of Energy Joint Genome Institute and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Authors (DD and NK) are also thankful to MNRE for their financial assistance. NK also gratefully acknowledges Department of Biotechnology (DBT), Government of India, for senior research fellowship. The authors from

IITKgp, India submitted the JGI-CSP project, analyzed the data and wrote the manuscript. The authors from DSMZ confirmed the strain identity and extracted high quality genomic DNA for sequencing. The authors from DoE-JGI, WC, USA, and LLNL, Livermore CA USA carried out the sequencing and annotation of the genome.

References

1. Das D, Khanna N, Veziroglu TN. Recent developments in biohydrogen production processes. *Chem Indus Chem Eng Quaterly* 2008; **14**:57-67. <http://dx.doi.org/10.2298/CICEQ0802057D>
2. Nandi R, Sengupta S. Microbial production of hydrogen: an overview. *Crit Rev Microbiol* 1998; **24**:61-84. <http://dx.doi.org/10.1080/10408419891294181>
3. Kumar N, Monga PS, Biswas AK, Das D. Modeling and simulation of clean fuel production by *Enterobacter cloacae* IIT-BT 08. *Int J Hydrogen Energy* 2000; **25**:945-952. [http://dx.doi.org/10.1016/S0360-3199\(00\)00017-3](http://dx.doi.org/10.1016/S0360-3199(00)00017-3)
4. Kumar N, Das D. Enhancement of hydrogen production by *Enterobacter cloacae* IIT-BT 08. *Process Biochem* 2000; **35**:589-593. [http://dx.doi.org/10.1016/S0032-9592\(99\)00109-0](http://dx.doi.org/10.1016/S0032-9592(99)00109-0)
5. Das D. Advances in biohydrogen production processes: An approach towards commercialization. *Int J Hydrogen Energy* 2009; **34**:7349-7357. <http://dx.doi.org/10.1016/j.ijhydene.2008.12.013>
6. Kumar N, Das D. Studies on molecular hydrogen production by *Enterobacter cloacae* IIT-BT 08. Paper presented at 9th European Congress on Biotechnology, Brussels, Belgium, 11–15 July, 1999.
7. Kumar N, Das D. The production of pollution free gaseous fuel. Patent application filed. 191/Cal/99, India.
8. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al.* The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. <http://dx.doi.org/10.1038/nbt1360>
9. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. <http://dx.doi.org/10.1073/pnas.87.12.4576>
10. Garrity GM, Bell JA, Lilburn T. Phylum XIV. Proteobacteria phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 2, Part B, Springer, New York, 2005, pp. 1.
11. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. *Int J Syst Evol Microbiol* 2005; **55**:2235-2238. <http://dx.doi.org/10.1099/ijs.0.64108-0>
12. Garrity GM, Bell JA, Lilburn T. Class III. Gammaproteobacteria class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.
13. Williams KP, Kelly DP. Proposal for a new class within the phylum Proteobacteria, Acidithiobacillia classis nov., with the type order Acidithiobacillales, and emended description of the class Gammaproteobacteria. *Int J Syst Evol Microbiol* 2013; **63**:2901-2906. <http://dx.doi.org/10.1099/ijs.0.049270-0>
14. Garrity GM, Holt JG. Taxonomic Outline of the Archaea and Bacteria. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 155-166.
15. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>

16. Rahn O. New principles for the classification of bacteria. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg* 1937; **96**:273-286.
17. Judicial Commission. Conservation of the family name *Enterobacteriaceae*, of the name of the type genus, and designation of the type species OPINION NO. 15. *Int Bull Bacteriol Nomencl Taxon* 1958; **8**:73-74.
18. Hormaeche E, Edwards PR. A proposed genus *Enterobacter*. *Int Bull Bacteriol Nomencl Taxon* 1960; **10**:71-74.
19. Sakazaki R. Genus VII. *Enterobacter* Hormaeche and Edwards 1960, 72; Nom. cons. Opin. 28, Jud. Comm. 1963, 38. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 324-325.
20. OPINION. 28 Rejection of the Bacterial Generic Name *Cloaca* Castellani and Chalmers and Acceptance of *Enterobacter* Hormaeche and Edwards as a Bacterial Generic Name with Type Species *Enterobacter cloacae* (Jordan) Hormaeche and Edwards. *Int Bull Bacteriol Nomencl Taxon* 1963; **13**:38.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. <http://dx.doi.org/10.1038/75556>
22. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. <http://dx.doi.org/10.1517/14622416.5.4.433>
23. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. <http://dx.doi.org/10.1101/gr.074492.107>
24. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:186-194. <http://dx.doi.org/10.1101/gr.8.3.175>
25. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175-189. <http://dx.doi.org/10.1101/gr.8.3.175>
26. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. <http://dx.doi.org/10.1101/gr.8.3.195>
27. Anonymous. Prodigal Prokaryotic Dynamic Programming Gene-finding Algorithm. Oak Ridge National Laboratory and University of Tennessee 2009 <http://compbio.ornl.gov/prodigal>
28. Pati A, Ivanova N, Mikhailova, N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes
29. Markowitz V, Mavromatis K, Ivanova N, Chen IM, Chu K, Kyrpides N. Expert Review of Functional Annotations for Microbial Genomes. *Bioinformatics* 2009; **25**:2271-2278. <http://dx.doi.org/10.1093/bioinformatics/btp393>
30. Sauter M, Böhm R, Böck A. Mutational analysis of the operon (hyc) determining hydrogenase 3 formation in *Escherichia coli*. *Mol Microbiol* 1992; **6**:1523-1532. <http://dx.doi.org/10.1111/j.1365-2958.1992.tb00873.x>
31. Zinoni F, Birkmann A, Stadtman TC, Böck A. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc Natl Acad Sci USA* 1986; **83**:4650-4654. <http://dx.doi.org/10.1073/pnas.83.13.4650>
32. Rossmann RG, Sawers G, Böck A. Mechanism of regulation of the formate-hydrogen lyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Mol Microbiol* 1991; **5**:2807-2814. <http://dx.doi.org/10.1111/j.1365-2958.1991.tb01989.x>
33. Leonhartsberger S, Korsia I, Böck A. A Review The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol* 2002; **4**:269-276.
34. Leonhartsberger S, Korsia I, Böck A. The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol* 2002; **4**:269-276.
35. Böhm R, Sauter M, Böck A. Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenlyase components. *Mol Microbiol* 1990; **4**:231-243. <http://dx.doi.org/10.1111/j.1365-2958.1990.tb00590.x>
36. Jacobi A, Rossmann R, Böck A. The hyp operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in *Escherichia coli*. *Arch Microbiol* 1992; **158**:444-451. <http://dx.doi.org/10.1007/BF00276307>

-
37. Schlensog V, Böck A. Identification and sequence analysis of the gene encoding the transcriptional activator of the formate hydrogenlyase system of *Escherichia coli*. *Mol Microbiol* 1990; **4**:1319-1327. <http://dx.doi.org/10.1111/j.1365-2958.1990.tb00711.x>
38. Yoshida A, Nishimura T, Kawaguchi H, Inui M. Enhanced Hydrogen Production from Formic Acid by Formate Hydrogen Lyase-Overexpressing *Escherichia coli* Strains. *Appl Environ Microbiol* 2005; **71**:6762-6768. <http://dx.doi.org/10.1128/AEM.71.11.6762-6768.2005>