Journal of Applied Biology & Biotechnology Vol. 13(Suppl 1), pp. 97-104, July-August, 2025
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2025.227583

Whole genome sequencing of pollution thriving *E. coli* and its fermentation ability of biomass for bioethanol production

Kathiresan Subramanian[1,2*], Kagne Suresh[3], Neelam Yadav[4], Narinderpal Kaur[5], Samrendra Singh Thakur[6]

[1]Hospira Healthcare Pvt. Ltd., Chennai, Tamil Nadu, India.
[2]Badrinarayan Barwale Mahavidhyalaya, Jalna, Maharashtra, India.
[3]Department of Microbiology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.
[4]Centre for Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India.
[5]Chitkara Centre for Research and Development, Chitkara University, Baddi, Himachal Pradesh, India.
[6]Department of Biotechnology, School of Biological Sciences, Dr Harisingh Gour Vishwavidyalaya (A Central University), Sagar, Madhya Pradesh, India.

ARTICLE INFO

Article history:
Received on: 16/12/2024
Accepted on: 21/04/2025
Available Online: 03/07/2025

Key words:
Bioethanol, *E. coli*, functional analysis, gas chromatography, whole genome sequencing

ABSTRACT

An attempt was made to isolate *Escherichia coli* strains from a polluted environment to explore the potentiality of the production of bioethanol. The whole genome sequencing (WGS) confirmed the strain as *E. coli.* The whole genome provided insight into structural and function annotations and mined into their potent enzymes for ethanol production. The WGS of the *E. coli* strain contributed nearly 23% of total genes involved in carbohydrate metabolism, and the highest Clusters of Orthologous Genes (COGs) were recorded around 447 Carbohydrate transport and metabolism genes. Additionally, *E. coli* enzymes, namely protease, alcohol dehydrogenase, and lyase enzymes, were observed, and each could potentially play a crucial role in ethanol production. Despite their importance in ethanol production, structural information for these enzymes from the microorganism remains unavailable. In the current investigation, genomic data of *E. coli* genome from three sequences were selected. Subsequently, the 3D structure of protease, alcohol dehydrogenase, and lyase enzymes were modeled and validated using structural bioinformatics methodologies. The gas chromatography of the fermented byproducts using this strain was analyzed, and it was seen that 2-butanol had the highest quantification of 31.055%, while ethanol resulted in 7.907%. This study provides the real-world applicability of the wild *E. coli* strain in bioethanol production.

1. INTRODUCTION

Fossil fuels are the backbone of the global energy system; they play a crucial role in the world's economic growth and prosperity; however, due to high consumption, the reservoirs of fossil fuels are continuously depleted. It is predicted that fossil fuel reserves will almost run off at the beginning of the 22nd century [1]. Future energy requirements need a promising and sustainable solution to tackle this global issue. In this context, there is an urgent need for sustainable energy solutions that have driven significant interest in bioethanol production [2,3]. Bioethanol can act as a renewable alternative to fossil fuels. This sustainable solution can counter air pollution and be a sustainable energy source. Recently, *Escherichia coli* have emerged as a promising candidate for various industrial biotechnological applications, including bioethanol production [4]. Due to their rapid growth rate, genetic tractability, and metabolic versatility, *E. coli* can be a potential candidate for biofuel production [5,6]. *Escherichia coli* is traditionally used on laboratory strains and genetic engineering most favorable model [7]. Moreover, the wild strain *E. coli* potentially codes the vast arsenals of enzymes and metabolic diversity they may have inherited.

The wild *E. coli* strains, which have adapted to diverse environmental conditions over millions of years, possess unique enzymes and regulatory networks that can be harnessed in bioethanol production [8]. These strains may offer enhanced activity under industrial conditions due to their ability to sustain high temperatures, low pH, presence of heavy metals, and the presence of toxic compounds, which are often encountered during biomass conversion processes [9,10]. The *Saccharomyces cerevisiae* is widely used for bioethanol production. However, recent studies have shown that *E. coli* can be substituted where it can tolerate high glycerol concentrations (up to 10 g/l), high salt concentrations [11], and low concentrations of heavy metals (<0.1 mM) [12]. Atsumi *et al.* [13] have explored a metabolic engineering approach using *E. coli* for the high-yield production of isobutanol in the fermentation process. Generally, most of the *E. coli* strains cannot be able to convert biomass into bioethanol. However, wild-type strains of *E. coli* may contain various biomass-degrading enzymes. This is due to their enhancement by the genetic transformation that aids them in their natural or artificial

*Corresponding Author
Kathiresan Subramanian, Hospira Healthcare Pvt. Ltd., Chennai, Tamil Nadu, India. E-mail: kathirsubramn@gmail.com

competence to enhance their capacity in utilizing other hydrocarbons or cellulose-based materials [14].

In the present study, whole genome sequencing (WGS) and the functional annotation approach were used to explore the entire genetic repertoire of wild *E. coli* strains for bioethanol production. This approach helps to identify novel enzymes, which may be an essential consideration, as the genes can help degrade complex plant biomass efficiently and further convert sugars to ethanol with improved yields [9,15]. In addition, validating the structure of enzymes identified from wild *E. coli* strains is crucial. These enzymes are used to understand their functional capabilities. Therefore, in this study, AlphaFold, an advanced tool, was used to predict protein structures as the tool helps to predict accurate models and validate the three-dimensional structures of enzymes. Also, this tool reveals critical insights into enzyme functionality, substrate interactions, and catalytic mechanisms, enhancing the understanding of how these enzymes contribute to bioethanol production. Structural validation ensures that the identified enzymes have the desired properties and supports the development of more effective biocatalysts [16].

Furthermore, the integration of antibiotic-resistance genes into *E. coli* strains used for bioethanol production presents additional opportunities and challenges. These genes can enhance the robustness of *E. coli* under harsh industrial conditions, providing resistance to stresses such as high temperatures and toxic inhibitors. Some antibiotic resistance genes encode enzymes with unique catalytic properties, including the ability to degrade complex substrates or detoxify inhibitory compounds, thereby supporting the efficient conversion of biomass to ethanol. Additionally, incorporating these genes can facilitate metabolic engineering efforts, optimizing *E. coli* strains for specific bioethanol production [5,7,8]. However, the application of antibiotic-resistance genes in bioethanol production must be cautiously approached due to the potential risks associated with their spread to pathogenic bacteria. The inadvertent transfer of these genes could contribute to the growing public health threat of antibiotic resistance [17,18]. Therefore, strict containment measures, such as secure laboratory environments and the use of genetically engineered organisms incapable of surviving outside controlled settings, are critical to prevent accidental release [19].

Generally, the *E. coli* strain is not able to convert biomass into bioethanol. Moreover, previous studies showed that *E. coli* and *S. cerevisiae* strains convert biomass into isobutanol [5,12,13]. The wild strains of the *E. coli* strains have the unique enzymes and metabolic potentials to convert biomass into bioethanol. Therefore, the present study aims to explore the potential of wild *E. coli* strains in bioethanol production by integrating WGS and functional annotations to uncover novel enzymes and pathways that enhance biomass degradation and bioethanol yield. By tapping into the natural enzymatic repertoire and metabolic diversity of wild *E. coli* and utilizing AlphaFold for enzyme structure validation, this research seeks to advance the development of more efficient and sustainable bioethanol production processes. The findings of this study could potentially address key challenges associated with large-scale biofuel production, broadening the goal of a sustainable energy future and sparking further research and development in this field.

2. MATERIALS AND METHODS

2.1. Sample Collection and Isolation of Bacteria

The soil samples were collected from the bank of the Cooum River in Chennai (13°05′52.6″N; 80°17′23.78″E). The isolation of bacteria was performed using a serial dilution technique, briefly, soil samples were diluted in MilliQ water, and the resulting dilutions were plated onto MacConkey agar. Individual bacterial colonies were isolated and grown separately to obtain pure cultures. After which the isolates were screened in an enriched lauryl sulphate-aniline blue MacConkey agar medium for coliform bacteria namely *E. coli.*

2.2. Biomass Utilization

Wet biomass (Plant litter) was collected from the Western Ghats (10°10′00.1″N; 77°04′00.1″E). The biomass was tested for its suitability as a growth substrate for the isolated bacteria. The growth was monitored at multiple time points: 3, 6, 9, 12, 18, 24, 32, 48, 56, and 64 hours.

2.3. Whole Genome Sequencing

The bacterial strain that showed the best growth was selected for further study. The Genomic DNA of this strain was isolated and sequenced to obtain 1 gigabyte (Gb) of data using the Illumina NovaSeq 6000 platform using 150 × 2 base pairs (bp) paired-end sequencing chemistry.

2.4. Assembly and Functional Analysis

Genome assembly was performed using the Spades tool v4.0. The annotated assembly was further used for functional analysis using the Prokka tool. This tool facilitated the assembly of the sequencing data and allowed for detailed analysis of the genomic content of the isolated strain. The genome's circular visualization and functional results were visualized using the GenoVi tool.

2.5. Protein Modelling

The Chimera X version 1.4 software package [20,21] utilized the AlphaFold2 tool for 3D structure prediction based on the amino acid sequences. The prediction process involved selecting the AlphaFold2 option within the Tools section of the Structure Prediction tab. Subsequently, the amino acid sequence was entered into the query box and submitted with default parameters. The resulting predicted tertiary structure was then input into the PROCHECK tool [22] to generate the Ramachandran plot, assessing the stereochemical quality of the protein structure for validation. The PrankWeb server, available online at https://prankweb.cz/, was utilized to identify the target proteins' active sites. Subsequently, the topmost binding pocket was chosen for the subsequent docking analysis [23].

2.6. Fermentation

Escherichia coli cultures were aerated using the shake-flask technique. In this experiment, we used a minimal medium containing biomass as substrate. One milliliter of cultures was inoculated on the grown in conical flasks (500 cm^3) on a shaker within a controlled environment chamber. To optimize oxygenation, larger surface areas were employed using T-flasks and roller bottles. After sterilizing the lab-scale fermenter, *E. coli* was inoculated into the sterile culture medium in a batch fermentation process with 25 g of biomass. During fermentation, cell concentrations were monitored by measuring optical density (OD) at 600 nm using ultraviolet-visible spectrophotometer. The concentration of minerals and salts was measured using Inductively Coupled Plasma Optical Emission Spectrometry (ICPOES). Byproducts were monitored using the gas chromatography-mass spectrometry (GC-MS) analysis; a detailed method is described in "Gas Chromatography (GC)" section. The bioethanol was collected after the fermentation

Table 1. Brief information related of alcohols standard used in the present study.

Sl. No.	Name of the compounds	Batch number	Purity (%)
1	Ethanol	F204325	99.9
2	Methanol	3085112913	99.9
3	1-Butanol	FPC/LM/1008	99.8
4	2-Butanol	B0705	99.0
5	1-Propanol	279544	97.8
6	2-Propanol	GUUOM	99.8

process. Continuous aeration was maintained to support aerobic culture and facilitate the removal of gaseous byproducts such as CO_2 [24,25]. The entire fermentation process was run for 82 hours at a constant 35°C temperature.

2.7. Gas Chromatography

The fermented byproducts were analyzed using GC (Shimadzu GC-2014). The analysis was performed with the following conditions: a DB624 column (60 m × 0.32 mm ID 1.8 µm film Capillary column) was used, with a flow rate of 1 ml/min (constant flow) and a total run time of 30 min, hydrogen flow 45 ml/min and air flow 450 ml/min [26,27]. The temperature program settings were as follows; the column was initially held at 40°C for 15 min, then increased to 220°C at the rate of 15°C per min and maintained at 220°C for 3 min. The detector temperature was set at 260°C. The injection mode was split and the injection time was 0.1 min.

The peaks of alcohols were identified by comparing retention time to those of standard reference compounds. The brief information related to the preparation of standards is mentioned in Table 1.

3. RESULTS

3.1. Isolation of Bacteria and WGS

The colony confirmed that coliforms appeared smooth, circular, white, or greyish colonies, primarily confirming the typical characteristic of *E. coli* in the selective medium. The sequencing result confirmed the *E. coli*, which resembled a genome size of around 4.6 million bp. The genome comprised 9,375,256 bp and 102 contigs. In the assembly, the GC content was 51%, the n50 value was 110 kilobase pairs (Kbp), and the genome completeness was 93%. The Genomic map included the number and functions of genes (Fig. 1A). The genes were plotted in the circular genome along with many genes like *arcB* and *evgS*. The CDS/open reading frame (ORF), tRNA, rRNA, GC content, and GC skew were analyzed (Fig. 1A).

3.2. Functional Annotations

The functional annotations were performed for the *E. coli* genome by two approaches. The first approach was performed by matching the annotations with the KEGG (Kyoto Encyclopaedia of Genes and Genomes) database. In the second approach to predict the functions of the genes, we used Clusters of Orthologous Groups database to classify the genes based on clusters of orthologous genes (COGs) id of proteins (Fig. 1B and D). The COGs recorded 41.4% for metabolism-related genes, followed by Cellular Processes and Signalling, which recorded 28.2%. Upon further classification into the clusters, the G: carbohydrate transport and metabolism showed the highest classified genes of 447, followed by E: amino acid transport and metabolism genes of 385 genes. Under information, storage, and processing, K: transcription showed the highest orthologous of 373 genes. Nearly 10.6%, *i.e.*, R: general function prediction only; S: function unknown, recorded 297 and 218 genes, respectively (Fig. 1D).

The metabolic pathways and associated gene counts from the KEGG database for *E. coli* showed that Carbohydrate metabolism had the highest number of genes (334), followed by environmental information processing (328) and Genetic information processing (170) genes, and amino acid metabolism (146) genes were recorded (Fig. 1C). The genes of the *E. coli* involved in pathways, namely pyruvate, and glycolysis, and further glyconeogenesis were also plotted and, their contribution to the enzymes in green color was shown (Fig. 1E and F). The *E. coli* genes with their enzyme ID were plotted in both the pathways involved in ethanol biosynthesis.

3.3. Structure Predictions

Utilizing the amino acid sequence with default parameters, the AlphaFold2 tool successfully predicted the protein structure (Fig. 2). For the protease model, the predicted local distance difference test calculated a mean confidence value of 92.4%, while the alcohol dehydrogenase and lyase models exhibited values of 90.8% and 91.5%, respectively. A mean confidence value of 90 and above is consistent with an experimental structure [16]. The protease-active-site residues are THR-184, TYR-208, LYS-209, SER-210, ALA-211, ASP-228, TRP-231, ALA-336, MET-338, PRO-375, GLY-376, GLY-377, SER-378, VAL-379, MET-406, ALA-408, SER-409, GLY-410, LEU-428, SER-431, GLY-433, ILE-434, PHE-435, PHE-488, and VAL-511. Similarly, the alcohol dehydrogenase -active-site residues comprise CYS-40, HIS-41, SER-42, HIS-45, GLN-46, TRP-51, HIS-62, CYS-88, MET-110, GLY-112, THR-113, TYR-114, ASN-115, SER-116, CYS-158, THR-162, VAL-181, GLY-182, ILE-183, GLY-184, GLY-185, LEU-186, PHE-204, THR-205, THR-206, LYS-210, SER-225, THR-244, VAL-245, ALA-246, VAL-267, GLY-268, ALA-269, MET-293, ILE-294, MET-331, GLY-334, VAL-336, TYR-338, and ARG-339. Additionally, the lyase-active-site residues include SER-71, MET-72, ARG-114, ILE-116, LEU-118, LEU-126, GLY-128, THR-130, PRO-151, LEU-152, LEU-191, and GLU-193 (Fig. 2A–F).

3.4. Structure Validation

The evaluation of the modeled protein structure was done using the Ramachandran plot in the PROCHECK tool (Fig. 2G–I). In the protease model, 92.8% of residues were situated in the most favored regions (A, B, and L), 7.2% in the additional allowed regions (a, b, l, and p), and 0% in the generously allowed regions (~a, ~b, ~l, and ~p), with no residues in the disallowed regions (Fig. 2G). Similarly, the alcohol dehydrogenase model exhibited 92.9% of residues in the most favored regions, 7.1% in the additional allowed regions, 0% in the generously allowed regions, and 0.7% in the disallowed regions (Fig. 2H). The lyase model displayed 89.7% of residues in the most favored regions, 8.9% in the additional allowed regions, 0.7% in the generously allowed regions, and no residues in the disallowed regions (Fig. 2I). Consequently, the distribution of amino acid residues indicates that the predicted cellulase structure possesses high-quality characteristics.

3.5. Fermentation and GC Chromatogram Estimation of Bioethanol

Using GC, the byproducts of fermentation were analyzed, with each chromatogram representing a different compound. Ethanol production was 7.9%, and 2-butanol was the highest byproduct, at nearly 31.055%

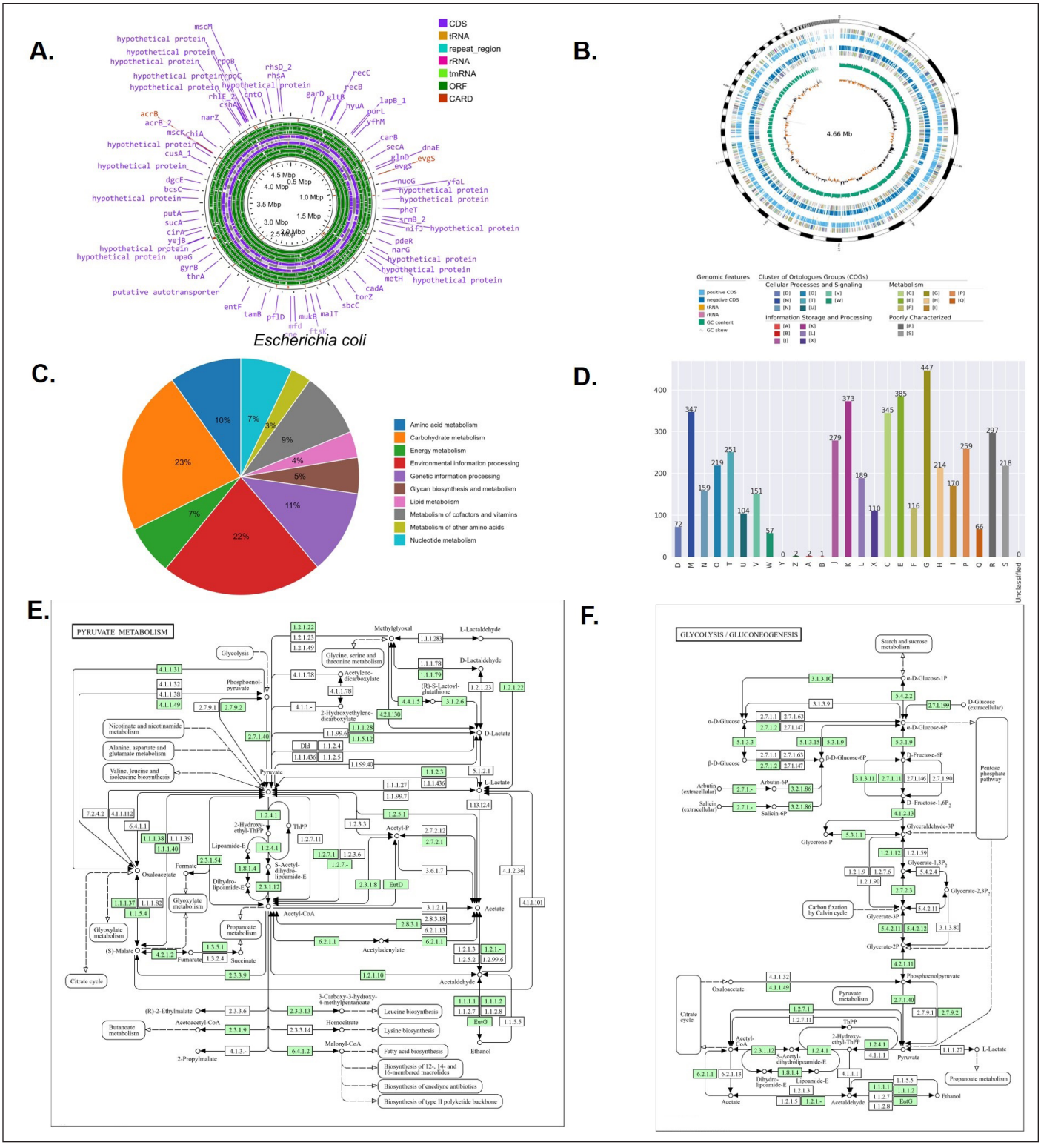

Figure 1. *Escherichia coli* (A): Whole genome assembly and the gene annotations representation and (B): annotations of cluster of orthologues groups; (C): Pie chart representing KEGG pathway enrichment analysis and (D): COG functional distributions of genes :-A: RNA processing and modifications, B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, and chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure, and biogenesis; K: Transcription; L: Replication, recombination, and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport, and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; X: Mobilome: prophages, transposons; Z: Cytoskeleton). KEGG Pathway contribution of genes by *E. coli*. (E): Pathway for Pyruvate metabolism. (F): Glycolysis and gluconeogenesis. The green boxes are the genes present in the *E. coli* genome contributing in the pathways.

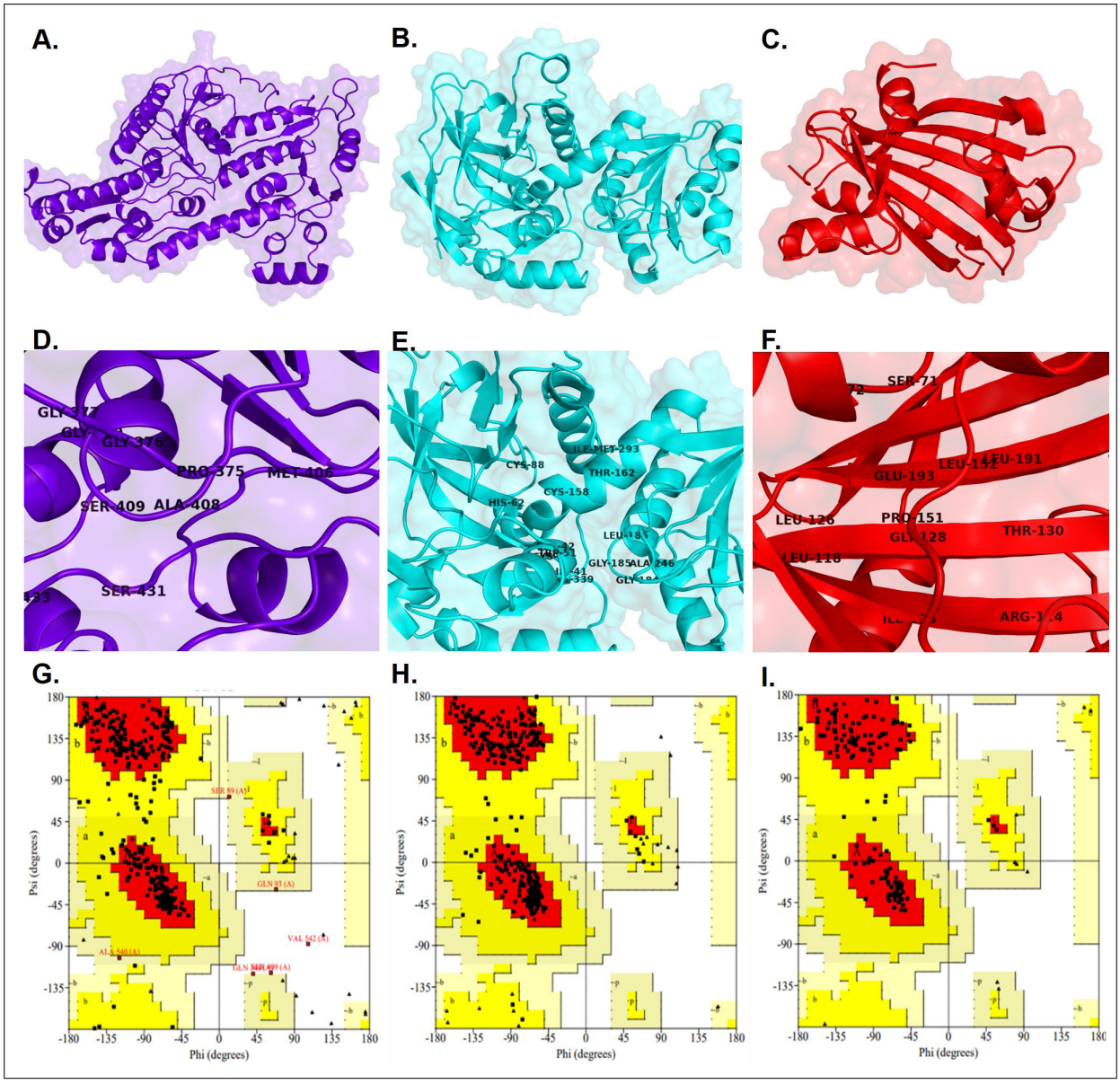

Figure 2. Prediction of a protein structure (A, B, C) and their active sites (D, E, F) of *E. coli* for protease (A & D), Alcohol dehydrogenase (B & E) and Lyase (C & F), respectively. Ramachandran plot in the PROCHECK tool (G) protease, (H) Alcohol dehydrogenase, and (I) lyase.

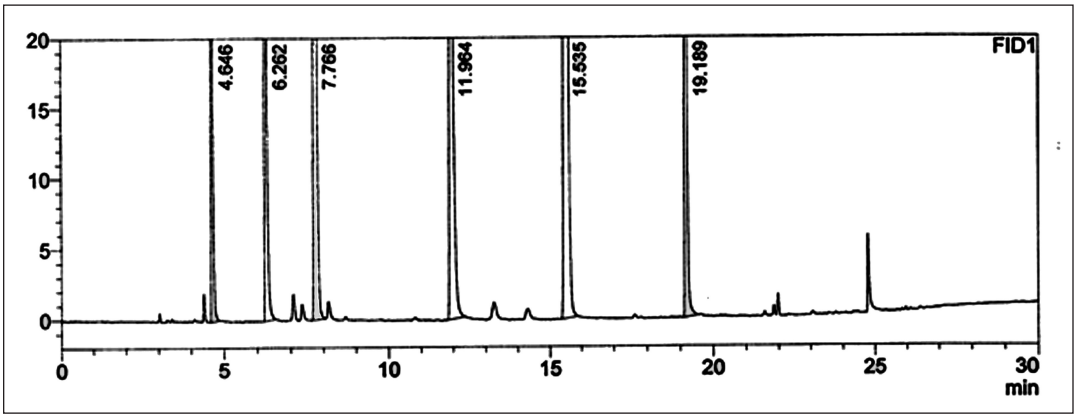

Figure 3. Result of GC, the *X*-axis showing the retention time while *Y*-axis showing the mass m/z of the peak. Where *m* is mass and *z* are charge.

Table 2. Peak table of fermented byproducts from GC.

Peak	Retention time	Area	Area%	Name of byproduct
1	4.646	126,045	4.445	Methanol
2	6.262	224,229	7.907	Ethanol
3	7.766	484,997	17.102	2-Propanol
4	11.964	423,721	14.941	1-Propanol
5	15.535	880,708	31.055	2-Butanol
6	19.189	696,257	24.551	1-Butanol
Total		2,835,956	100.00	--

(Fig. 3; Table 2). The other products are also mentioned in Table 2, with their retention time and area percentage.

4. DISCUSSION

This study aimed to isolate *an E. coli* bacterium using a selective medium from a harshly polluted area. *Escherichia coli* may have a more natural transformation of various genes for survival. Natural transformation allows for the integration of extra-chromosomal DNA, which can enter the cell through membrane pores and is influenced by environmental factors such as heat shock. Usually, *E. coli* does not utilize cellulose, as it primarily originates from human environments. However, wild-type *E. coli* can utilize cellulose and other sugars from wet plant biomass when adapted to become competent for various substrates [14]. These isolates are tolerant and capable of utilizing different sugars and can convert renewable biomass into biofuel by degrading the plant litter. Therefore, a WGS of *E. coli* was performed, and its structural and function genes were studied. The protein structure of these genes into enzymes and evaluating their active site and completeness were studied. After confirmation, we fermented plant litter using *E. coli* for bioethanol production, and the quantity was calculated using GC.

It was seen that *E. coli* K-12 reported a genome consisting of 4,639,221 bases of circular duplex DNA [28]. While in the present study reported *E. coli* strain also reported 4,658,037 bases represented in circular duplex DNA with positive and negative ORF. The previous study based on the functional potential of *E. coli* K-12 strain identified genes related to transport and binding proteins (6.55%) and energy synthesis (5.67%) [28]. In the present investigation, it was carbohydrate metabolism (23%) and environmental information processing (22%). This proves that our *E. coli* strain is different from *E. coli* K1, which involves the degradation of various sugars and has a wide range of adaptation capabilities in harsh environments. The COGs namely, carbohydrate transport and metabolism, recorded the highest group of proteins with 447 genes. The organism possesses a range of enzymes, including alcohol dehydrogenase, protease, and lyase, that are essential for ethanol production. However, these enzymes' lack of structural information has impeded a comprehensive understanding of their metabolic functions. To address this, computational tools and comparative genomics can be employed to elucidate genome structure and identify protein isoforms at the gene level, providing a more complete depiction of the organism's metabolic machinery. Moreover, some of the mutants of *E. coli* strains produce higher amounts of alcohol dehydrogenase enzyme, allowing them to use ethanol as a carbon source or to meet acetate needs [14,29]. The active sites of the protease, alcohol dehydrogenase, and lyase enzymes play an important role in the various enzymatic processes such as catalytic cleavage of peptide bonds, cleavage, and formation of chemical bonds without hydrolysis or oxidation. These processes are directly involved in biomass conversion [30]. In the present study, several active sites of the three enzymes were recorded in wild *E. coli* strains genome further confirms the applicability of this strain in bioethanol production.

Several studies have reported to utilizing *E. coli* [4] and *S. cerevisiae* [31] for bioethanol production using various lignocellulosic feedstocks for partial degradation substrate degradation. For example, *E. coli* B pLOI297 was used to ferment aspen, which was reported to have undergone a pretreatment process with SO_2 [32]. This fermentation method resulted in an ethanol concentration of 31 g/l, yielding of 0.26 g of ethanol per gram of sugar consumed [33]. In another study, *E. coli* KO11 fermented maple biomass, pretreated with 1% sulfuric acid at 160°C for 10 min. After the pretreatment, over liming was used, resulting in an ethanol concentration of 84 g/l and a yield of 0.4 g ethanol per gram sugar [34]. In the present study, the strain has produced 7.9% (equal to 291 g/l) of ethanol and confirmed the ability to produce bioethanol from 1 kg of biomass (calculated).

In addition, *S. cerevisiae* strains have been applied to similar processes. *Saccharomyces cerevisiae* MTCC-36, together with *Pichia stipitis* NCIM-3498, fermented *Shorea robusta* (sal) biomass, pretreated with 1% hydrochloric acid at 121°C for 30 min. This fermentation yielded 19.1 g/l of ethanol, with a high yield of 0.39 g ethanol per gram of sugar [35]. Another study that involved *S. cerevisiae* MEC1133 has fermented paulownia wood. It was treated through autohydrolysis and subsequent sulfuric acid pretreatment and resulted in 47.6 g/l of ethanol with a yield of 0.26 g ethanol per g sugar [36]. Though *S. cerevisiae* is wildly used for bioethanol production, recent studies have shown that *E. coli* may be more suitable in conditions where it can tolerate high glycerol concentrations (up to 10 g/l), high salt concentrations [11], and low concentrations of heavy metals (<0.1 mM) [12,37].

The present study is based on the WGS of the *E. coli* strain, regarding its functional potential to convert plant biomass to bioethanol. The computational and bioinformatic analysis were used to find the functional applicability of this strain in bioethanol production. Despite the high-end bioinformatic analysis used in the present study, the work has certain limitations. Notably, the structural prediction of the proteins and the enzymes was entirely based on bioinformatic analysis, with no validation of wet lab conditions. However, the GC-MS-based experiment, which was based on the wet lab work, showed that the byproducts of the fermentation were several alcohols, including butanol, ethanol, methanol, and propanol. Although the study focused on the ethanol production, we found that butanol was the main byproduct. The conversion of the ethanol to butanol is easier; however, the conversion of butanol to bioethanol is a complex process, and there are no known industrial-scale processes currently for the conversion of butanol to ethanol. Therefore, future studies should be directed to find out the pathways to convert butanol to ethanol at a larger industrial scale. The present work has its own limitations related to the proper correlation between enzymes and protein data with GC-MS data, specifically based on wet laboratory work. Therefore, future work should incorporate experimental validation with proper correlation for the real-world application of this strain in bioethanol production.

5. CONCLUSION

An attempt was made to isolate *E. coli* strains from the natural environment to explore the potentiality of the production of bioethanol. It was also evident from the WGS that the *E. coli* strain contributes nearly 23% of the total genes involved in carbohydrate metabolism. Therefore, they can metabolize sugars from biomass and contribute to bioethanol production. The GC (Shimadzu GC-2014) of the fermented byproducts was analyzed, and it was seen that ethanol resulted in 7.907% on the pilot scale for initial confirmation. Further, this strain

can be used for gene knockout studies to enhance the robustness of bioethanol productivity.

6. AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be an author as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

7. FUNDING

There is no funding to report.

8. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

9. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

10. PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of the publisher, the editors and the reviewers. This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

11. USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declares that they have not used artificial intelligence (AI)-tools for writing and editing of the manuscript, and no images were manipulated using AI.

12. DATA AVAILABILITY

The raw sequence of the present study was deposited into the NCBI sequence Read archive (SRA) database under the BioProject accession number PRJNA1138183.

REFERENCES

1. Holechek JL, Geli HM, Sawalhah MN, Valdez R. A global assessment: can renewable energy replace fossil fuels by 2050? Sustainability 2022;14(8):4792; doi: https://doi.org/10.3390/su14084792
2. Aditiya HB, Mahlia TMI, Chong WT, Nur H, Sebayang AH. Second generation bioethanol production: a critical review. Renew Sust Energ Rev 2016;66:631–53; doi: https://doi.org/10.1016/j.rser.2016.07.015
3. McMillan JD. Bioethanol production: status and prospects. Renew Ener 1997;10(2–3):295–302; doi: https://doi.org/10.1016/0960-1481(96)00081-X
4. Koppolu V, Vasigala VKR. Role of *Escherichia coli* in biofuel production. Microbiol Insights 2016;9:MBI-S10878; doi: https://doi.org/10.4137/MBI.S108
5. Ingram LO, Conway T, Clark DP, Sewell GW, Preston J. Genetic engineering of ethanol production in *Escherichia coli*. Appl Environ Microbiol 1987;53(10):2420–5; doi: https://doi.org/10.1128/aem.53.10.2420-2425.1987
6. Yomano LP, York SW, Ingram LO. Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production. J Ind Microbiol Biotechnol 1998;20(2):132–8; doi: https://doi.org/10.1038/sj.jim.2900496
7. Yomano LP, York SW, Zhou S, Shanmugam KT, Ingram LO. Re-engineering *Escherichia coli* for ethanol production. Biotechnol Lett 2008;30:2097–103; doi: https://doi.org/10.1007/s10529-008-9821-3
8. Chang D, Islam ZU, Zheng J, Zhao J, Cui X, Yu Z. Inhibitor tolerance and bioethanol fermentability of levoglucosan-utilizing *Escherichia coli* were enhanced by overexpression of stress-responsive gene *ycfR*: the proteomics-guided metabolic engineering. Synth Syst Biotechnol 2021;6(4):384–95; doi: https://doi.org/10.1016/j.synbio.2021.11.003
9. Horinouchi T, Tamaoka K, Furusawa C, Ono N, Suzuki S, Hirasawa T, *et al*. Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress. BMC Genomics 2010;11:1–11; doi: https://doi.org/10.1186/1471-2164-11-579
10. Munjal N, Mattam A, Pramanik D, Srivastava P, Yazdani SS. Modulation of endogenous pathways enhances bioethanol yield and productivity in *Escherichia coli*. Microb Cell Fact 2012;11:1–12; doi: https://doi.org/10.1186/1475-2859-11-145
11. Adnan NAA, Suhaimi SN, Abd-Aziz S, Hassan MA, Phang LY. Optimization of bioethanol production from glycerol by *Escherichia coli SS1*. Renew Energ 2014;66:625–33; doi: https://doi.org/10.1016/j.renene.2013.12.032
12. Bautista VM, Hernández SC, González LR, Jiménez LD. *Saccharomyces cerevisiae* vs *Escherichia coli* in the valorization of crude glycerol to produce ethanol. Bioresour Technol Rep 2023;24:101634; doi: https://doi.org/10.1016/j.biteb.2023.101634
13. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 2008;451(7174):86–9; doi: https://doi.org/10.1038/nature06450
14. Ohmiya K, Sakka K, Kimura T. Anaerobic bacterial degradation for the effective utilization of biomass. Biotechnol Bioprocess Eng 2005;10:482–93; doi: https://doi.org/10.1007/BF02932282
15. Wang J, Chen L, Tian X, Gao L, Niu X, Shi M, *et al*. Global metabolomic and network analysis of *Escherichia coli* responses to exogenous biofuels. J Proteome Res 2013;12(11):5302–12; doi: https://doi.org/10.1021/pr400640u
16. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873):583–9; doi: https://doi.org/10.1038/s41586-021-03819-2
17. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol 2019;65(1):34–44; doi: https://doi.org/10.1139/cjm-2018-0275
18. Levy SB. Antibiotic resistance-the problem intensifies. Adv Drug Deliv Rev 2005;57(10):1446–50; doi: https://doi.org/10.1016/j.addr.2005.04.001
19. Askari-khorasgani O, Pessarakli M. Safety assessment of genetically modified crops for yield increase and resistance to both biotic and abiotic stresses and their impact on human and environment. Adv Plants Agric Res 2018;8(2):109–12; doi: https://doi.org/10.15406/apar.2018.08.00300
20. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, *et al*. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 2021;30(1):70–82; doi: https://doi.org/10.1002/pro.3943
21. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, *et al*. UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci 2018;27(1):14–25; doi: https://doi.org/10.1002/pro.3235
22. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26(2):283–91; doi: https://doi.org/10.1107/S0021889892009944
23. Jendele L, Krivak R, Skoda P, Novotny M, Hoksza D. PrankWeb: a web server for ligand binding site prediction and visualization.

Nucleic Acids Res 2019;47(W1):W345–9; doi: https://doi.org/10.1093/nar/gkz424
24. Lawford HG, Rousseau JD. Ethanol production by recombinant *Escherichia coli* carrying genes from *Zymomonas mobilis*. Appl Biochem Biotechnol 1991;28:221–36; doi: https://doi.org/10.1007/BF02922603
25. Rao K, Chaudhari V, Varanasi S, Kim DS. Enhanced ethanol fermentation of brewery wastewater using the genetically modified strain *E. coli KO11*. Appl Microbiol Biotechnol 2007;74:50–60; doi: https://doi.org/10.1007/s00253-006-0643-8
26. Lindsay SE, Bothast RJ, Ingram LO. Improved strains of recombinant *Escherichia coli* for ethanol production from sugar mixtures. Appl Microbiol Biotechnol 1995;43:70–5; doi: https://doi.org/10.1007/BF00170625
27. Dombek KM, Ingram LO. Determination of the intracellular concentration of ethanol in *Saccharomyces cerevisiae* during fermentation. Appl Environ Microbiol 1986;51(1):197–200; doi: https://doi.org/10.1128/aem.51.1.197-200.1986
28. Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, *et al*. The complete genome sequence of *Escherichia coli K-12*. Science 1997;277(5331):1453–62; doi: https://doi.org/10.1126/science.277.5331.1453
29. Clark D, Cronan Jr JE. *Escherichia coli* mutants with altered control of alcohol dehydrogenase and nitrate reductase. J Bacteriol 1980;141(1):177–83; doi: https://doi.org/10.1128/jb.141.1.177-183.1980
30. Wang KY, Zhang J, Hsu YC, Lin H, Han Z, Pang J, *et al.* Bioinspired framework catalysts: from enzyme immobilization to biomimetic catalysis. Chem Rev 2023;123(9):5347–420; doi: https://doi.org/10.1021/acs.chemrev.2c00879
31. Buijs NA, Siewers V, Nielsen J. Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. Curr Opin Chem Biol 2013;17(3):480–8; doi: https://doi.org/10.1016/j.cbpa.2013.03.036
32. Lawford HG, Rousseau JD. Loss of ethanologenicity in *Escherichia coli* B recombinants pLOI297 and KO11 during growth in the absence of antibiotics. Biotechnol Lett 1995;17:751–6; doi: https://doi.org/10.1007/BF00130363
33. Lawford HG, Rousseau JD. Mannose fermentation by an ethanologenic recombinant *Escherichia coli*. Biotechnol Lett 1993;15:615–20; doi: https://doi.org/10.1007/BF00138551
34. Okuda N, Ninomiya K, Takao M, Katakura Y, Shioya S. Microaeration enhances productivity of bioethanol from hydrolysate of waste house wood using ethanologenic *Escherichia coli* KO11. J Biosci Bioeng 2007;103(4):350–7; doi: https://doi.org/10.1263/jbb.103.350
35. Raina N, Slathia PS, Sharma P. Response surface methodology (RSM) for optimization of thermochemical pretreatment method and enzymatic hydrolysis of deodar sawdust (DS) for bioethanol production using separate hydrolysis and co-fermentation (SHCF). Biomass Convers Biorefin 2020;12: 5175–95; doi: https://doi.org/10.1007/ s13399-020-00970-0
36. Domínguez E, del Río PG, Romaní A, Garrote G, Domingues L. Hemicellulosic bioethanol production from fast-growing Paulownia biomass. Processes 2021;9(1):173; doi: https://doi.org/10.3390/pr9010173
37. Vargas-Bautista C, Rahlwes K, Straight P. Bacterial competition reveals differential regulation of the pks genes by *Bacillus subtilis*. J Bacteriol 2014;196(4):717–28; doi: https://doi.org/10.1128/jb.01022-13

How to cite this article:
Subramanian K, Suresh K, Yadav N, Kaur N, Thakur SS. Whole genome sequencing of pollution thriving *E. coli* and its fermentation ability of biomass for bioethanol production. J Appl Biol Biotech. 2025;13(Suppl 1): 97–104. DOI: 10.7324/JABB.2025.227583.