Distribution of dye-decolorizing peroxidase (DyP) genes among bacteria based on the NCBI database

Irfan Mustafa[1*], *Suharjono*[1], *Tri* Ardyati[1] and *Yoga Dwi* Jatmiko[1]

[1]Biology Department, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Jalan Veteran, Malang, Indonesia

Abstract. Synthetic textile dyes, such as anthraquinones, are durable and persistent to microbial degradation, making them toxic to the environment. Certain bacteria have been reported to possess dye-decolorizing peroxidase (DyP) enzymes, which show potential for degrading anthraquinones. The isolation of local bacteria with DyP genes is crucial to addressing this issue in Indonesia. Therefore, to increase the likelihood of obtaining these potential bacteria, it is essential to gather information on the distribution of DyP genes across different bacterial species reported to date. This study was conducted in silico to investigate the distribution of DyP-coding genes in various bacterial species using the NCBI database. The results revealed that DyP enzymes are mostly found in members of the phylum Actinobacteria compared to those in the phyla Firmicutes and Proteobacteria. These findings serve as a foundational reference for guiding methods in the isolation of dye-degrading bacteria possessing DyP enzymes.

1 Introduction

Dyes are an integral part of the fashion and textile industries. However, during the dyeing process, not all dye molecules bind to the fabric. Approximately 20% of the applied dyes are released into the environment through wastewater due to inefficiencies in the dyeing process [1]. Moreover, to meet market demands for high-quality textile coloration, dyes are synthesized with properties that ensure stability against exposure to light, temperature, water, detergents, and bleaching agents, making them resistant to biological degradation. This situation leads to environmental pollution, as such dyes also exhibit toxic properties [2].

The successful decomposition of recalcitrant compounds relies on the ability of microorganisms to produce oxidative enzymes. Peroxidase enzymes, which require peroxide as a co-substrate, have been reported as excellent oxidizing agents for decomposing dyes. Several types of peroxidases, such as lignin peroxidase (LiP), manganese peroxidase (MnP), versatile peroxidase (VP), and dye-decolorizing peroxidase (DyP), are commonly found in fungi and bacteria [3]. Bacteria offer advantages in ease of culture handling compared to fungi, making them more suitable for the decomposition of synthetic textile dyes.

* Corresponding author: irfan@ub.ac.id

The chemical structure of synthetic dyes is categorized based on their chromophore groups, namely azo, anthraquinone, and triphenylmethane. Anthraquinone dyes are the second most used dyes after azo dyes due to their low cost, wide availability, and excellent dyeing performance. The dyes possess complex and highly stable structures, making them more toxic to microorganisms and human cells compared to azo dyes. [4].

Anthraquinone is a biologically recalcitrant compound, making its degradation particularly challenging. Therefore, isolating bacteria with the potential for anthraquinone bioremediation requires an appropriate approach. One effective strategy is to focus on bacterial taxa in which most members harbor the DyP gene, allowing for targeted isolation within specific taxonomic groups. This study aims to analyze the distribution of the DyP gene among bacterial species available in the NCBI database and find the most abundant phylum.

2 Methods

The nucleotide database search on NCBI was conducted in June 2022 using the keywords of both “dye oxidizing peroxidase” and “dye oxidizing peroxidase bacteria”. The coding sequences (CDS) for the DyP enzyme genes from each registered strain were downloaded individually in FASTA format and compiled into a single Notepad file. The composition of the search results was tabulated using Microsoft Excel and visualized as a diagram. The nucleotide sequences in FASTA format were aligned using ClustalX2. The output from ClustalX2 was converted into a format compatible with MEGA7 software for further analysis, enabling the construction of a phylogenetic tree using the Neighbor-Joining algorithm. For simplification, the phylogenetic tree only shows the names of bacterial species that represent various strains beneath them. Species with partial CDS and uncultured status were excluded from the phylogenetic tree. Partial DNA sequences were omitted because their functions were not fully confirmed, while uncultured organisms were excluded as they were largely not identified. The average length of the DyP gene CDS analyzed was approximately 1,250 base pairs.

3 Results and discussion

The search results obtained using the keyword “dye oxidizing peroxidase bacteria” in the NCBI nucleotide database yielded 773 items, while the keyword “dye oxidizing peroxidase” produced 847 items, all of which were whole genome sequences. This discrepancy suggests that DyP-encoding genes are also found in non-bacterial organisms, including fungi and animals. In the fungal category, 55 search results were identified, distributed across 23 different genera. In the animal category, ten search results were found, originating from 7 invertebrate genera belonging to the phylum Mollusca, class Insecta, and class Cestoda, as well as one vertebrate genus from the family Salmonidae. These findings suggest that the enzyme is most frequently found in prokaryotic microorganisms, specifically bacteria, while other prokaryotes, such as Archaea, did not appear in the search results.

A review article reported that DyP enzymes are mostly found in eukaryotic microorganisms, particularly fungi. DyP enzymes from bacteria remain relatively understudied [4]. This is unsurprising, as fungi have a well-established history of producing hydrolytic enzymes capable of degrading various complex compounds in nature. Notably, the first DyP protein was isolated from the soil fungus *Thanatephorus cucumeris*, which was shown to effectively decolorize dyes [5].

3.1 Bacterial phyla harboring the DyP gene

Within the bacterial category, the search results revealed that DyP genes were found in 127 genera across three phyla: Actinobacteria, Firmicutes, and Proteobacteria (Figure 1). Actinobacteria accounted for the largest proportion, exceeding three-quarters of the search results, while Firmicutes and Proteobacteria contributed 2.3% and 14.1%, respectively. Actinobacteria are Gram-positive bacteria with high guanine and cytosine content in their DNA. These unicellular prokaryotes exhibit unique characteristics not found in other bacteria, such as the formation of non-septate mycelia. Additionally, members of this phylum play a crucial role in the decomposition of organic compounds such as cellulose and chitin [6]. Firmicutes, also Gram-positive, can be found in both rod and cocci forms. On the other hand, Proteobacteria is comprised of a diverse group of Gram-negative bacteria. These three phyla are recognized as among the most abundant groups of bacteria found across various environments on Earth.

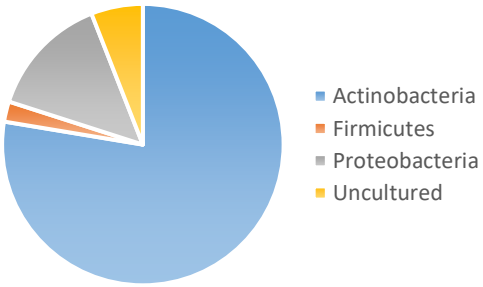


Fig. 1. Composition of bacterial phyla possessing the DyP gene based on the nucleotide database search results in NCBI.

Members of these three phyla have been reported to possess DyP enzymes with potential applications in the treatment of dyeing wastewater, such as *Streptomyces lividans* SLI4211 from Actinobacteria, *Bacillus subtilis* KCTC2023 (Firmicutes), and *Pseudomonas fluorescens* Pf-5 (Proteobacteria) [5, 7]. The search results revealed that most species identified as possessing the DyP gene belong to the phylum Actinobacteria. This significant difference in the proportion may be contributed by the massive registration of DNA sequences from the whole-genome sequencing project of 1000 Actinobacteria strains, submitted by the DOE-Joint Genome Institute to the NCBI database [8].

3.2 Bacterial genera harboring the DyP gene

The composition of bacterial genera with an abundance greater than 1% from the search results is shown in Figure 2. *Escherichia coli* is the only species outside Actinobacteria displayed in the figure, whereas Streptomyces is the Actinobacteria genus with the highest number of strains (31%) whose genomes have been registered. All bacterial genera detected in the search results are associated with complete whole-genome sequences, meaning that the DyP gene sequences presented are complete CDS. Whole-genome sequencing provides comprehensive genomic information, including the detection of all CDS along with 16S rRNA gene sequences, enabling the accurate identification of bacterial species. [9]

A chart sector in Figure 2 labeled "uncultured" corresponds to DyP gene sequences partially obtained from uncultured bacteria; thus, the species harboring these genes remain unidentified. These sequences are often derived from DyP gene clones obtained from

environmental samples amplified by the polymerase chain reaction [10]. Additionally, "uncultured" may also originate from whole-genome shotgun sequencing assembled from DNA of uncultured bacterium.

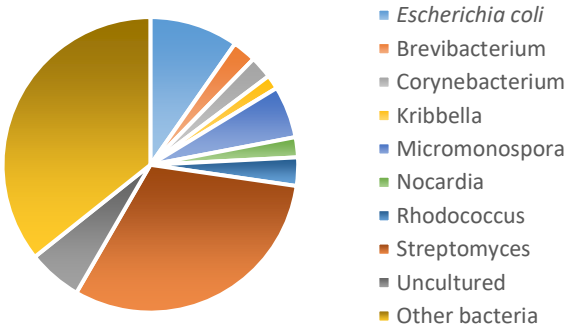


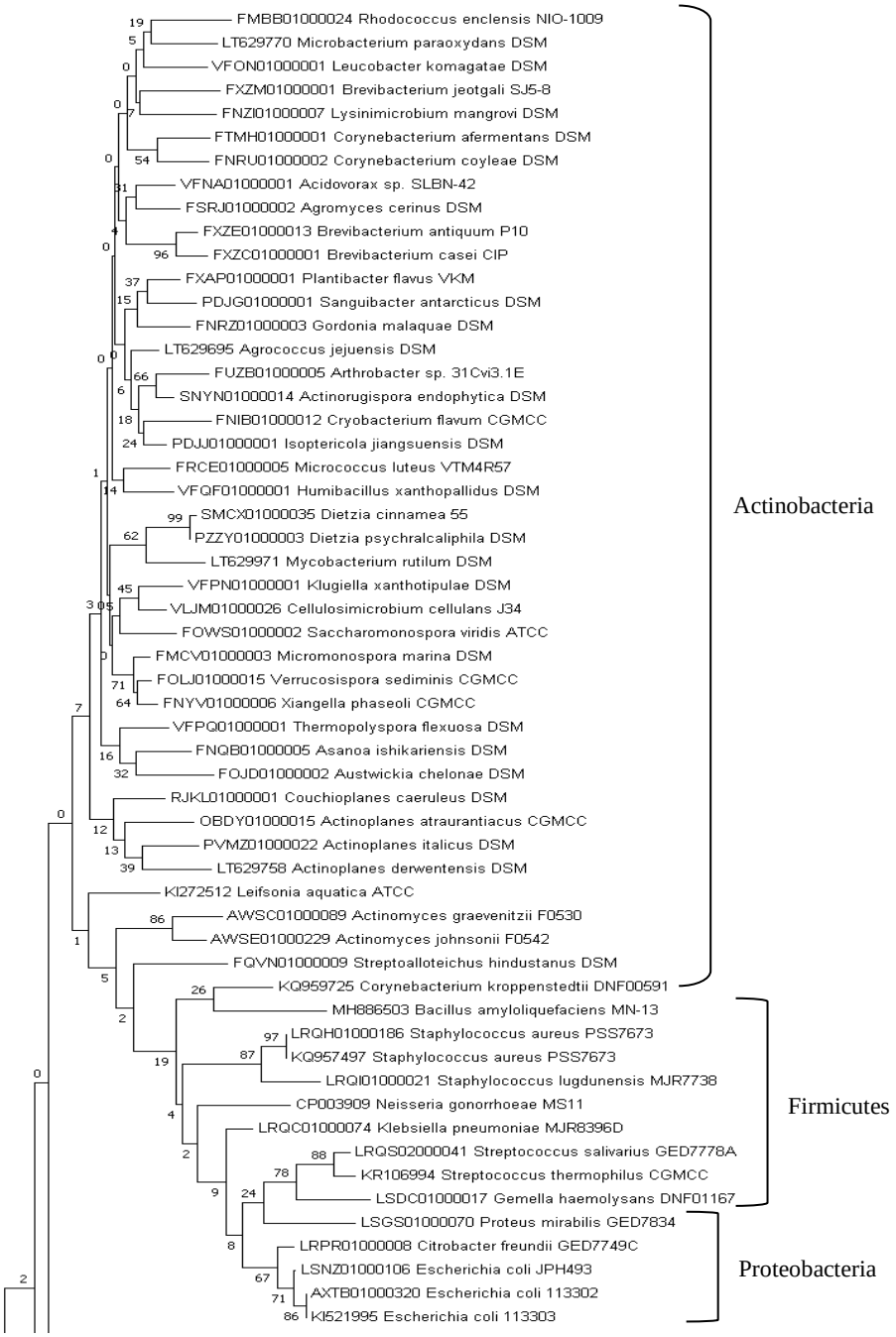
Fig. 2. Composition of bacterial genera possessing the DyP gene based on the nucleotide database search results in NCBI.

3.3 Phylogenetic tree of DyP-gene harboring bacteria

The relationships among the obtained species are presented in the form of a phylogenetic tree based on their DyP gene sequences (Figure 3). The tree divides the sequences into five major clusters. The first, third, and fifth clusters consist of members of the phylum Actinobacteria, while the second cluster includes members of both phyla Firmicutes and Proteobacteria. The fourth cluster is exclusively composed of Proteobacteria. This suggests that the grouping of DyP gene sequence is consistent with the classification of phyla based on 16S rDNA sequences. According to [11], closely related bacterial species often exhibit high DNA sequence similarity, reflecting their shared evolutionary history.

The widespread presence of the DyP gene in the phylum Actinobacteria suggests its broad distribution within this group. Therefore, screening for potential DyP-producing bacteria can be directed to Actinobacteria, using anthraquinone compounds as the sole carbon source to enhance growth medium selectivity.

Dye oxidizing peroxidase is known to have diverse substrates, including other complex aromatic compounds and lignin-derived compounds. This versatility allows DyP enzymes to be widely applied in wastewater treatment, bioremediation, and lignocellulose biorefineries. In lignocellulose biorefineries, they play a crucial role in degrading lignin to access cellulose and hemicellulose, which are then utilized for biofuel and biomaterial production [12, 13]. These characteristics highlight the potential of DyP enzymes as a valuable tool for sustainable industrial processes.


FMBB01000024 Rhodococcus enclensis NIO-1009
LT629770 Microbacterium paraoxydans DSM
VFON01000001 Leucobacter komagatae DSM
FXZM01000001 Brevibacterium jeotgali SJ5-8
FNZI01000007 Lysinimicrobium mangrovi DSM
FTMH01000001 Corynebacterium afermentans DSM
FNRU01000002 Corynebacterium coyleae DSM
VFNA01000001 Acidovorax sp. SLBN-42
FSRJ01000002 Agromyces cerinus DSM
FXZE01000013 Brevibacterium antiquum P10
FXZC01000001 Brevibacterium casei CIP
FXAP01000001 Plantibacter flavus VKM
PDJG01000001 Sanguibacter antarcticus DSM
FNRZ01000003 Gordonia malaquae DSM
LT629695 Agrococcus jejuensis DSM
FUZB01000005 Arthrobacter sp. 31Cvi3.1E
SNYN01000014 Actinorugispora endophytica DSM
FNIB01000012 Cryobacterium flavum CGMCC
PDJJ01000001 Isoptericola jiangsuensis DSM
FRCE01000005 Micrococcus luteus VTM4R57
VFQF01000001 Humibacillus xanthopallidus DSM
SMCX01000035 Dietzia cinnamea 55
PZZY01000003 Dietzia psychralcaliphila DSM
LT629971 Mycobacterium rutilum DSM
VFPN01000001 Klugiella xanthotipulae DSM
VLJM01000026 Cellulosimicrobium cellulans J34
FOWS01000002 Saccharomonospora viridis ATCC
FMCV01000003 Micromonospora marina DSM
FOLJ01000015 Verrucosispora sediminis CGMCC
FNYV01000006 Xiangella phaseoli CGMCC
VFPQ01000001 Thermopolyspora flexuosa DSM
FNQB01000005 Asanoa ishikariensis DSM
FOJD01000002 Austwickia chelonae DSM
RJKL01000001 Couchioplanes caeruleus DSM
OBDY01000015 Actinoplanes atraurantiacus CGMCC
PVMZ01000022 Actinoplanes italicus DSM
LT629758 Actinoplanes derwentensis DSM
KI272512 Leifsonia aquatica ATCC
AWSC01000089 Actinomyces graevenitzii F0530
AWSE01000229 Actinomyces johnsonii F0542
FQVN01000009 Streptoalloteichus hindustanus DSM
KQ959725 Corynebacterium kroppenstedtii DNF00591
MH886503 Bacillus amyloliquefaciens MN-13
LRQH01000186 Staphylococcus aureus PSS7673
KQ957497 Staphylococcus aureus PSS7673
LRQI01000021 Staphylococcus lugdunensis MJR7738
CP003909 Neisseria gonorrhoeae MS11
LRQC01000074 Klebsiella pneumoniae MJR8396D
LRQS02000041 Streptococcus salivarius GED7778A
KR106994 Streptococcus thermophilus CGMCC
LSDC01000017 Gemella haemolysans DNF01167
LSGS01000070 Proteus mirabilis GED7834
LRPR01000008 Citrobacter freundii GED7749C
LSNZ01000106 Escherichia coli JPH493
AXTB01000320 Escherichia coli 113302
KI521995 Escherichia coli 113303
Actinobacteria
Firmicutes
Proteobacteria

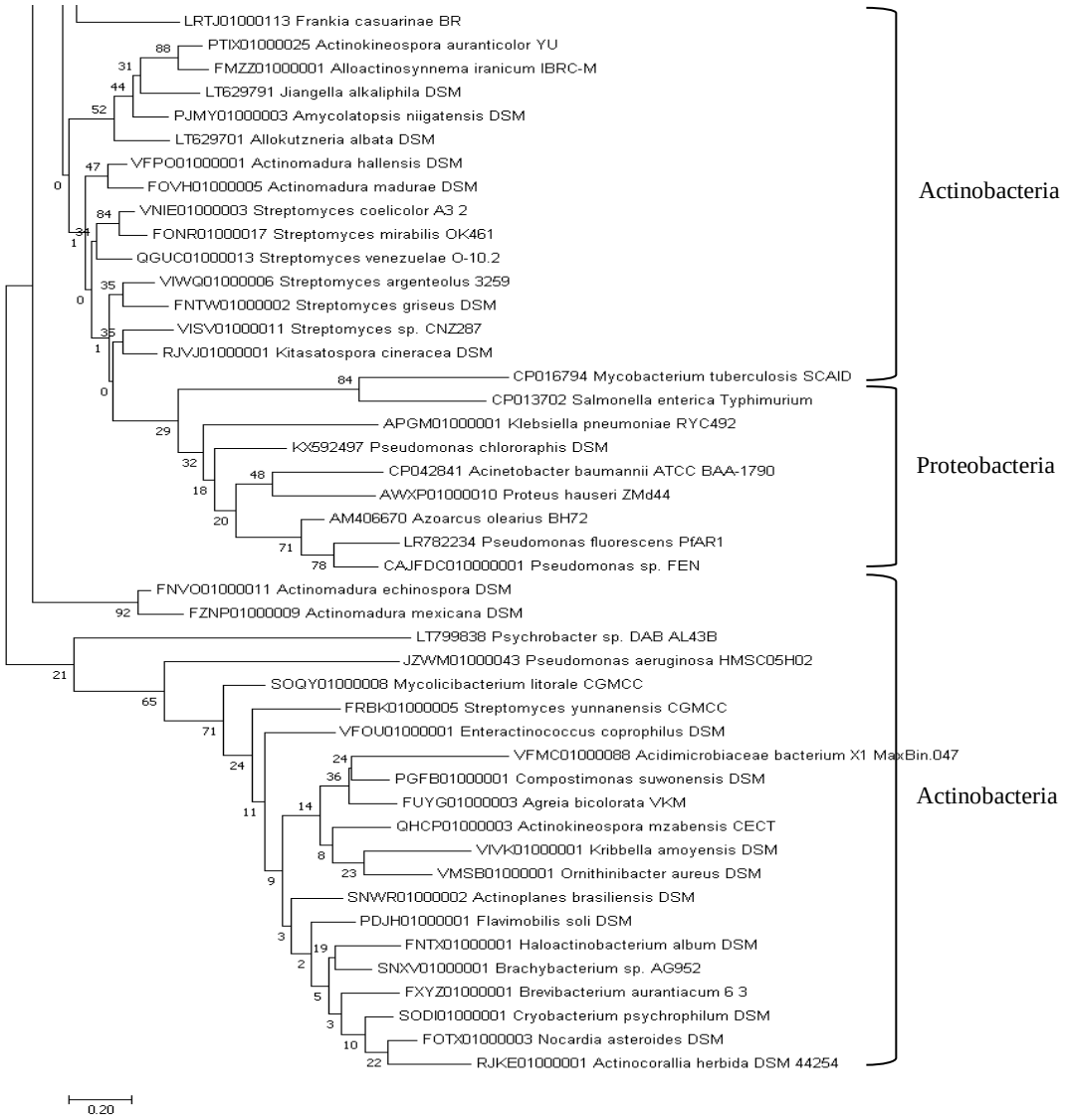


Fig. 3. Neighbor-joining phylogenetic tree showing relationships among DyP-gene harboring bacteria from NCBI database. The scale bar in the bottom represents 0.20 nucleotide substitutions per nucleotide position.

4 Conclusion

The search for the presence of the DyP gene in the NCBI nucleotide database identified 776 bacterial strains. Actinobacteria emerged as the most frequently detected phylum, with Streptomyces being the most represented genus. These findings suggest that isolation of anthraquinones oxidizing bacteria should focus on employing media optimized to support the growth of Actinobacteria.

I would like to express my sincere appreciation to FMIPA Universitas Brawijaya for their financial support through Research Internal Funding Program of this research project and to the members of the research team who contributed to the successful completion of this study.

References

1. F. M. D. Chequer, G.A.R. De Oliveira, E.R.A. Ferraz, J.C. Cardoso, Textile Dyes: Dyeing Process and Environmental Impact, Chapter 6 (IntechOpen, 2013)
2. M. A. Hassaan, A. El Nemr, Health and environmental impacts of dyes: mini review. Am. J. Environ. Sci. Eng. **1**, 3 (2017)
3. N. Bansal, S.S. Kanwar, Review article: peroxidase(s) in environment protection. Sci. World J. **2013** (2013)
4. H. Li, Y. Wang, Y. Wang, H. Wang, K. Sun, Z. Lu, Bacterial degradation of anthraquinone dyes. J. Zhejiang Univ-Sci B (Biomed & Biotechnol) **20**, 6 (2019)
5. C. Chen, T. Li, Bacterial dye-decolorizing peroxidases: biochemical properties and biotechnological opportunities. Phys. Sci. Rev. **1**, 9 (2016)
6. R. Anandan, D. Dharumadurai, G.P. Manogaran, Chapter 1: An Introduction to Actinobacteria (IntechOpen, 2016)
7. A. K. Chaplin, M.T. Wilson, J.A.R. Worrall, Kinetic characterisation of a dye decolourising peroxidase from *Streptomyces lividans*: new insight into the mechanism of anthraquinone dye decolourisation. Dalton Trans. **46** (2017)
8. Joint Genome Institute, Exploiting Actinobacteria genomes, https://jgi.doe.gov/ Accessed on June 2022
9. V. V. Kadnikova, A.V. Mardanova, A.V. Beletskya, Y.A. Frankc, O.V. Karnachukc, N.V. Ravin, Complete genome sequence of an uncultured bacterium of the candidate Phylum Bipolaricaulota. Microbiol. **88**, 4 (2019)
10. M. Soman, M. Mini, S. Joseph, J. Thomas, N. Chacko, T.G. Sumithra, R. Ambily, B.K. Mani, R. Balan, Cloning and sequence analysis of a partial CDS of leptospiral ligA gene in pET-32a – *Escherichia coli* DH5α system. Vet. World **11** (2018)
11. N. Han, W. Yu, Y. Qiang, W. Zhang, T4SP database 2.0: an improved database for type iv secretion systems in bacterial genomes with new online analysis tools. Comput. Math. Methods Med. **2016** (2016)
12. D. Silva, C.F. Rodrigues, C. Lorena, P.T. Borges, L.O. Martins, Biocatalysis for biorefineries: the case of dye-decolorizing peroxidases. Biotechnol. Adv. **65** (2023)
13. T. Yoshida, Y. Sugano, Unexpected diversity of dye-decolorizing peroxidases. Biochem. Biophys. Rep. **33** (2023)