

Article

Computational Interpretation of Functional Divergence of VOC Family Catechol Dioxygenases in *Bacillus thuringiensis* HHY919: Insights from Homology Modeling and Molecular Docking

Liwei Yang [†], Hongyan Hou [†], Wenjie Zhang, Weibing Zhang, Wei Zhang, Yulong Zhao , Feier Ren, Shuaijiang Guo and Zhonghao Wang ^{*} 

Key Laboratory of Livestock and Poultry Healthy Breeding Technology in Northwest China, Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Xinjiang Agricultural Vocational and Technical University, Changji 831100, China

^{*} Correspondence: wzhh1402@163.com or zhonghaowang@xjvut.edu.cn

[†] These authors contributed equally to this work.

Abstract

Catechol 2,3-dioxygenase (C23O) is the rate-limiting enzyme in the meta-cleavage pathway of aromatic compound degradation, yet its functional annotation within the structurally conserved VOC superfamily remains challenging due to high sequence homology. Here, we isolated a catechol-degrading strain from bovine feces and identified it as *Bacillus thuringiensis* HHY919 via whole-genome sequencing. Genome annotation revealed four VOC genes sharing the same COG annotation (“catechol 2,3-dioxygenase”) but divergent KO annotations (two as glyoxalases and two as C23O), suggesting functional divergence. Using homology modeling and molecular docking, we compared their binding affinities toward catechol and 11 derivatives. All four proteins showed typical meta-cleavage binding energies (−4.9 to −5.4 kcal/mol), with slightly more favorable binding than an ortho-cleavage control. Notably, gene4224 exhibited the broadest and strongest predicted affinities in virtual screening against 212 compounds, particularly for trichlorophenol (−6.0 kcal/mol) and complex natural products (qvina_score ≤ −7.6 kcal/mol). Phylogenetic analysis and docking results jointly identified gene3355 and gene4224 as computationally prioritized C23O candidates, with gene4224 recommended as the top candidate for future enzyme engineering and bioremediation studies. This study provides a computational workflow for resolving functional ambiguity in VOC family enzymes and generating testable hypotheses for experimental validation.

Keywords: *Bacillus thuringiensis*; catechol 2,3-dioxygenase; VOC superfamily; homology modeling; molecular docking; substrate specificity; functional divergence; bioremediation



Academic Editor: Maurizio Muniz-Miranda

Received: 6 August 2026

Revised: 27 August 2026

Accepted: 31 August 2026

Published: 8 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Aromatic compounds (phenol, toluene, chlorobenzenes, biphenyls, polycyclic aromatic hydrocarbons, etc.) are persistent organic pollutants widely distributed in the environment, primarily originating from industrial activities, for example, the petrochemical, pesticide, dye, and plastic industries [1–3]. Due to their carcinogenic, teratogenic, and mutagenic properties, they pose serious threats to ecosystems and human health. Physical and chemical remediation methods are often costly and prone to secondary pollution, but microbial degradation is considered a sustainable and cost-effective approach [4–7].

In microbial degradation pathways, catechol 2,3-dioxygenase (C23O, EC 1.13.11.2) catalyzes the meta-cleavage of the aromatic ring—a rate-limiting step that determines the overall efficiency of pollutant mineralization [8–10]. C23O belongs to the vicinal oxygen chelate (VOC) superfamily, which features a conserved fold and a metal-coordinating motif [11–13]. VOC family members catalyze diverse reactions, including dioxygenation, isomerization, and nucleophilic addition, while sharing a common mechanism: substrate coordination to the divalent metal ion via adjacent oxygen atoms [13].

A key challenge in VOC family annotation is the discordance between different homology-based classification systems. COG (Clusters of Orthologous Groups) annotation reflects broad structural domain membership [14], whereas KO (KEGG Orthology) annotation provides finer biochemical functional assignments [15]. When these two annotations conflict, it suggests possible functional divergence or annotation ambiguity—a problem that has not been systematically addressed using structure-based approaches.

Bacillus thuringiensis is widely distributed in environments such as soil, water, and food, and multiple studies have previously reported its ability to degrade various aromatic pollutants, including phenol and benzoic acid [16]. Whole-genome sequencing analyses have identified key enzyme genes associated with aromatic compound degradation, including catechol 2,3-dioxygenases, in multiple *Bacillus* strains [12].

In this study, we isolated a catechol-degrading *Bacillus thuringiensis* strain from environmental samples. Whole-genome annotation revealed four VOC family genes (*gene1627*, *gene2450*, *gene3355*, and *gene4224*) sharing the same COG annotation (“catechol 2,3-dioxygenase”), but divergent KO annotations: *gene1627* and *gene2450* as glyoxalases (K01759), and *gene3355* and *gene4224* as C23O (K07104). To resolve this functional ambiguity, we employed homology modeling and molecular docking to compare the substrate-binding affinities of the four proteins. We further assessed the broad-spectrum binding potential of the top candidate through large-scale virtual screening against 212 natural products. Our integrated approach aims to: (1) determine whether the four proteins exhibit computable differences in substrate binding that correlate with KO annotations; (2) identify structural features underlying predicted functional divergence; and (3) prioritize the most promising C23O candidate for future experimental validation and bioremediation applications.

2. Results

2.1. 16S rRNA Gene Analysis of Strain HHY919

A 1454 bp fragment of the 16S rRNA gene was obtained. Comparison with known type strains via the EzBioCloud website (<http://www.ezbiocloud.net/eztaxon/>, accessed on 30 August 2026) showed that strain HHY919 had the highest similarity (100%) to *Bacillus toyonensis* BCT-7112^T and *Bacillus mobilis* 0711P9-1^T. Similarities to other type strains are provided in Supplementary Table S1. A phylogenetic tree based on 16S rRNA gene sequences was constructed for strain HHY919 and related taxa (Figure 1A).

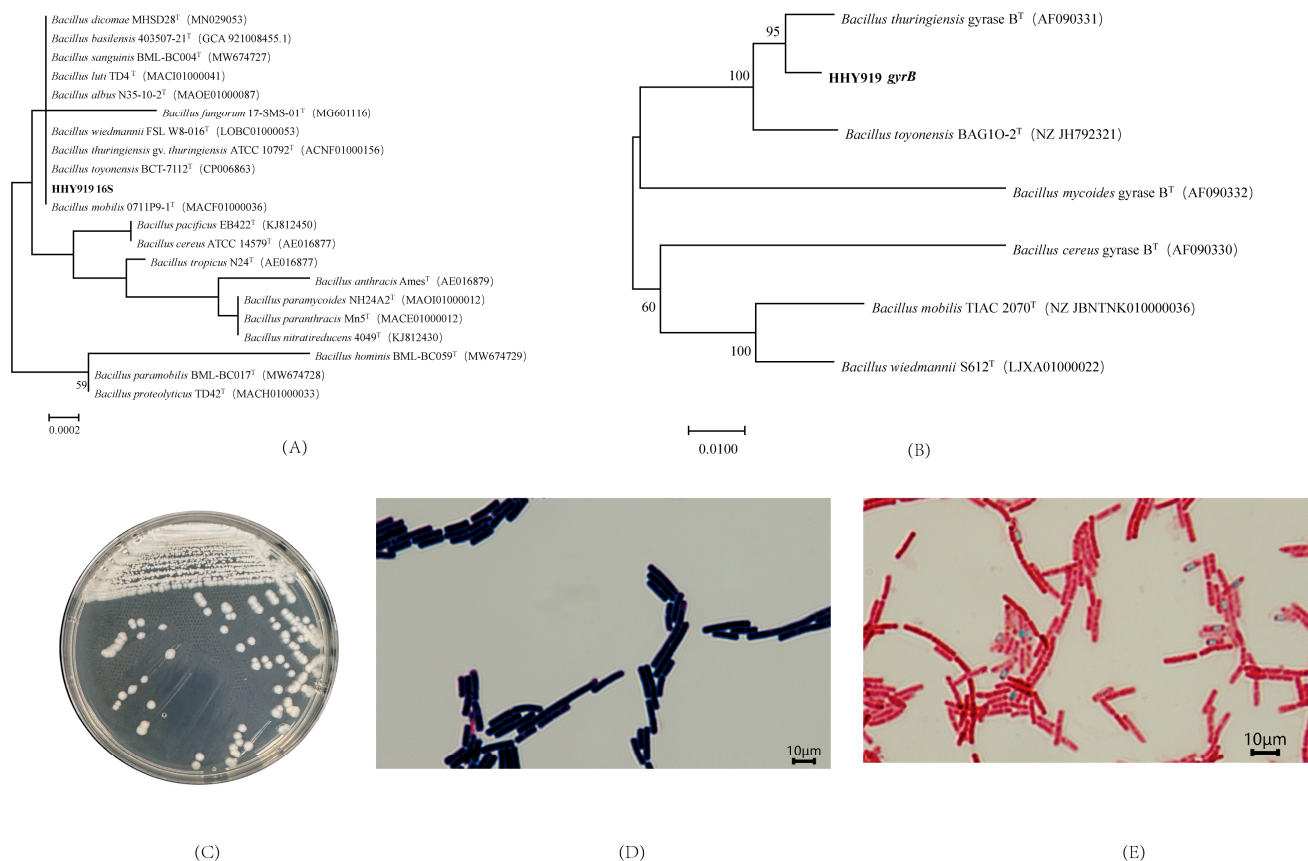


Figure 1. (A) Phylogenetic analysis based on 16S rRNA gene sequences of strain HHY919; (B) phylogenetic analysis based on gyrB gene sequences of strain HHY919; (C) colony morphology of strain HHY919; (D) Gram-staining micrograph of strain HHY919; (E) spore-staining micrograph of strain HHY919.

2.2. gyrB Gene Sequence Analysis of Strain HHY919

A 1131 bp fragment of the gyrB gene was obtained. NCBI BLAST (2.14.0) results showed that strain HHY919 had the highest similarity (98.85%) to *Bacillus thuringiensis* ATCC 10792^T. A phylogenetic tree of strain HHY919 and its close relatives was constructed (Figure 1B), demonstrating that strain HHY919 stably clustered with *Bacillus thuringiensis*.

2.3. Colony Morphological Characteristics of Strain HHY919

After 24 h of incubation on TSA medium at 30 °C, colonies of strain HHY919 appeared pale yellow, opaque, and dull (Figure 1C). Gram staining was positive, and cells were rod-shaped (Figure 1D). The results of pore staining are shown in Figure 1E. These morphological and Gram-staining characteristics are consistent with those of the genus *Bacillus*.

2.4. Physiological and Biochemical Characteristics of Strain HHY919

The fatty acid profiles determined by the MIDI system showed that the major fatty acids of strain HHY919 were C15:0 iso (27.51%), C13:0 iso (12.42%), C16:0 iso (7.65%), and C14:0 iso (7.50%). Detailed fatty acid results, consistent with the major cellular fatty acid characteristics of the genus *Bacillus*, are presented in Table 1.

API 50CH test results showed positive reactions for D-ribose, D-glucose, D-fructose, N-Acetylglucosamine, arbutin, esculin, salicin, D-Cellulose, D-Maltose, D-Trehalose, Starch, and Gluconat; weakly positive reactions for glycerinum, Inositol, and Amygdalin; and negative reactions for the remaining substrates (see Table 2). These results are consistent with the biochemical metabolic characteristics of the genus *Bacillus*.

Table 1. Fatty acid composition of strain HHY919.

Peak Name	Percent	Peak Name	Percent
11:0 iso	0.13	16:0	5.11
12:0 iso	1.62	15:0 2OH	0.83
12:0	0.64	17:1 iso ω 10c	2.89
13:0 iso	12.42	17:1 iso ω 5c	2.93
13:0 anteiso	1.29	17:1 anteiso A	0.48
13:0	0.14	17:0 iso	5.18
14:0 iso	7.50	17:0 anteiso	0.84
14:0 anteiso	0.09	17:1 ω 9c	0.07
14:0	4.12	17:1 ω 7c	0.11
13:0 iso 3OH	0.11	17:0	0.24
15:0 iso	27.51	18:0 iso	0.12
15:0 anteiso	3.48	18:1 ω 9c	0.18
15:1 ω 5c	0.36	18:0	0.31
16:1 ω 7c alcohol	1.49	17:0 iso 3OH	0.12
16:0 iso	7.65	19:0 iso	0.17
16:1 ω 11c	0.46		
Peak Name	Percent	Comment1	Comment2
Summed Feature 2	3.18	12:0 aldehyde	unknown 10.9525
Summed Feature 3	8.23	16:1 ω 7c/16:1 ω 6c	16:1 ω 6c/16:1 ω 7c

Table 2. API 50CH test results for strain HHY919.

Analytical Profile Index (API)	Result	Analytical Profile Index (API)	Result
Control	—	Esculin	+
Glycerinum	w	Salicoside	+
Erythrose	—	D-Cellobiose	+
D-arabinose	—	D-Maltose	+
L-arabinose	—	D-Lactose	—
D-Ribose	+	D-Melibiose	—
D-xylose	—	D-Sucrose	—
L-xylose	—	D-Trehalose	+
D-Adonitol	—	Inulin	—
β -methyl-D-xyloside	—	D-Melezitose	—
D-Galactose	—	D-Raffinose	—
D-Glucose	+	Starch	+
D-Fructose	+	Glycogen	+
D-Mannose	—	Xylitol	—
L-Sorbose	—	Gentiobiose	—
L-Rhamnose	—	D-Turanose	—
Dulcitol	—	D-Lyxose	—
Inositol	w	D-Tagatose	—
D-Mannitol	—	D-Fucose	—
D-Sorbitol	—	L-Fucose	—
α -methyl-D-mannosidase	—	D-Arabitol	—
α -methyl-D-glucoside	—	L-Arabitol	—
N-Acetylglucosamine	+	Gluconat	+
Amygdalin	w	2-keto-gluconate	—
Arbutin	+	5-keto-gluconate	—

Note: +, positive; —, negative; w, weakly positive.

2.5. Genomic Analysis

As shown in Table 3, the whole-genome sequence had a total length of 5,657,846 bp (Figure 2A), with a GC content of 35.26%, and contained 5567 total genes. The ANI value with *Bacillus_A_thuringiensis* was 99.1752%. These genomic data further confirmed that this strain belongs to *Bacillus thuringiensis*. COG annotation was used for detailed evolutionary and functional classification of the amino acid sequences of *B. thuringiensis* HHY919

(Figure 2B). Gene Ontology (GO) annotation provided detailed functional categorization across different biological levels (Figure 2C), while KEGG pathway annotation revealed the metabolic pathways involved in aromatic compound degradation in *B. thuringiensis* HHY919 (Figure 2D). The 16S rRNA gene sequences are highly conserved among closely related species within the *Bacillus cereus* group and cannot effectively distinguish individual members. Therefore, based on *gyrB* gene sequence similarity (98.85%) and whole-genome ANI values (99.1752%), the strain was conclusively identified as *Bacillus thuringiensis*.

Table 3. Genomic analysis of *Bacillus thuringiensis* HHY919.

Item	Value
Sample name	<i>Bacillus thuringiensis</i> HHY919
Genome size	5,657,846 bp
Whole-genome G+C content	35.26%
Total gene count	5567
Number of plasmids	1
ANI	99.1752%

2.6. Identification and Sequence Characteristics of VOC Family Target Genes

The phylogenetic tree (Figure 3A) exhibited two major independent evolutionary clades, and the branch divisions were highly consistent with the catalytic functions indicated by the downstream reactions (Figure 3B–E). This clustering reflects phylogenetic relatedness to *Pseudomonas* homologs. We emphasize that phylogenetic topology indicates evolutionary relatedness and does not independently prove catalytic function; rather, it provides a framework for interpreting the docking results detailed below.

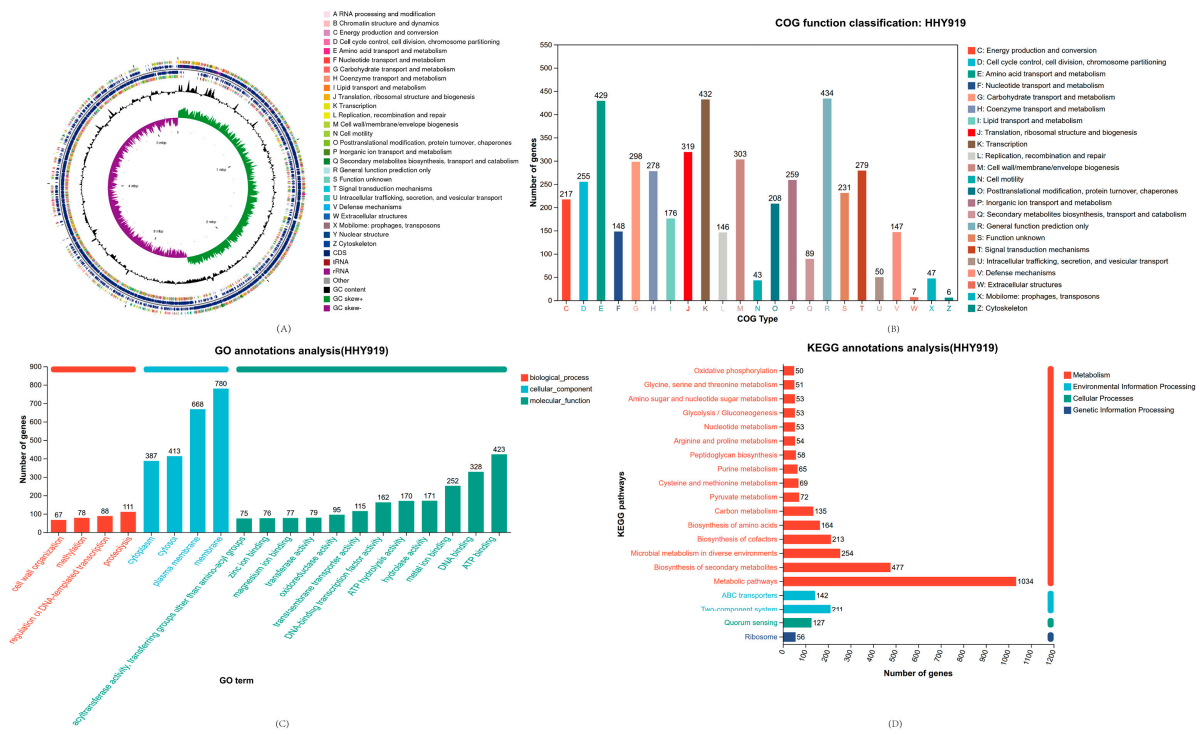


Figure 2. (A) Whole-genome circular map of *Bacillus thuringiensis* HHY919; (B) COG annotation results and analysis; (C) GO annotation statistics (Top 10); (D) KEGG annotation results and analysis.

Clade I (upper main cluster): *Pseudomonas*-type meta-cleavage functional cluster. This clade contained gene1627, gene2450, and the reference proteins Q04285 (Figure 3C) and P08127 (Figure 3D) as a monophyletic group with high bootstrap support (>85%). This clustering reflects the phylogenetic relatedness of gene1627 and gene2450 to *Pseudomonas*-derived

C23O homologs. However, we emphasize that phylogenetic topology indicates evolutionary relatedness rather than independently proving catalytic function; it provides a framework for interpreting the docking results in the context of known enzymes.

Clade II (lower main cluster): *Bacillus*-type meta-cleavage functional cluster (with distal ortho-cleavage reference outgroup).

Within this large clade, gene4224 and gene3355 formed a tight sister subclade with the reference protein P54721 (Figure 3B). This indicates that gene4224 and gene3355 share close evolutionary relatedness to Gram-positive (*Bacillus*-type) meta-cleavage dioxygenases, forming a phylogenetic subtype distinct from the *Pseudomonas* type in Clade I.

Basal distant lineage (ortho reference outgroup): The reference protein P07773 (Figure 3E) was located at the basal-most position outside this large clade, with considerable genetic distance from gene4224/gene3355. Figure 3E clearly shows that P07773 catalyzes ortho-cleavage (catechol 1,2-dioxygenase activity). Its position in the tree suggests that this large clade retains an ancient ortho-cleavage ancestral lineage, while gene4224 and gene3355 have diverged along the meta-cleavage evolutionary direction, with no direct functional relationship to the ortho function of P07773.

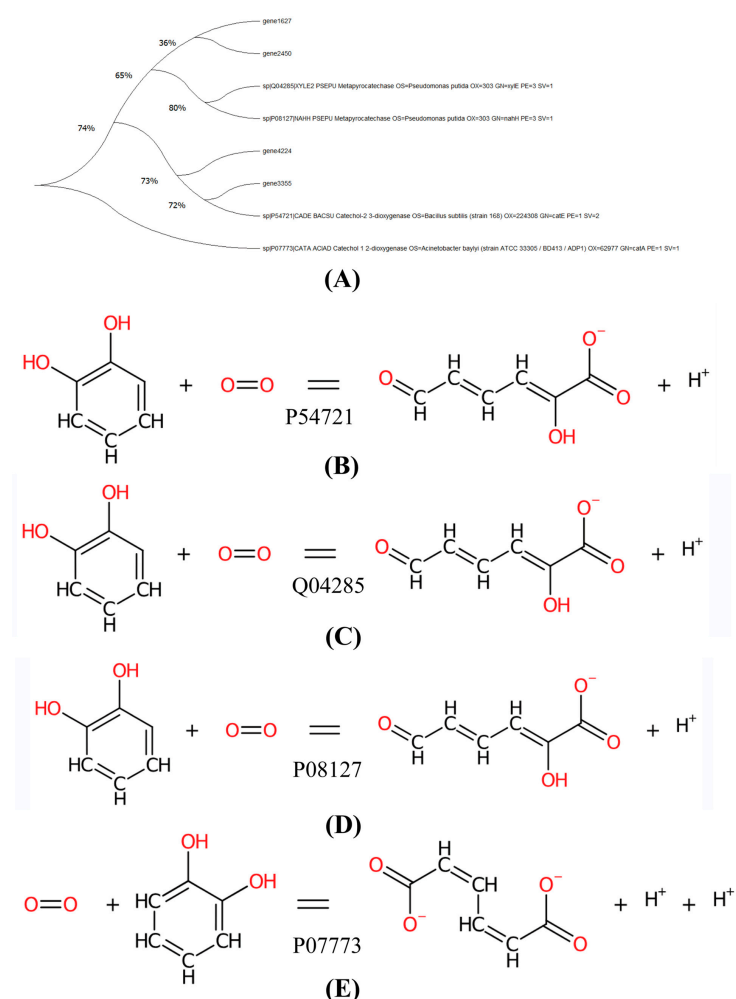


Figure 3. (A) Phylogenetic relationships of the four HHY919 VOC proteins and reference enzymes. The tree resolves the four proteins into two distinct clades: Clade I (*Pseudomonas*-type meta-cleavage, containing gene1627/gene2450 with Q04285 and P08127) and Clade II (*Bacillus*-type meta-cleavage, containing gene3355/gene4224 with P54721). The ortho-cleavage enzyme P07773 (catechol 1,2-dioxygenase) serves as the outgroup. The ring-opening activities of reference enzymes P54721 (B), Q04285 (C), P08127 (D), and P07773 (E) are shown, confirming the functional assignments inferred from phylogenetic clustering.

2.7. Molecular Docking Analysis: Binding Modes of Catechol to the Active Sites

To further validate the functional typing revealed by the phylogenetic tree (Figure 3) and to evaluate the affinity and binding mechanisms of each enzyme toward catechol, three-dimensional structures were constructed using SWISS-MODEL homology modeling, and molecular docking was performed using CB-Dock3. Model quality and binding energy assessments showed that all homology models of the native genes exhibited high structural reliability (Table 4: GMQE $\geq$ 0.95, sequence identity up to 90.32–98.59%), providing a solid foundation for subsequent docking analyses.

Table 4. Homology modeling quality and catechol binding energies.

Protein	Binding Energy (kcal/mol) (Fe ²⁺)	Seq Identity	GMQE	Source
gene1627	−5.2	90.32%	0.96	HHY919
gene2450	−5.0	92.05%	0.97	HHY919
gene3355	−4.9	97.02%	0.95	HHY919
gene4224	−5.4	98.59%	0.96	HHY919
P54721	−5.5	a	a	meta (<i>Bacillus</i> -derived)
Q04285	−5.5	a	a	meta (<i>Pseudomonas</i> -derived)
P08127	−5.6	a	a	meta (<i>Pseudomonas</i> -derived)
P07773	−4.7	a	a	ortho control

Note: “a” indicates reference protein models obtained directly from UniProt.

As shown in Figure 4, the binding energies of the four native genes to catechol (−4.9 to −5.4 kcal/mol) were highly similar to those of the classical meta-cleavage reference enzymes (P54721, P08127, Q04285; −5.5 to −5.6 kcal/mol) and more favorable than that of the ortho-cleavage control P07773 (−4.7 kcal/mol). These binding energies are consistent with all native genes possessing catechol-binding capabilities and preferentially adopting meta-cleavage enzyme binding modes.

2.8. Molecular Docking Results: Extended Substrate Profiling

To systematically interpret the substrate recognition patterns underlying the binding energy data, the 12 ligands were categorized into three groups based on substituent properties: alkyl/methyl class (4-methylcatechol, o-/m-/p-cresol), halogenated/polychlorinated class (4-chlorocatechol, di-/tri-/pentachlorophenol), and monocyclic/simple aromatics (phenol, benzene, nitrophenol). Quantitative analysis revealed that the binding energy patterns of the four native genes were clearly consistent with their phylogenetic branch assignments (Table 5).

Table 5. Molecular docking results (binding energies in kcal/mol).

Substrate	gene1627	gene2450	gene3355	gene4224	P54721	Q04285	P08127	P07773
Phenol	−4.9	−4.6	−4.5	−4.7	−4.6	−4.9	−4.6	−4.6
Benzene	−4.5	−4.0	−3.5	−4.1	−3.8	−4.2	−3.9	−4.2
Nitrophenol	−5.3	−5.3	−5.4	−5.6	−6.0	−5.6	−5.9	−5.4
4-Methylcatechol	−5.4	−5.3	−5.4	−5.8	−5.9	−6.0	−5.9	−5.3
4-Chlorocatechol	−5.1	−5.1	−5.4	−5.5	−5.8	−5.1	−5.7	−5.3
Dichlorophenol	−5.3	−4.9	−5.1	−5.5	−5.3	−4.9	−5.0	−5.2
Trichlorophenol	−5.0	−4.8	−4.9	−6.0	−5.6	−4.7	−4.8	−5.1
Pentachlorophenol	−5.2	−5.1	−4.9	−5.7	−6.0	−4.5	−5.2	−5.1
o-Cresol	−5.3	−5.0	−5.0	−5.6	−5.2	−5.1	−5.2	−4.8
m-Cresol	−5.3	−5.0	−4.8	−5.7	−5.3	−6.0	−5.5	−5.1
p-Cresol	−5.5	−4.7	−4.7	−5.3	−5.1	−5.4	−5.4	−4.8

Overall trend: The catechol (ortho-diphenol) scaffold is a structural prerequisite for high-affinity binding.

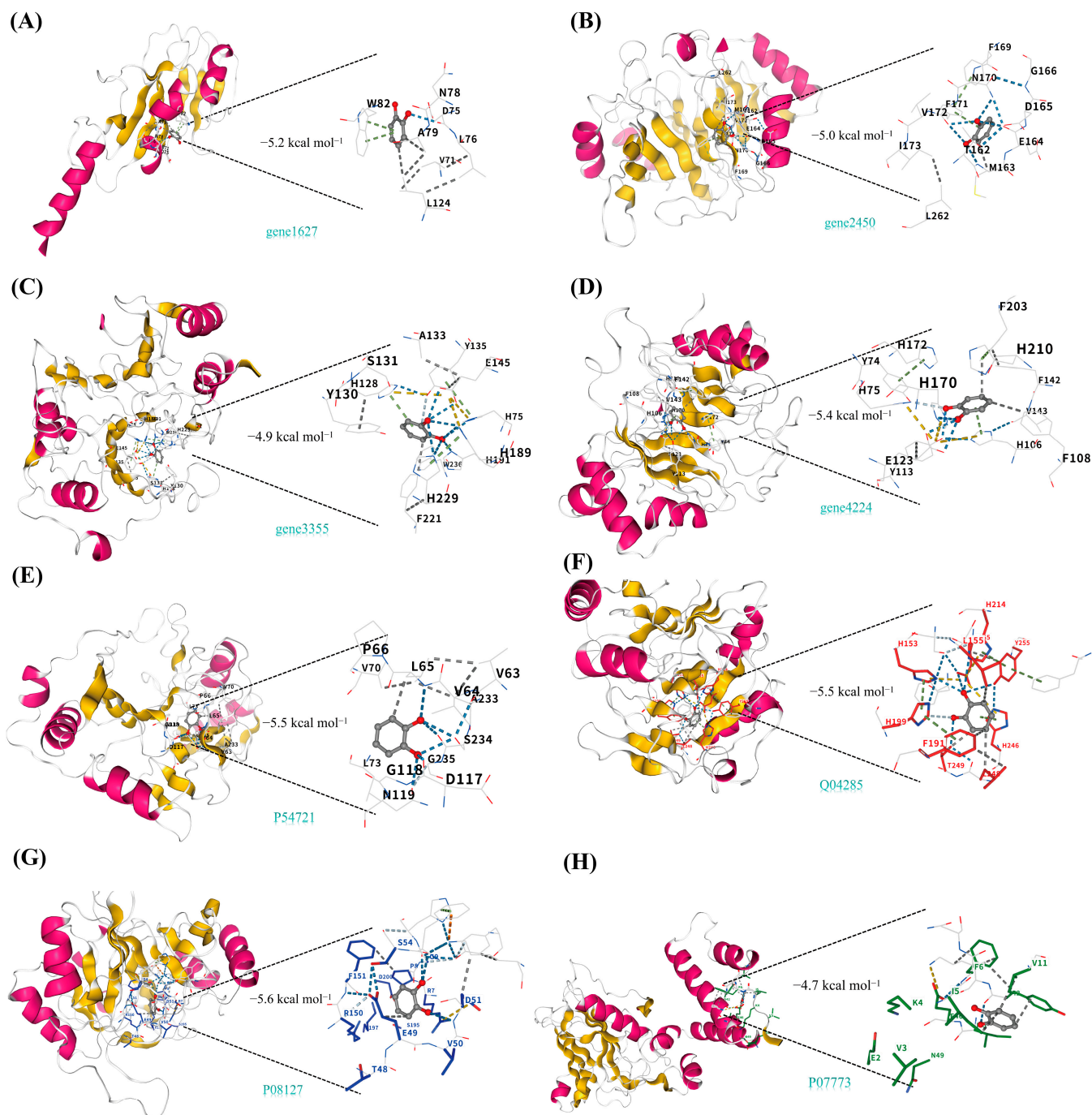


Figure 4. Binding conformations and interaction diagrams of catechol with the four target proteins (A–D) and reference enzymes (E–H).

All tested enzymes (including reference enzymes) showed their lowest binding energies toward benzene (-3.5 to -4.5 kcal/mol) and phenol (-4.5 to -4.9 kcal/mol), indicating that these enzymes are not broad-spectrum aromatic-adsorbing proteins; their high-affinity binding strictly depends on the catechol (ortho-diphenol) core scaffold or its substituted derivatives.

(1) Stepwise changes in binding energy: Non-hydroxylated benzene (average ~ -4.0) $\rightarrow$ monohydroxylated phenol (average ~ -4.7), with a binding energy improvement of

~0.7 kcal/mol; phenol → 4-methylcatechol (average ~−5.6), with a further increase of 0.9–1.0 kcal/mol. This stepwise trend clearly demonstrates that the coordination anchoring of the two hydroxyl groups (ortho-diphenol) to the active-site metal ion is the primary source of binding energy; monohydroxylated or non-hydroxylated aromatic rings cannot be efficiently captured.

Overall, all tested enzymes showed more favorable binding toward catechol-scaffold derivatives (−4.8 to −6.0) than toward monohydroxylated phenols (−4.5 to −4.9) and non-hydroxylated benzene (−3.5 to −4.5), consistent with the essential requirement of the ortho-diphenol structure for high-affinity binding.

(2) Comparison among native genes: Quantitative evidence for functional divergence.

Among the four native genes, gene4224 exhibited the most broad-spectrum and robust binding capability, with the lowest (most negative) binding energy for 9 out of 12 tested substrates, reaching the table-wide maximum (−6.0) for trichlorophenol, suggesting broader adaptability to complex multi-substituted pollutants.

gene1627 showed a relative preference for methyl-substituted compounds, with binding energies of −5.3 to −5.5 for cresols and the highest binding energy among the four for benzene (−4.5), suggesting its pocket has good basal tolerance for non-polar small molecules.

In contrast, gene2450 and gene3355 showed clear selective disadvantages: gene2450 had lower binding energies toward polychlorinated compounds (trichlorophenol −4.8, dichlorophenol −4.9), while gene3355 had a binding energy of only −3.5 for unsubstituted benzene, differing by 1.0 kcal/mol from gene1627 (−4.5), indicating the strongest steric exclusion of flat small molecules at its active-site entrance.

(3) Cross-comparison with reference enzymes: Energetic boundaries between meta and ortho cleavage.

Compared with the classical meta-cleavage reference enzymes (P54721, Q04285, P08127), gene4224 approached or matched the computed binding levels for five complex substrates (4-methylcatechol, o-/m-cresol, pentachlorophenol, trichlorophenol), particularly for halogenated substrates, suggesting that its binding pocket may have acquired broader-spectrum adaptability.

Compared with the ortho-cleavage control P07773, all native genes showed more favorable binding on halogenated/nitro-substituted substrates (0.5–1.2 kcal/mol). This predicted quantitative gap defines the differential affinity landscape of meta- versus ortho-cleavage enzymes toward aromatic substrates and provides computational support consistent with the evolutionary branch assignments of the two enzyme classes in Figure 3.

2.9. HelixVS Virtual Screening Identifies the Natural Product Binding Profile of gene4224

To explore the potential binding capability of gene4224 toward complex natural products and to validate the substrate preference conclusions derived from simple derivatives, we performed virtual screening using the HelixVS platform against 212 C7–C9 benzene-containing compounds from The Natural Products Atlas database. Using the SWISS-MODEL homology structure of gene4224 (GMQE = 0.96) as the receptor, with all active-site residues retained, the DTI_score distribution of screened compounds (post_screening = True, consecutive_ring = True) ranged from 16.12 to 34.19, and qvina_scores ranged from −7.6 to −4.7 kcal/mol. To further analyze high-confidence binding molecules, we selected compounds with DTI_score ≥ 28 (Top 10%) for structural clustering and binding mode analyses. The structural features and binding energy distributions of top-ranking compounds are shown in Table 6. A positive correlation between DTI_score and qvina_score was observed: although the two scoring algorithms differ (the former being a composite affinity score and the latter a Vina energy term), the

overall trends were consistent. For example, C₈H₈N₂O₄ (qvina = −7.6) and C₉H₈O₅ (lig_23, qvina = −7.1) were both high-scoring representatives. Substituent diversity was evident: top-ranking molecules encompassed nitro (C₉H₈N₂O₇, −6.1), halogen (C₈H₉BrO₃, −6.5; C₈H₇ClN₂O₅, −6.9), lactone ring (C₉H₈O₅, lig_19, −6.7; C₉H₈O₅, lig_23, −7.1), amide/urea (C₈H₈N₂O₄, −7.6), and various other functional groups, validating the broad-spectrum compatibility of the gene4224 (Figure 5) active pocket toward multi-site substituted groups.

Table 6. Key parameters of representative high-ranking compounds.

Ligand	Molecular_Formula	DTI_Score ^a	Qvina_Score (kcal/mol) ^b	Structure
lig_23	C ₉ H ₈ O ₅	34.19	−7.1	
lig_59	C ₉ H ₁₂ O ₄	34.14	−6.2	
lig_35	C ₉ H ₈ N ₂ O ₇	33.54	−6.1	
lig_190	C ₇ H ₅ O ₄	32.91	−6.0	
lig_112	C ₈ H ₉ BrO ₃	32.87	−6.5	
lig_136	C ₈ H ₇ ClN ₂ O ₅	32.52	−6.9	
lig_19	C ₉ H ₈ O ₅	32.48	−6.7	
lig_122	C ₈ H ₈ N ₂ O ₆	31.21	−7.4	
lig_123	C ₈ H ₈ N ₂ O ₄	30.77	−7.6	

^a DTI_score is a unitless composite affinity score produced from a deep learning model; higher values indicate more promising predicted interactions. ^b qvina_score is the binding energy (kcal/mol) estimated by QuickVina2; more negative values indicate stronger predicted binding affinity.

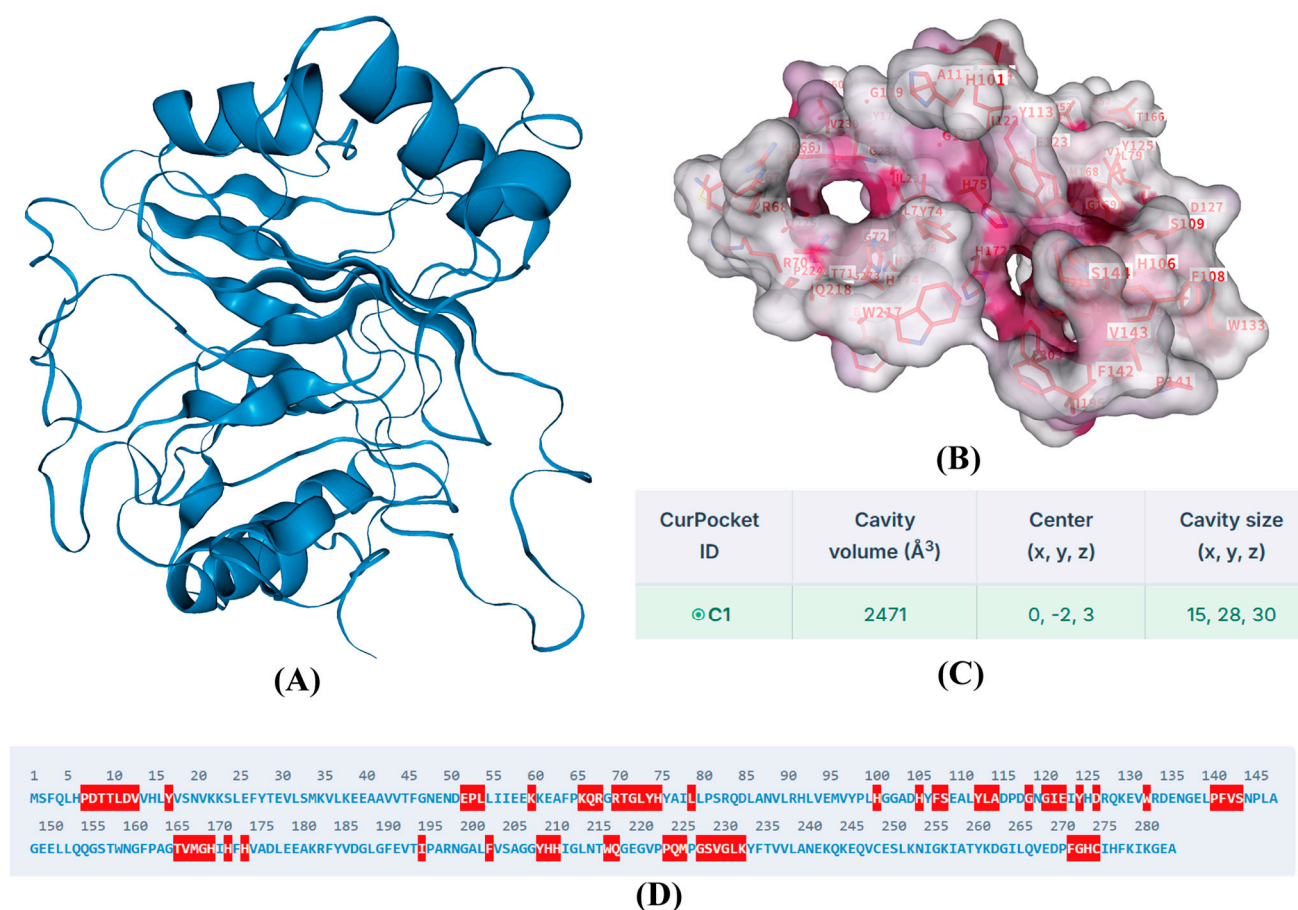


Figure 5. Structural characterization and active-site analysis of the gene4224 protein. (A) Overall three-dimensional structure of gene4224 generated by homology modeling, shown in cartoon representation. (B) Surface view of the substrate-binding pocket, illustrating the pocket shape and electrostatic potential distribution. (C) Location of the binding pocket within the global protein structure (indicated by a semitransparent surface). (D) Detailed view of the key amino acid residues lining the active pocket, with coordinating and substrate-interacting residues labeled, the red parts indicate the amino acids involved in cavity design.

3. Discussion

3.1. Structural Basis of Functional Divergence Within the VOC Family

The core finding of this study is that four VOC family proteins sharing identical COG annotations exhibit significant functional divergence in substrate binding capability, and this divergence is highly consistent with their KO annotations. Specifically, gene3355 and gene4224 (KO-annotated as C23O) showed high binding affinities toward catechol and various aromatic derivatives (binding energies -4.9 to -6.0 kcal/mol), whereas gene1627 and gene2450 (KO-annotated as glyoxalases), although retaining some binding capacity toward catechol (-4.9 to -5.0 kcal/mol), exhibited considerably more limited binding capabilities toward complex substituted substrates, especially polychlorinated compounds and methyl isomers.

This finding challenges the common assumption that “COG annotation equates to biochemical function.” COG annotation is an evolutionary classification system based on sequence homology [13], and its “catechol 2,3-dioxygenase” entry actually encompasses all VOC superfamily members possessing the $[\beta\alpha\beta\beta]$ domain, rather than providing precise enzymatic functional annotation. By contrast, KO annotation, through providing finer classification of orthologous gene clusters, can distinguish subfunctions such as C23O (K07104) and glyoxalase (K01759). The “COG-consistent, KO-divergent” phenomenon

observed in this study is a typical manifestation of the “structurally conserved, functionally diverse” evolutionary feature of the VOC family.

From a structural biology perspective, the reason why the four proteins retain similar binding capacities toward catechol (-4.9 to -5.4) lies in their shared conserved fold characteristic of the VOC family, which includes the His residues required for iron coordination and the general framework of the substrate-binding pocket. However, key residue substitutions at the entrance of the active pocket in gene1627 and gene2450 lead to significantly reduced recognition of non-polar/hydrophobic substituents. Specifically, the arrangement of hydrophobic residue clusters in gene1627 renders the pocket reasonably compatible with methyl substituents, as supported by the active-site conformational analysis in Figure 4, but bulky substituents such as those in polychlorinated compounds may cause steric clashes. This is reflected in the pentachlorophenol binding energy of only -5.2 (a limited improvement over catechol at -4.9), whereas gene2450 has smaller or slightly more polar residues at the corresponding positions, further reducing its affinity for benzene (-4.0) and simple phenols. For gene3355, the active entrance is relatively constricted, rendering unsubstituted benzene (-3.5) almost unable to bind efficiently—the most extreme case of “selective exclusion” among the four proteins.

Notably, although gene1627 and gene2450 have KO annotations as glyoxalases (the substrate of which is methylglyoxal rather than aromatic compounds), our molecular docking results show that they still possess non-zero binding capacities toward catechol and its derivatives. This phenomenon is not contradictory but rather reflects the “functional pleiotropy” or “substrate promiscuity” of VOC family proteins: a single fold can accommodate multiple substrates with different chemical types, but under physiological conditions [17], only one type of reaction is catalyzed. Gene1627 and gene2450 may have evolved from C23O toward glyoxalase function, while retaining partial affinity toward aromatic substrates—providing computational evidence consistent with the evolutionary plasticity of the VOC family.

3.2. *gene4224: Molecular Mechanisms and Application Potential of Broad-Spectrum Substrate Binding*

Among the four native genes, gene4224 showed the broadest predicted binding profile in our docking analysis, with the lowest computed binding energy for 9 out of 12 tested substrates. Its binding energies toward trichlorophenol (-6.0) and various complex natural products (qvina_score up to -7.6) were more favorable than those of the other three native genes and some reference enzymes.

The structural basis for this broad-spectrum binding capability can be attributed to the following three features: (1) An intact iron coordination environment: gene4224 retains the VOC family C23O coordination residues, ensuring stable anchoring of ortho-diphenols. (2) Synergistic polar network: gene4224 possesses a flexible hydrogen-bonding network in its active pocket that can undergo conformational fine-tuning in response to the electronegativity and steric volume of substrate substituents, enabling “adaptive” recognition—this is key to maintaining high affinity toward polar substituents such as nitro ($-\text{NO}_2$), halogens ($-\text{Cl}$, $-\text{Br}$), and carbonyl ($\text{C}=\text{O}$). (3) Tolerant hydrophobic core: the hydrophobic residues provide a moderately sized anchoring platform that can accommodate both small substituents like methyl and larger substrates such as pentachlorophenol through side-chain adjustments.

The large-scale HelixVS screening results further validated this mechanism. Among the 212 C7–C9 natural products, all high-ranking molecules ($\text{DTI_score} \geq 28$) possessed a catechol or pyrogallol core, with diverse additional substituents (nitro, halogen, lactone, amide, alcoholic hydroxyl, etc.). This finding has important application implications: gene4224 is capable not only of binding industrial pollutants (chlorinated phenols) but also

of recognizing complex natural products, suggesting that it may contribute to the strain's ecological adaptation to mixed aromatic compound environments through broad-spectrum substrate binding.

From a biocatalysis application perspective, gene4224 is the most promising candidate for enzyme engineering among the four genes. Its predicted substrate range encompasses simple catechol, methyl/chloro/nitro-substituted compounds, and even natural lactones, providing a promising starting point for developing multifunctional biocatalysts for degrading mixed pollutants or synthesizing high-value intermediates. Future studies should focus on: (1) site-directed mutagenesis to optimize the active-site entrance and further improve binding to low-affinity substrates (benzene, phenol); (2) chimera construction with reference enzymes such as P54721 [18] to explore optimal combinations of synergistic hydrophobic–polar networks; and (3) in vitro enzymatic characterization to validate computationally predicted catalytic efficiencies.

3.3. Evolutionary Co-Ordination Between Phylogenetic Branches and Substrate Profiles

Our results clearly reveal a co-evolutionary relationship between phylogenetic topology (Figure 3) and substrate binding capability (Table 5). The “*Bacillus*-type meta-cleavage functional cluster” (Clade II), containing gene3355 and gene4224, clusters tightly with P54721; members of this branch show broad compatibility with various substituents. In contrast, although the “*Pseudomonas*-type meta-cleavage functional cluster” (Clade I) containing gene1627 and gene2450 is phylogenetically homologous to Q04285 [19]/P08127 [20], it exhibits significant narrowing (especially for gene2450 and gene3355 on specific substrates).

This incomplete evolutionary–functional match suggests that phylogenetic tree topology reflects the evolutionary history of the entire protein sequence [21], while substrate binding capability is more determined by a limited number of key residues in the active pocket [22,23]. When a few key residues mutate, the biochemical function may have already substantially diverged, even when overall sequence similarity remains high [24]. This finding has methodological implications for mining new enzymes from genomic data: one cannot rely solely on phylogenetic branching or a single annotation system (COG or KO); structural biology tools such as homology modeling and molecular docking should be integrated for comprehensive assessment. Compared to the well-characterized *Bacillus* C230 P54721, gene4224 exhibited more favorable predicted binding for trichlorophenol ($\Delta G = -6.0$ vs. -5.6 kcal/mol) and comparable or more favorable scores on most substituted catechols, suggesting that gene4224 may present with expanded substrate tolerance.

3.4. Performance Benchmarking Against Homologous Reference Enzymes

Cross-comparison of the four native genes with classical reference enzymes allows for more precise positioning of their predicted binding profiles within the known enzymatic landscape. Within the *Pseudomonas*-type branch, Q04285 (XylE) shows an extreme preference for m-cresol (-6.0), consistent with its physiological role in the upstream toluene degradation pathway. P08127 (NahH) shows high affinity for nitrophenol (-5.9) and 4-chlorocatechol (-5.7), consistent with its known involvement in the naphthalene/salicylate degradation pathway. Although gene1627 performed well on methyl-substituted compounds, its binding to nitro/chloro-substituted compounds was noticeably weaker than the two reference enzymes, suggesting that it is more biased toward natural alkylated substrates.

Within the *Bacillus*-type branch, P54721 (CatE) showed high affinity for pentachlorophenol (-6.0) and 4-methylcatechol (-5.9), consistent with the physiological function of *Bacillus subtilis* in degrading complex organic matter in soil environments. gene4224 reached or exceeded the levels of P54721 on most complex substrates, particularly

excelling in binding natural products (HelixVS qvina_score up to -7.6), suggesting that gene4224 may have evolved substrate-binding capabilities that surpass its homologous reference enzyme. We caution that any evolutionary interpretations regarding adaptation or selection pressures remain speculative and require further genomic or experimental evidence for support.

The ortho-cleavage control P07773 [25] (CatA) was at a disadvantage for all 12 substrates (differences of 0.5 – 1.2 kcal/mol), the differential affinity landscape between meta- and ortho-cleavage enzymes toward aromatic substrates. Comparison with ortho-cleavage control P07773 shows a general gap that is supportive of the meta/ortho classification but does not constitute biochemical evidence of cleavage regioselectivity.

3.5. Limitations and Future Directions

While the computational predictions await experimental validation, the strong consistency between phylogenetic clustering, KO annotation, and substrate-binding profiles across multiple independent docking analyses supports the robustness of our functional assignments. Future studies will focus on heterologous expression and kinetic characterization of the top candidate gene4224.

4. Materials and Methods

4.1. Reagents and Consumables

Bacillus medium (HB8786, Haibo Biotechnology Co., Ltd., Qingdao, China), Petri dishes, anhydrous ethanol, a LiChrospher[®] RP-18 HPLC column ($5\ \mu\text{m}$ particle size, $L \times \text{I.D.}$ $25\ \text{cm} \times 4.6\ \text{mm}$; Sigma-Aldrich, St. Louis, MO, USA), acetonitrile (HPLC grade), methanol (HPLC grade), catechol (Sigma-Aldrich, St. Louis, MO, USA, C9510-100g), Tris buffer, $0.2\ \mu\text{m}$ organic syringe filters, syringes, 2 mL autosampler vials, and physiological saline were used in the experiments.

4.2. Isolation and Screening of the Catechol-Degrading Strain

Fresh bovine fecal samples (100 g) were weighed and thoroughly extracted with sterile physiological saline, followed by shaking at $37\ ^\circ\text{C}$ and 100 r/min for 1 h to release endogenous microorganisms. Gradient dilution and spread-plate techniques were employed, and the diluted extracts were plated on a Bacillus screening medium containing 50 mg/L catechol and incubated at constant temperature for 24 h. Based on colony morphology and size differences, 10 morphologically distinct single colonies were preliminarily isolated and subjected to liquid enrichment culture. After enrichment, re-screening was performed using the streak-plate method. Each purified single colony was inoculated into Bacillus liquid medium containing 50 mg/L catechol, and the catechol-degrading ability of each strain was determined. The strain with optimal degradation performance, achieving $88.9\% \pm 2.1\%$ catechol degradation within 72 h (mean $\pm$ SD, $n = 3$ biological replicates), was selected. Uninoculated controls and heat-killed cell controls showed no significant degradation, confirming that degradation was biologically mediated. The target strain was submitted to the Agricultural Microbiology Center of the China Committee for Culture Collection of Microorganisms for polyphasic taxonomic identification, and whole-genome sequencing and subsequent bioinformatics analyses were entrusted to Shanghai Majorbio Bio-Pharm Technology Co., Ltd (Shanghai, China).

4.3. Catechol Detection Method

Catechol concentrations were determined using a high-performance liquid chromatography (HPLC) system (Agilent Technologies, Santa Clara, CA, USA) equipped with a LiChrospher[®] RP-18 HPLC column ($5\ \mu\text{m}$ particle size, $L \times \text{I.D.}$ $25\ \text{cm} \times 4.6\ \text{mm}$). The

mobile phase consisted of methanol (A) and 0.1% phosphoric acid solution (B) at a ratio of 20:80 (*v/v*). The flow rate was 0.8 mL/min, column temperature was 30 °C, injection volume was 10 µL, detection wavelengths were 210 nm and 280 nm, and total run time was 15 min. Under these conditions, the retention time of catechol was 5.8 min. Calibration curves were linear ($R^2 = 0.983$) over the range of 0.5–100 mg/L.

4.4. Molecular Biological Characterization

4.4.1. Primer Design for 16S rRNA Gene Analysis of Strain HHY919

The 16S rRNA gene sequence was determined using the universal bacterial primers 27F (5'-GTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTACGACTT-3').

4.4.2. Primer Design for *gyrB* Gene Sequence Analysis of Strain HHY919

The *gyrB* gene was amplified using the universal bacterial primers UP-1 (5'-GAAGTCATCATGACCGTTCTGCAYGCNNGGNGGNAARTTYGA-3') and UP-2r (5'-AGCAGGATACGGATGTGCGAGCCRTCNCRCNCGRTCNGTCAT-3').

4.4.3. Morphological and Staining Characteristics

The purified strain HHY919 was streaked onto TSA plates and incubated at 30 °C for 24 h. Colony morphology (shape, elevation, margin, surface wrinkles, color, etc.) was observed macroscopically and photographed. Strain HHY919 was subjected to Gram and spore staining, and its cell morphology, Gram reaction, and spore formation were observed under a light microscope and photographed.

4.4.4. Physiological and Biochemical Characterization

The API 50CH test kit was used to determine the strain's ability to utilize 49 carbohydrates. Incubation was carried out at 30 °C, and the results for each carbon source were recorded as positive (+), weakly positive (w), or negative (−) according to the manufacturer's instructions. Strain HHY919 was cultured in TSB liquid medium to stationary phase, and cells were harvested by centrifugation. Fatty acids were extracted, saponified, and methylated following the MIDI standard protocol and fatty acid profiles were analyzed by gas chromatography and compared with typical *Bacillus* fatty acid characteristics for chemotaxonomic classification. (Fatty acid analysis and API 50CH tests were entrusted to the Agricultural Culture Collection of China (ACCC)).

4.5. Whole-Genome Sequencing and Analysis

High-quality genomic DNA was extracted from strain HHY919 using a bacterial genomic DNA extraction kit, and a paired-end library with an insert size of approximately 350 bp was constructed. Whole-genome sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) (PE150 mode), yielding approximately 100× genome coverage. Raw reads were quality-filtered using Trimmomatic (version 0.39) (removing adapters and low-quality bases with $Q < 20$), and de novo assembly was performed using SPAdes v3.15. The final assembly comprised 5,657,846 bp with a GC content of 35.26%, 5567 predicted genes, and one plasmid. Gene prediction was performed using Prodigal, and functional annotation was conducted against the NR, COG, KEGG, and CAZy databases to mine genes involved in carbohydrate metabolism and aromatic pollutant degradation and to dissect the genetic basis of catechol degradation in this strain. The genome assembly has been deposited in the Genome Warehouse (GWH) at the National Genomics Data Center under BioSample accession number SAMC8345804.

4.6. Target Gene Identification and Sequence Analysis

From the whole-genome annotation results, genes with COG descriptions containing “catechol 2,3-dioxygenase” were selected, yielding four target genes. Multiple sequence alignments of the four target protein amino acid sequences were performed using Clustal Omega (version 1.2.4; Conway Institute, University College Dublin, Dublin, Ireland), with emphasis on the conservation and variability of active-site key residues (iron-coordinating residues and substrate-binding pocket residues). A phylogenetic tree was constructed using the Neighbor-Joining (NJ) method in MEGA 11 (jointly developed by Temple University, USA, and Tokyo Metropolitan University, Japan), with bootstrap support calculated from 1000 replicates. The sequence alignment covered 298 amino acid positions, and gaps were treated by partial deletion with a 95% site coverage cutoff. To confirm topology stability, the NJ tree was cross-verified with a maximum-likelihood (ML) analysis using the LG model; both methods yielded identical branching patterns for the core clades. The reference set included experimentally characterized C23Os from *Pseudomonas* (XylE, Q04285; NahH, P08127), *Bacillus* (CatE, P54721), other bacterial C23Os, glyoxalases, and an ortho-cleavage catechol 1,2-dioxygenase (CatA, P07773) as an outgroup. Accession numbers for all reference sequences are provided in Supplementary Table S2.

4.7. Homology Modeling of Three-Dimensional Protein Structures

Three-dimensional structural models of the target proteins (gene1627, gene2450, gene3355, and gene4224) were constructed using homology modeling via the SWISS-MODEL online server (<https://swissmodel.expasy.org>; accessed on 2 August 2026). Model quality was evaluated primarily by GMQE (Global Model Quality Estimation) and sequence identity. GMQE scores range from 0 to 1, with higher scores indicating greater expected model accuracy, confirming that the constructed protein models were structurally reliable for subsequent molecular docking analyses. The catalytic metal center in all four homology models is Fe^{2+} (ferrous iron), consistent with the metal requirement of the template structures (P54721, Q04285, and P08127), which are well-characterized Fe^{2+} -dependent catechol 2,3-dioxygenases.

Reference protein structures (P54721, Q04285, P08127, P07773) were downloaded from UniProt (<https://www.uniprot.org/>; accessed on 2 August 2026).

4.8. Ligand Preparation

Catechol and 11 of its derivatives (4-methylcatechol, 4-chlorocatechol, phenol, benzene, o-/m-/p-cresol, pentachlorophenol, nitrophenol, dichlorophenol, trichlorophenol) were used as ligand molecules in this study. Ligand structure files were obtained from the PubChem database and uploaded to the CB-Dock3 platform in SDF format. Docking parameters were set to server defaults. For each protein–ligand combination, multiple docking poses were generated, and the pose with the lowest (most negative) binding energy was selected as the optimal binding mode for subsequent analysis. More negative binding energies indicate stronger binding affinities. We explicitly note that docking scores were used as relative binding affinity indicators and that differences of <1.0 kcal/mol are within the typical error range of empirical scoring functions. The results should be interpreted as trends rather than precise affinity rankings.

4.9. Molecular Docking

To evaluate the molecular recognition mechanisms and binding affinities between each protein and aromatic substrates, protein–ligand molecular docking was performed using the CB-Dock3 online server (<https://cadd.labshare.cn/cb-dock3/>; accessed on 3 August 2026). The CB-Dock3 docking workflow includes four main stages: (1) automated domain

detection—the server automatically analyzes the target protein structure and identifies and segments domains, with the user able to choose docking against the full protein or a specific biological domain; (2) dual-mode binding site identification—the system identifies potential binding pockets by detecting surface concavities using the curvature-based cavity detection algorithm CurPocket while also searching the BioLiP2 homologous template library for topologically similar ligand–protein complex templates; (3) molecular docking calculation—within the detected binding pockets, molecular docking calculations are performed using a custom box size and the AutoDock Vina (v1.2.7) docking engine; and (4) result output—the system returns binding energies (in kcal/mol) for each docking pose and provides three-dimensional binding mode visualization.

To further expand the substrate screening scope and validate the previous docking conclusions, the HelixVS virtual screening platform was also employed for large-scale screening against a natural product library using gene4224. HelixVS integrates multi-level docking precision (coarse-to-fine screening) and comprehensive scoring functions (DTI_score and qvina_score), and its screening efficacy has been validated in DUD-E benchmark tests, identifying on average 159% more active molecules than the open-source tool AutoDock Vina, with a 28-fold increase in speed. The screening database was The Natural Products Atlas (containing tens of thousands of bacterial/fungal natural products), and 212 C7–C9 benzene-containing compounds were selected for screening. Screening parameters were set to platform defaults, with DTI_score ≥ 28 used as the high-confidence hit threshold.

4.10. Data Analysis

16S rRNA and gyrB gene sequence homology searches were performed using the EzBioCloud and NCBI databases, respectively [10]. Phylogenetic trees were constructed using the neighbor-joining method in MEGA 11 software (jointly developed by Temple University, USA, and Tokyo Metropolitan University, Japan), with 1000 bootstrap replicates to test branch support. Whole-cell fatty acid profiles were quantified using the MIDI system software, and carbon source utilization characteristics were recorded based on API 50CH color reactions. Raw whole-genome sequencing reads were quality-controlled with FastQC, assembled with SPAdes, and functionally annotated against NR, COG, KEGG, and CAZy databases. Minimum binding energies (ΔG , kcal/mol) were extracted from AutoDock Vina output results as the primary metric for binding affinity.

5. Conclusions

Based on the experimental and computational results, the following conclusions can be drawn: (1) Experimentally, strain HHY919 was identified as *Bacillus thuringiensis* and degraded 88.9% of 50 mg/L catechol within 72 h, confirming its aromatic-compound-degrading capacity. (2) Computationally, the four VOC genes share structural folds but differ in predicted substrate-binding profiles. gene3355 and gene4224 (KO-annotated as C23O) exhibited binding scores consistent with meta-cleavage reference enzymes, while gene4224 exhibited the broadest and most favorable predicted affinity, particularly toward halogenated and complex natural products. (3) Therefore, gene4224 is recommended as a top candidate for future experimental validation as a C23O enzyme and for potential application in bioremediation. gene1627 and gene2450, despite retaining catechol-binding capacity, showed narrower profiles consistent with their glyoxalase annotation. (4) This study demonstrates that integrating COG/KO annotations with homology modeling and docking provides a useful computational pipeline for resolving functional ambiguity in structurally conserved enzyme families and for generating testable hypotheses for experimental biology.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/catal16090811/s1>, S1: Similarities to other type strains; S2: Protein modeling; S3: C7–C9 substances containing benzene rings; S4: Helix result.

Author Contributions: The manuscript was written with contributions from all authors. H.H.: conceptualization; W.Z. (Wenjie Zhang): methodology; W.Z. (Weibing Zhang): software; Y.Z.: validation; Z.W.: formal analysis, writing—original draft, investigation, project administration, funding acquisition; S.G.: resources; L.Y.: data curation; W.Z. (Wei Zhang): writing—review and editing; visualization; F.R.: supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was Supported By Science and Technology Program of Xinjiang Uyghur Autonomous Region (2025LQ02006) and the Earmarked Fund for XJARS (No. XJARS-09-12).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the non-invasive nature of the sampling. Samples consisted of manure collected from the ground after natural defecation; because no animal handling or intervention was required, the study meets the Institutional Animal Care and Use Committee (IACUC) criteria for exemption.

Data Availability Statement: All detailed data are provided in the attachment.

Acknowledgments: We would like to extend our heartfelt gratitude to all of the unlisted teachers and classmates who generously contributed their time and expertise to assist in our experiments. Their invaluable support was instrumental to our success.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yabalak, E.; Aminzai, M.T.; Gizir, A.M.; Yang, Y. A Review: Subcritical Water Extraction of Organic Pollutants from Environmental Matrices. *Molecules* **2024**, *29*, 258. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Souaf, B.; Methneni, N.; Toumi, D.; Albergamo, A.; Beltifa, A.; Turco, V.L.; Hassani, R.; Fathallah, S.; Mansour, H.B.; Bella, G.D. Persistent organic pollutants in abiotic and biotic matrices from an anthropized area: Investigation of their occurrence and associated health risks. *Mar. Pollut. Bull.* **2025**, *218*, 118071. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Sarah, P.; Silvia, G.; Alina, P.; Rossano, P.; Giorgio, B.L.; Andrea, G.; Luca, G. PAHs, PCBs, PBDEs, and OCPs trapped and remobilized in the Lake of Cavazzo (NE Italy) sediments: Temporal trends, quality, and sources in an area prone to anthropogenic and natural stressors. *Environ. Res.* **2022**, *213*, 113573. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Mateescu, C.; Lungulescu, E.M.; Nicula, N.O. Effectiveness of Biological Approaches for Removing Persistent Organic Pollutants from Wastewater: A Mini-Review. *Microorganisms* **2024**, *12*, 1632. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Patel, A.B.; Shaikh, S.; Jain, K.R.; Desai, C.; Madamwar, D. Polycyclic Aromatic Hydrocarbons: Sources, Toxicity, and Remediation Approaches. *Front. Microbiol.* **2020**, *11*, 562813. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Sahith, V.N.; Kumar, J.A.; Sruthi, V.S.; Sundararaman, S.; Prabu, D.; Venkatesan, D.; Renita, A.A. Bioremediation of polycyclic aromatic hydrocarbons contaminated soils/water for environmental remediation. *Biodegradation* **2025**, *37*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Liu, B.; Ge, W.; Liu, X.; Zhang, X.; Chen, Q.; Wu, J.; Li, Z.; Chai, C. Bioremediation of polycyclic aromatic hydrocarbons from an aged contaminated agricultural soil using degrading bacteria and soil amendments. *Biorem. J.* **2022**, *26*, 305–317. [\[CrossRef\]](#)
8. Olukunle, O.F. Molecular Identification of Crude Oil-Degrading Bacteria and Screening for Catechol 2, 3 Dioxygenase (C23O) Gene. *Biotechnol. J. Int.* **2020**, *23*, 1–14. [\[CrossRef\]](#)
9. Ghosh, S.; Qureshi, A.; Purohit, H.J. Enhanced expression of catechol 1,2 dioxygenase gene in biofilm forming *Pseudomonas mendocina* EGD-AQ5 under increasing benzoate stress. *Int. Biodeterior. Biodegrad.* **2017**, *118*, 57–65. [\[CrossRef\]](#)
10. Wang, Z.; Sun, J.; Yang, P.; Zhang, W.; Jiang, Y.; Liu, Q.; Yang, Y.; Hao, R.; Guo, G.; Huo, W.; et al. Molecular Analysis of Indole and Skatole Decomposition Metabolism in *Acinetobacter piscicola* p38 Utilizing Biochemical and Omics Approaches. *Microorganisms* **2024**, *12*, 1792. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Colabroy, K.L.; Horwitz, A.D.; Basciano, V.R.; Fu, Y.; Travitz, K.M.; Robinson, M.K.; Shimanski, B.A.; Hoffmann, T.W. A New Way of Belonging: Active-Site Investigation of L-DOPA Dioxygenase, a VOC Family Enzyme from Lincomycin Biosynthesis. *Biochemistry* **2019**, *58*, 4794–4798. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Duffner, F.M.; Müller, R. A novel phenol hydroxylase and catechol 2,3-dioxygenase from the thermophilic *Bacillus thermoleovorans* strain A2: Nucleotide sequence and analysis of the genes. *FEMS Microbiol. Lett.* **2006**, *161*, 37–45. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Galperin, M.Y.; Kristensen, D.M.; Makarova, K.S.; I Wolf, Y.; Koonin, E.V. Microbial genome analysis: The COG approach. *Brief. Bioinform.* **2019**, *20*, 1063–1070. [[CrossRef](#)] [[PubMed](#)]
14. Zhuang, Y.; Kim, N.Y.; Fritsche, L.G.; Mukherjee, B.; Lee, S. Incorporating functional annotation with bilevel continuous shrinkage for polygenic risk prediction. *BMC Bioinform.* **2024**, *25*, 65. [[CrossRef](#)] [[PubMed](#)]
15. Suenaga, H.; Fujihara, H. Molecular basis for adaptive evolution of aromatic degradation enzymes in bacteria revealed by metagenomics. *Front. Microbiol.* **2026**, *17*, 1795400. [[CrossRef](#)] [[PubMed](#)]
16. Wu, S.; Zhong, J.; Lei, Q.; Song, H.; Chen, S.-F.; Wahla, A.Q.; Bhatt, K.; Chen, S. New roles for *Bacillus thuringiensis* in the removal of environmental pollutants. *Environ. Res.* **2023**, *236*, 116699. [[CrossRef](#)] [[PubMed](#)]
17. Yang, L.; Ma, Z.; Cao, X.; Xu, L.; Ma, X.; Zhu, J.; Xiao, M.; Jiang, X. Molecular mechanism of the substrate promiscuity in GT1 glucosyltransferases: Insights from molecular dynamics simulations and experimental validation. *Int. J. Biol. Macromol.* **2025**, *336*, 149450. [[CrossRef](#)] [[PubMed](#)]
18. Yamamoto, H.; Uchiyama, S.; Fajar, A.N.; Ogasawara, N.; Sekiguchi, J. Determination of a 12 kb nucleotide sequence around the 76° region of the *Bacillus subtilis* chromosome. *Microbiology* **1996**, *142*, 1417–1421. [[CrossRef](#)] [[PubMed](#)]
19. Benjamin, R.C.; Voss, J.A.; Kunz, D.A. Nucleotide sequence of xylE from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7. *J. Bacteriol.* **1991**, *173*, 2724–2728. [[CrossRef](#)] [[PubMed](#)]
20. Harayama, S.; Rekik, M.; Wasserfallen, A.; Bairoch, A. Evolutionary relationships between catabolic pathways for aromatics: Conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. *Mol. General. Genet. MGG* **1987**, *210*, 241–247. [[CrossRef](#)] [[PubMed](#)]
21. Moi, D.; Bernard, C.; Steinegger, M.; Nevers, Y.; Langleib, M.; Dessimoz, C. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. *Nat. Struct. Mol. Biol.* **2025**, *32*, 2492–2502. [[CrossRef](#)] [[PubMed](#)]
22. Ma, X.; Cao, X.; Ma, Z.; Zhu, J.; Yang, L.; Xiao, M.; Jiang, X. Molecular Dynamics Simulation Reveals the Mechanism of Substrate Recognition by Lignin-Degrading Enzymes. *Int. J. Mol. Sci.* **2025**, *26*, 9378. [[CrossRef](#)] [[PubMed](#)]
23. Ma, N.; Zhu, J.; Wang, H.; Shen, B.; Qian, M.C.; Xiao, Z. Revealing the interaction mechanisms of characteristic aroma compounds in apples (*Malus × domestica*) by S-curve modeling, molecular docking, and molecular dynamics simulation. *Bioact. Carbohydr. Diet. Fibre* **2026**, *35*, 100523. [[CrossRef](#)]
24. Wright, M.; Downard, K.M. Prevalence of Prime Mass Residues in Single Point Mutations and their Importance to Protein Stability, Function and Evolution. *BioSystems* **2025**, *258*, 105638. [[CrossRef](#)] [[PubMed](#)]
25. Valérie, B.; David, V.; Nuria, F.; Annett, K.; Sophie, O.; Laurent, L.; Stéphane, C.; Catherine, R.; Simone, D.; Patrick, W.; et al. Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **2004**, *32*, 5766–5779. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.