Iranian J Biotech. 2019 December;17(4): e2250 DOI: 10.30498/IJB.2019.95299

Research Article

The *In Silico* Characterization of a Salicylic Acid Analogue Coding Gene Clusters in Selected *Pseudomonas fluorescens* Strains

Muhammad Shehroz[1], Muneeba Aslam[1], Munazza Ali Khan[1], Sara Aiman[1], Sahib Gul Afridi[2], Asifullah Khan[2*]

[1] Department of Biochemistry, Abdul Wali Khan University Mardan, Khyber Pakhtunkhwa, Pakistan
[2] Department of Biochemistry, Abdul Wali Khan University Mardan, Pakistan

[*] *Corresponding author:* Asifullah Khan, Department of Biochemistry, Abdul Wali Khan University Mardan, Pakistan. Tel: 92-937-542195, E-mail: asif@awkum.edu.pk

Background: The microbial genome sequences provide solid *in silico* framework for interpretation of their drug-like chemical scaffolds biosynthetic potentials. *Pseudomonas fluorescens* strains are metabolically versatile and producing therapeutically important natural products.
Objectives: The key objective of the present study was to mine the publically available data of *P. fluorescens* strains genomes for putative drug-like metabolites identification.
Materials and Methods: We implemented the computational biology resources of AntiSMASH and BAGEL3 for the secondary metabolites prediction from *P.* fluorescens strains genome sequences. The predicted secondary metabolites were evaluated using drug discovery chemoinformatics resources, like Drugbank database search and molecular docking inspection.
Results: The analyses unveiled a wide array of chemical scaffolds biosynthesis in different *P. fluorescens* strains. Subsequently, the drug-like potential evaluation of these metabolites identified few strains, i*ncluding P. fluorescens* PT14, *P. fluorescens* A5O6, and *P. fluorescens* FW300-N2E3 that *harbor* the biosynthetic gene clusters for salicylic acid-like metabolite biosynthesis. The molecular docking inspection of this metabolite against human cyclooxygenase and aldo-keto reductase targets revealed its feasible inhibitory potentials like other salicylate compounds.
Conclusion: The computational biology and drug discovery analyses identified different gene clusters in *P. fluorescens* genomes coding for salicylic acid-like chemotypes biosynthesis. These gene clusters may worthy to target through metabolic engineering for the massive production of salicylates-like chemical scaffolds from microbial resources.
Keywords: Bacterial genomes; *P. fluorescens*; Biosynthetic gene clusters; Secondary metabolites; Salicylic acid

1. Background

Microbes produce secondary metabolites for survival and development of ecological interactions with surrounding organisms in the environment (1, 2). These microbial secondary metabolites have wide range of applications in pharmaceutical and agricultural industries (3). The microbial secondary metabolites include non-ribosomal peptides (NRP), polyketides (PK), post-translationally modified peptides (RiPP), terpenoids, saccharides, and their hybrids. These metabolites biosynthesis is controlled by a set of genes located closely on microbial chromosome termed as Biosynthetic Gene Cluster (BGC). This organization enables the joint expression of these biosynthetic enzymes, regulators and transporters (4). The rich genetic diversity of microbial BGCs eventually brings high chemical diversity in underlying coding secondary metabolites (5, 6).

The gram-negative *Pseudomonas fluorescens* is a ubiquitous soil bacterium, harbors metabolic versatility (7, 8). The species of pseudomonads produce the fluorescent, which play essential role in promotion of plant health as well as diseases suppression (7). The species is particularly important for natural products biosynthesis as because of its versatile metabolic feature and existence of miscellaneous enzymatic machinery (9). Therefore, recombinant strains of *P. fluorescens* have been used for biosynthesis of several natural products, including antibiotics, antitumor, anti-inflammatory, antifungal, and cholesterol-lowering drugs (10).

The recent advancement in genome-level DNA sequencing approaches in the form of next-generation sequencing (NGS) enabled the researchers to easily understand genome-based ecological fitness and metabolic versatility of microbes (11). Nowadays, the robust genome-level DNA sequencing availability of microbial species and growth in bioinformatics resources greatly facilitated the discoveries of novel secondary metabolites from rational and predictive

perspectives (12). Likewise, the BGCs organization in microbial genomes in the form of single locus enable the accurate *in silico* prediction of these genomic motifs through bioinformatics principles (13). Despite of the recent developments of these resources, the identification of microbial secondary metabolites from genome-guided information is still limited compare to the bioassay screening methodologies.

2. Objectives

The key objective of the study was to mine the publicly available genomes sequences of different *P. fluorescens* strains to anticipate their drug-like secondary metabolites coding potential.

3. Materials and Methods

3.1. Genome Sequences Retrieval

The 17 complete and 6 chromosome-level genome sequences of *P. fluorescens* (available till September 2017) were acquired from NCBI resource (https://www.ncbi.nlm.nih.gov/) (14) and the database of Integrated Microbial Genome and Microbiomes (https://img.jgi.doe.gov/) (15) **(Table S1)**.

3.2 Data Mining for Potential Secondary Metabolites Identification

The prediction models implemented as bioinformatic resources were used for BGCs identification from genome datasets. These include AntiSMASH 4.0 (https://antismash.secondarymetabolites.org/#!/start) (16) and BAGEL3 (http://bagel.molgenrug.nl) (17). The AntiSMASH resource holds known secondary metabolite BGCs from the MIBiG (Minimum Information about a Biosynthetic Gene cluster) project (https://mibig.secondarymetabolites.org/) (18). The integration of additional modules like ClusterBlast and ClusterFinder enabled the AntiSMASH current version to putatively identify the unknown BGCs on the basis of similarities and homologies to previously characterized BGCs with known metabolic products (16).

3.3. Drug-like Potential Evaluation of Putative Metabolites

The chemical structures of the secondary metabolites were designed using ChemDraw (https://www.perkinelmer.com/category/chemdraw) (19). The chemical structures were converted into 3D and SMILES format to conduct additional analyses. The MOE (Molecular Operating Environment) from Chemical Computing Group Inc. Canada, 2016 (https://www.chemcomp.com/Products.htm) was employed to evaluate the drug-likeness of coding metabolites according to Lipinski's rule of five (20, 21). Lipinski's principle is commonly implemented to evaluate the drug-likeness potential of a chemical compound. The rule is important to consider during drug discovery when a biologically active lead compound is prioritized. According to Lipinski's rule the drug-like compound should not have H-bond donors >5, neither the H-bond acceptors >10. Besides, the molecular weight should be <500 Daltons and the partition coefficient of octanol-water calculated as log P should not be >5 (20, 21). The predicted secondary metabolites were evaluated for drug-like compound similarity and drug target search using Drugbank database (https://www.drugbank.ca/) (22).

3.4. Molecular Docking Analysis

The computational method of molecular docking can predict the ligands binding capabilities to receptor proteins. The analyses of molecular docking usually performed to test the molecular binding affinity of putatively identified salicylic acid analogue against salicylates drug targets (23). The MOE resource was used for this purpose. The template including complex structure of salicylate bound human COX-2 and AKR1C1 targets were acquired from the protein data bank (PDB) (https://www.rcsb.org/) under IDs, 5F1A and 3C3U, respectively. The complex 3D protonation and energy minimization was achieved by MOE for downstream analyses. The protein-ligand docking procedure implemented in the MOE was used to elucidate the molecular binding potential of ligand molecule with receptor. The binding affinity of the ligand was estimated via generalized Born/volume integral (GB / VI) method implemented in the MOE (24). The ligand binding free energy was estimated via London dG scoring method in the MOE. The receptor atoms residing in the ligand locality and binding site were retained flexible. The ligand atoms were set free to move in the binding pocket. The top docking conformations were selected based on "S" score and binding energy.

3.5. Phylogenetic Analysis

The phylogenetic analysis among different *P. fluorescens* strains was performed based on genomic variations in multilocus sequence typing (MLST) loci. This scheme defines microbial strains based on seven housekeeping genes sequences. MLST method is well-known to determine the evolutionary relationship among bacterial strains (25). The seven housekeeping genes for *P. fluorescens* strains typing are *glnS* (glutamine tRNA ligase), *gyrB* (DNA gyrase beta subunit), *ileS* (isoleucine tRNA ligase), *nuoD* (NADH-quinone oxidoreductase, C/D subunit), *recA* (recombinase A), *rpoB* (DNA-directed RNA polymerase, beta subunit), and *rpoD* (RNA

polymerase sigma factor) (26). The DNA sequences of these seven housekeeping genes were combined and Maximum Likelihood-based rooted phylogenetic tree was generated from corrected multiple sequence alignment comprises of 3,286 nucleotides characters from each strain. We utilized MEGA7 software resource (https://www.megasoftware.net/) (27) to perform this analysis. The tree with the highest log likelihood (i.e., -35342.71) is shown. Besides, the phylogenetic analysis was performed on the basis of few conserved biosynthetic proteins (i.e., enlisted in **Table S4**) of salicylic acid analogue coding BGCs. The genes sequences of these biosynthetic proteins were concatenated from their respective strains and a neighbor-joining based phylogenetic tree was constructed from multiple sequence alignment of 23,210 nucleotides characters using the MEGA7 resource (27).

4. Results

4.1. Putative Secondary Metabolites

The analyses of different strains of *P. fluorescens* genomes assemblies inferred the BGCs coding for non-ribosomal peptide synthetases (NRPs), type III polyketide synthases (PKs), bacteriocins, homoserine lactones (HSL), siderophores, phenazines, arylpolyenes, terpenes, and ectoines metabolites (Table S2; Table S3). Moreover, some hybrid secondary metabolites, i.e. originating from BGCs comprise of distinct classes of secondary metabolites coding sub-clusters were predicted. These include butyrolactone-HSL, phenazine-NRP, bacteriocins-NRP, terpene-NRP, HSL-phenzaine, siderophore-ectoine, HSL-ladderane, and HSL-NRP. Besides, several BGCs of unknown biosynthetic capabilities were predicted. A pyoverdine-coding BGC was putatively identified for all *P. fluorescens* strains genomes (**Table S2**).

4.2. Salicylic Acid Analogue Metabolite Identification

The putative secondary metabolites identified from genome mining of *P. fluorescens* strains were evaluated for necessary drug-like features using *in silico* drug discovery approaches. During this analysis, a single metabolite was found to fulfill the Lipinski's rules of five (20) (**Table 1**; Additional file 1). The scanning of DrugBank database identified this metabolite as salicylic acid analogue and its corresponding BGCs were characterized for genome sequences of *P. fluorescens* A5O6, *P. fluorescens* PT14, and *P. fluorescens* FW300-N2E3 strains (**Table 2**). The BLASTN analysis against NCBI genome repository identified the existence of salicylic acid analogue coding BGC in three additional *Pseudomonas* strains with significant query coverage and sequence identity (**Table 3**). This salicylic acid analogue is coding from NRP type of BGCs annotated from AntiSMASH version 4.0 resource analysis.

Table 1: The basic drug-like features of a salicylic acid analogue metabolite identified from *P. fluorescens* strains genomes

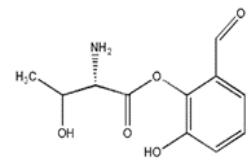
Lipinski's **rules of five values**	**Chemical scaffold of putatively identified compound**
Mol. weight:239.23 g/mol log P:0.18 Hydrogen Donor: 3 Hydrogen Accepter: 5 Toxic : No	

Table 2: The genomic position of salicylic acid analogue coding BGCs in different *P. fluorescens* strains

***P. fluorescens* strain**	**Genbank Accession number**	**Chromosome**	**Start position-End position**
A506	NC-017911.1	1	3,048,111-3,099,246
PT14	NZ-CP017296.1	1	1,942,951-1,994,085
FW300-N2E3	NZ-CP012830.1	1	5,251,222-5,323,220

Table 3: The NCBI Genbank BLASTN hits obtained for salicylic acid analogue coding BGC sequence from *P. fluorescens* A506

BLAST hit	**Strains**	**Query coverage**	**E-value**	**Identity**
1	*Pseudomonas sp.*WCS374	100%	0.0	99%
2	*Pseudomonas sp.* Bs2935	100%	0.0	99%
3	*Pseudomonas fluorescens* strain pt14	100%	0.0	99%

4.3. Molecular Docking Inspection of Putative Salicylic Analogue Against Cyclooxygenase (COX) Target

The COX (cyclooxygenase) is an enzyme responsible for the formation of thromboxane and prostaglandins. The mechanism of well-known anti-inflammatory salicylic acid such as aspirin and ibuprofen is based on pharmaceutical inhibition of COX. Aspirin acts upon both COX-1 and COX-2 isozymes to inhibit the prostaglandins formation. The aspirin covalently modifies the COX by acetylating its Ser-530 residue within active site. This eventually results in complete loss of COX-1 biological activity, while COX-2 acetylation leads to generation of monooxygenated 15(R)-hydroxyeicosatetraenoic acid (15R-HETE) (28). The analysis of molecular docking inferred the active binding of putative salicylic acid analogue within the aspirin binding pocket of human COX-2 **(Figure 1)**. Different conformations revealed that benzaldehyde group of the putative analogue compound forms molecular interactions with Ser-530, which was previously reported to affect the essential stereochemistry required for the addition of oxygen to prostaglandin end product (29). Based on this *in silico* binding potential (i.e. binding energy of -12.30Kcal. moL^{-1}), we assume that putatively identified salicylate analogue acts as a potential candidate to inactivate the COX-2 functioning as like aspirin.

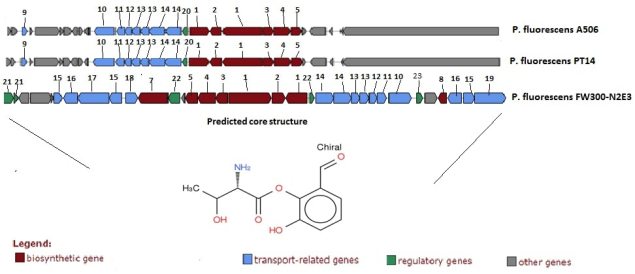


Figure 1: The putative salicylic analogue compound [(2S)-2-formyl-6-hydroxyphenyl 2-amino-3-hydroxybutanoate] identified for *P. fluorescens* strains A5O6 (cluster 5), PT14 (cluster 3) and FW300-N2E3 (cluster 9) genomes. The numbers depict different protein domains as; 1. Condensation domain containing protein 2. Lysine/ornithine N-monooxygenase 3. Isochorismate synthase 4. AMP dependent synthase and ligase 5. Decarboxylase pyridoxal dependent 7. Diguanylate cyclase 8. Short chain dehydrogenase/ reductase 9. Homoserine/threonine efflux protein 10. TonB dependent sidrophore receptor 11. Iron compound ABC transporter periplasmic 12. ABC transporter ATP binding protein 13. Transport system permease 14. ABC transporter related protein 15. RND family efflux transporter mfp subunit 16. RND efflux system outer membrane lipoprotein 17, 19. AcrBAcrD and AcrF family proteins 18. Major facilitator transporter 20. Extra cytoplasmic function sigma70-factor 21. Response regulator 22. LacI transcriptional regulator 23. TetR family transcriptional regulator.

4.4. Molecular Docking Against Aldo-keto Reductase (AKR1C1)

The human aldo-keto reductase 1C1 (AKR1C1) also called 20α-hydroxy steroid dehydrogenase plays key role in progesterone metabolism (30). The oncogenic AKR1C1 causes 20α-hydroxy steroid hyperactivity and leads to the development of endometriosis and breast cancer. Besides, the hyperactive AKR1C1 leads to a premature birth defect by conversion of active progesterone, i.e. essential for the maintenance of pregnancy, to an inactive 20α-hydroxy progesterone (31, 32). The salicylates including, 3, 5-Diiodosalicylic acid (DISA) is reported as a potent inhibitor of human AKR1C1 (32). During molecular docking analysis, the binding energies and chemical interactions pattern with key residues revealed the AKR1C1 inhibitory potential of putatively identified salicylic acid analogue like other salicylates compounds (**Figure 2; Table 4**).

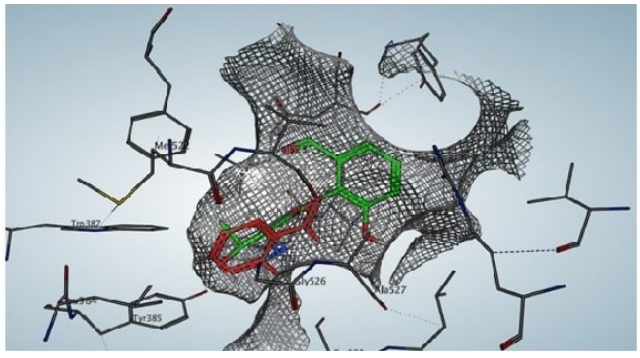


Figure 2: The molecular docking of putative salicylic acid analogue in the aspirin binding pocket of human COX-2. The putative analogue compound is shown with green legend while aspirin is depicted with red color in the binding pocket.

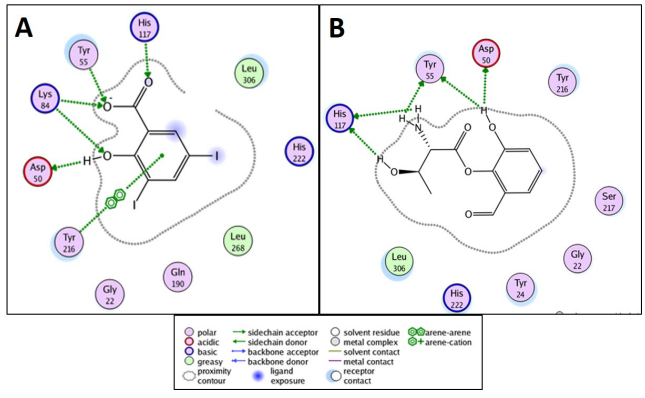
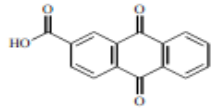
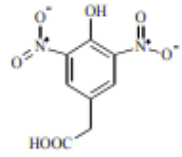
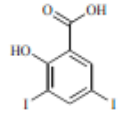
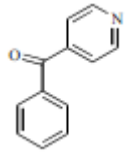
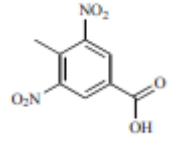
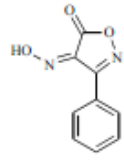
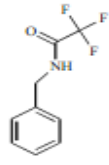
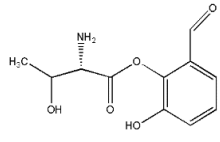


Figure 3: (A) The molecular docking poses of 3, 5-diiodosalicylic acid (DISA) and **(B)** putative analogue within the active site binding pocket of AKR1C1.

Table 4: The comparative docking results of putative salicylic analogue along with other salicylates compounds against AKR1C1 binding pocket.

S.No	Compound name	Structure	Binding energy (Kcal/mol)	Receptor-interaction residues
1	Anthraquinone-2-carboxylic acid		-11.1192	Asp50, Lys84,Tyr55
2	3,5-dinitro-4-hydroxyphenyl acetic acid		-12.5280	Tyr24, Asp50
3	3,5-diiodosalicylic acid		-10.8759	Asp50, Lys84, His117, Tyr55
4	4-benzoylpyridine		-8.1918	His117,Tyr55
5	3,5-dinitro-P-toluic acid		-10.7199	Tyr24, Asp50
6	3-phenyl-4,5-isoxazoledione-4-oxime		-9.1001	Asp50,Lys84
7	N-benzyl-2,2,2-trifluoroacetamide		-9.4832	His222
8	putative salicylic analogue compound (identified in this study)		-12.0593	Tyr55, Asp50, His117

5. Discussion

The microbial genomes mining through computational biology approaches became an important tool for drug-like novel metabolites discovery. The *in silico* analyses of available genome sequences of *P. fluorescens* strains revealed a diverse array of secondary metabolites biosynthetic potential in *P. fluorescens* strains. Few

BGCs were found to be strain-specific. For instance, the HSL-NRP coding from a hybrid BGC was detected only in *P. fluorescens* FW300-N2C3. We speculate that these strain-specific BGCs are coding for metabolites that attribute unique ecological competency to their respective *P. fluorescens* strains.

The chemoinformatic analysis results are crucial in the identification of few *P. fluorescens* strains, which harbor BGCs coding for salicylic acid like compound biosynthesis. The bio-control strains of *Acinetobacter* spp., *Pseudomonas* spp., and *Serratia* spp. produce salicylic acid to elicit the essential cellular signaling and induce plant defense mechanisms (7, 33). The orthologs distribution pattern of salicylic acid analogue coding BGCs was found similar between *P. fluorescens* A5O6 and *P. fluorescens* PT14, while in the case of *P. fluorescens* FW300-N2E3 the gene organization pattern was found somehow distinct (**Figure 3**). The AntiSMASH though predicted only the core chemical scaffold of salicylic acid analogue for these BGCs, while we speculate the biosynthesis of different salicylic acid like chemotypes with different aryl starter unit in these *P. fluorescens* strains.

The MLST analysis of salicylic coding BGCs protein domains identified the *P. fluorescens* A5O6 and *P. fluorescens* PT14 as phylogenetically closest species (**Figure 4**). In addition, the *P. fluorescens* FW300-N2E3, which harbor distinction in orthologs arrangement in salicylic acid analogue coding for BGC was found far phylogenetically from *P. fluorescens* A5O6 and *P. fluorescens* PT14 (**Figure 4**).

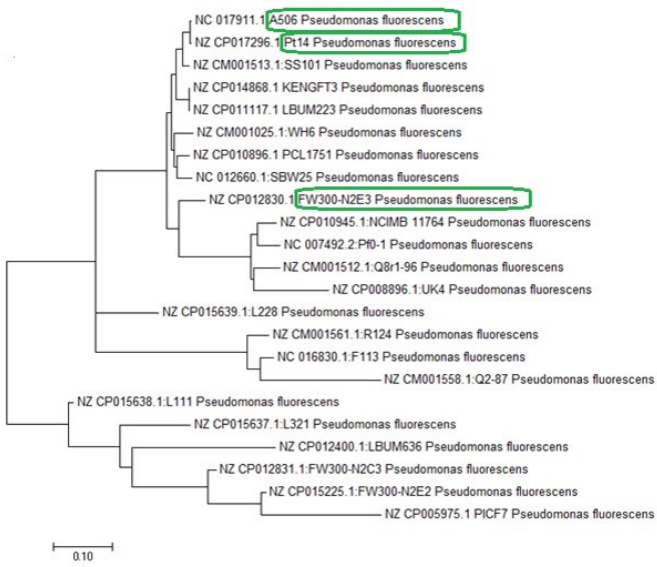


Figure 4: A phylogenetic analysis of *P. fluorescens* strains on the basis of seven housekeeping genes MLST approach. The strains boxed in green harbor BGCs for coding putative salicylic acid analogue metabolite.

A similar phylogenetic tree pattern was observed for these strains, while the analysis was performed on the basis of conserved genes sequences from their salicylic acid analogue coding for BGCs (**Fig.5; Table S4**). This speculates that different *P. fluorescens* strains share the molecular machinery for salicylic acid chemical scaffold biosynthesis. Moreover, some strains harbor linage-specific distinction in orthologs arrangement in these BGCs. This somehow reveals lineage-specific salicylic acid chemotypes biosynthesis in *P. fluorescens* strains.

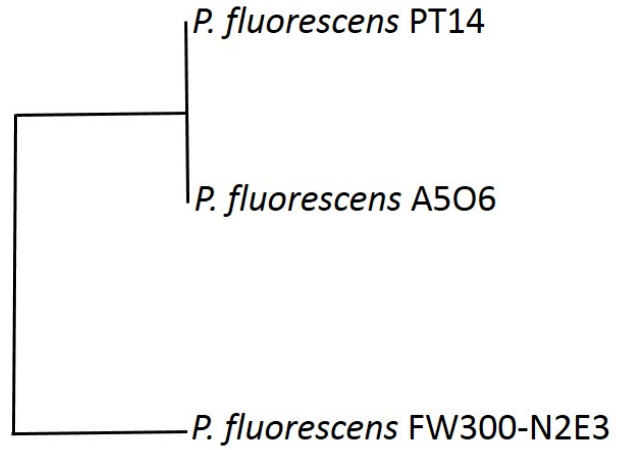


Figure 5: A phylogenetic analysis of P. fluorescens strains predicted to harbor salicylic acid coding BGCs. The analysis was based on concatenated conserved biosynthetic genes sequences from salicylic acid coding BGCs.

6. Conclusion

The genome mining for BGCs revealed that *P. fluorescens* strains encode a diverse array of secondary metabolites. The *in silico* drug discovery analysis identified few *P. fluorescens* strains, which hold molecular signatures for coding salicylic acid like metabolite. The analyses conducted here provide a solid framework for experimental characterization of salicylate coding gene clusters in *P. fluorescens* strains. We speculate that additional methodologies of metabolic engineering and strain-specific induction might be promising for the massive production of such important drug-like chemical scaffold. The salicylate-coding gene cluster characterized here can be up-regulated or activated by promoter exchange cloning and transformation to construct recombinant transformants *P. fluorescens* strains for large scale biological production of salicylate chemical scaffolds. Furthermore, the biological manipulation of these gene clusters by recombinant DNA approaches may produce additional novel and potent chemical derivatives of salicylic acid.

Competing Interests

The authors declared that no competing interests exist.

Acknowledgement

We acknowledge the Higher Education Commission (HEC) Pakistan for financial support under the National Research Program for Universities (NRPU) Project-5193, 2017.

References

1. Aiman S, Shehroz M, Munir M, Gul S, Shah M, Khan A. Species-Wide Genome Mining of *Pseudomonas putida* for Potential Secondary Metabolites and Drug-Like Natural Products Characterization. *J Proteomics Bioinform.* 2018;**11**(1):001-007. doi: 10.4172/jpb.1000460
2. Berdy J. Bioactive microbial metabolites, A personal review. *J. Antibiot.*2005; **58**, 1-26. DOI:10.1038/ja.2005.1
3. Huang T, Lin S. Microbial Natural Products: A Promising Source for Drug Discovery. *J Appl Microbiol Biochem*. 2017;**1**(2). doi:10.21767/2576-1412.100005
4. Diminic J, Starcevic A, Lisfi M, Baranasic D, Gacesa R, Hranueli D, Long PF, Cullum J, Zucko J. Evolutionary concepts in natural products discovery: what actinomycetes have taught us. J. Ind. *Microbiol. Biotechnol.* 2014;**41**(2),211-217. doi. org/10.1007/s10295-013-1337-8
5. Medema MH, Cimermancic P, Sali A, Takano E, Fischbach MA. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLOS Comput. Biol.* 2014;**10**(12),e1004016. doi.org/10.1371/journal.pcbi.1004016
6. Osbourn A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.*2010;**26**,449-457. doi: https://doi.org/10.1016/j.tig.2010.07.001
7. Haas D, Defago G. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol.* 2005;**3**,307–319. doi:10.1038/nrmicro1129
8. Palleroni NJ. Pseudomonadaceae. Bergey's Manual of Systematic Bacteriology. Krieg, N. R., and Holt, J. G., (editors) Baltimore: *The Williams and Wilkins Co*. 1984;141-199. doi: 10.1099/00207713-35-3-408.
9. Scales, BS, Dickson RP, Lipuma JJ, Huffnagle GB. Microbiology, genomics, and clinical significance of the *Pseudomonas fluorescens* species complex, an unappreciated colonizer of humans. *Clin Microbiol Rev*. 2014;**27**(4),927–948. doi: 10.1128/CMR.00044-14.
10. Loeschcke A, Thies S. *Pseudomonas putida* - a versatile host for the production of natural products. *Appl Microbiol Biotechnol*.2015;**99**,6197-6214. doi**:** https://doi.org/10.1007/s0025.
11. Li J, Fu R, Yang Y, Horz HP, Guan Y, Lu Y, Lou H, Tian L, Zheng S, Liu H, Shi M, Tang K., Wang S, Xu S. A metagenomic approach to dissect the genetic composition of enterotypes in Han Chinese and two Muslim groups. *Syst Appl Microbiol*. 2018;**41**(1),1-12. doi: https://doi.org/10.1016/j.syapm.2017.09.006
12. Li MH, Ung PM, Zajkowski J, Garneau S, Sherman DH. Automated genome mining for natural products. *BMC Bioinformatics* 2009;**10**,1-10. doi: https://doi.org/10.1186/1471-2105-10-185.
13. Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol*. 2015;**11**(9),639-648. doi:10.1038/nchembio.1884.
14. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, *et al*., Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* .2016;**44**(Database issue), D7-D19. doi: https://doi.org/10.1093/nar/gkv1290
15. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 2014;**2**,D560-D567. doi: https://doi.org/10.1093/nar/gkt963.
16. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA. Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 2017;**45**(W1):W36-W41. doi: https://doi.org/10.1093/nar/gkx319.
17. Van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res*. 2013;**41**(Web Server issue),W448-453. doi: https://doi.org/10.1093/nar/gkt391
18. Medema MH, Kottmann R, Yilmaz P, Cummings M, *et al*. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol*. 2015;**11**(9),625-631. doi: 10.1038/nchembio.1890.
19. Evans DA. History of the Harvard Chem Draw project. *Angew Chem Int Ed* 2014; **53**:11140-45. doi: https://doi.org/10.1002/anie.201405820.
20. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv Drug Deliv Rev*. 1997;**23**,3–25. http://doi.org/10.1016/S0169-409X(00)00129-0
21. Leeson P. Drug discovery: Chemical beauty contest. *Nature* 2012;**481**,455-456. doi:10.1038/481455a.
22. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*2006;1,**34** (Database issue), D668-72. doi: https://doi.org/10.1093/nar/gkj067.
23. Feher M, Williams CI. Numerical errors in minimization based binding energy calculations. *J Chem Inf Model.* 2012; **52,**3200-3212. doi: 10.1021/ci300298d.
24. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug* Des. 2011; **2**:146-157. **doi**: 10.2174/157340911795677602.
25. Gatei W, Das P, Dutta P, Sen A, Cama V, Lal AA, Xiao L. Multilocus sequence typing and genetic structure of *Cryptosporidium hominis* from children in Kolkata, India. *Infect. Genet. Evol.* 2007;**7**:197-205. doi: https://doi.org/10.1016/j.meegid.2006.08.006.
26. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, 2010;**11**:595.doi: https://doi.org/10.1186/1471-2105-11-595.
27. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 2016;**33**,1870-1874. doi: https://doi.org/10.1093/molbev/msw054.
28. Lucido MJ, Orlando BJ, Vecchio AJ, Malkowski MG. Crystal Structure of Aspirin-Acetylated Human Cyclooxygenase-2: Insight into the Formation of Products with Reversed Stereochemistry. *Biochemistry* 2016;**55**(8),1226-1238. doi**:** 10.1021/acs.biochem.5b01378

29. Zhang Y, Dufort I, Rheault P, Luu-The V. Characterization of a human 20alpha-hydroxysteroid dehydrogenase. *J. Mol. Endocrinol.* 2000;**25**(2),221-228. doi: 10.1677/jme.0.0250221.
30. Lewis MJ, Wiebe JP, Heathcote JG. Expression of progesterone metabolizing enzyme genes (AKR1C1, AKR1C2, AKR1C3, SRD5A1, SRD5A2) is altered in human breast carcinoma. *BMC Cancer* 2004;**4**(1),1-12. doi: https://doi.org/10.1186/1471-2407-4-27
31. Piekorz RP, Gingras S, Hoffmeyer A, Ihle JN, Weinstein Y. Regulation of progesterone levels during pregnancy and parturition by signal transducer and activator of transcription 5 and 20alpha-hydroxysteroid dehydrogenase. *Mol. Endocrinol*. 2005;**19**(2),431-440. doi: https://doi.org/10.1210/me.2004-0302.
32. Dhagat U, Carbone V, Chung RPT, Matsunaga T, Endo S, Hara A, El-Kabbani O. A salicylic acid-based analog discovered from virtual screening as a potent inhibitor of human 20alpha-hydroxysteroid dehydrogenase. *Med. Chem***.** 2007;**3**(6),546–50. **doi**: 10.2174/157340607782360399.
33. Indiragandhi, P, Anandham R, Madhaiyan M, Sa TM. Characterization of plant growth-promoting traits of bacteria isolated from larval guts of Diamondback moth *Plutellaxylostella* (Lepidoptera: Plutellidae). *Curr. Microbiol*. 2008;**56**(4),327-333. doi**:** https://doi.org/10.1007/s00284-007-9086-4.