Bioinformation
www.bioinformation.net

open access
Hypothesis

Metabolic pathway reconstruction of eugenol to vanillin bioconversion in *Aspergillus niger*

Suchita Srivastava[1], Suaib Luqman[1], Feroz Khan[2], Chandan S. Chanotiya[3], Mahendra P. Darokar[1*]

[1]Molecular Bioprospection Division, [2]Metabolic & Structural Biology Division, [3]Chemical Sciences Division, Central Institute of Medicinal & Aromatic Plants, (Council of Scientific & Industrial Research), Lucknow-226015 (UP) INDIA, Mahendra P. Darokar - E-mail: mpdarokar@yahoo.com * Corresponding Author

Received November 01, 2009; Revised November 18, 2009; Accepted December 13, 2009; Published January 23, 2010

Abstract:
Identification of missing genes or proteins participating in the metabolic pathways as enzymes are of great interest. One such class of pathway is involved in the eugenol to vanillin bioconversion. Our goal is to develop an integral approach for identifying the topology of a reference or known pathway in other organism. We successfully identify the missing enzymes and then reconstruct the vanillin biosynthetic pathway in *Aspergillus niger*. The procedure combines enzyme sequence similarity searched through BLAST homology search and orthologs detection through COG & KEGG databases. Conservation of protein domains and motifs was searched through CDD, PFAM & PROSITE databases. Predictions regarding how proteins act in pathway were validated experimentally and also compared with reported data. The bioconversion of vanillin was screened on UV-TLC plates and later confirmed through GC and GC-MS techniques. We applied a procedure for identifying missing enzymes on the basis of conserved functional motifs and later reconstruct the metabolic pathway in target organism. Using the vanillin biosynthetic pathway of *Pseudomonas fluorescens* as a case study, we indicate how this approach can be used to reconstruct the reference pathway in *A. niger* and later results were experimentally validated through chromatography and spectroscopy techniques.

Keywords: Metabolic Pathway, reconstruction, eugenol, vanillin, *Pseudomonas*, *Aspergillus*, GC-MS.

Background:
The identification of metabolic pathway's genes and proteins is an emerging area of great interest. The accumulation of data from different origins and the development of methods to mine that data create an opportunity to bridge the gap between the fragmentary view of genes and proteins and the more integrated approach of Bioinformatics **[1].** Moreover, increasing amounts of experimental data that can be mined for information about how proteins in cells assemble as metabolic and signal transduction pathways circuits are generated each day **[2-6].** Datasets available for such tasks include the primary literature, whole genome two hybrid screenings, large scale microarray experiments, full genome sequences and the patterns of conserved/non-conserved homologues and orthologues in them. Theoretical and experimental methods are being developed and used to analyze these different types of data and infer networks of proteins or genes that are involved in the similar type biological processes **[7].** In general, the networks derived by the computational analysis of these data are static, in the sense that they provide little information **[8].** This can be an important problem while assembling the network structure of either novel pathways or complex pathways with an unclear reaction network. Thus, it is a challenge to identify the correct bioconversion pathway that allows for the creation of sequence synteny based pathway network whose mechanism can be analyzed and tested against experimental observations. To achieve such a goal, strategies that combine the different theoretical and wet lab confirmations to identify proteins and generate a set of plausible pathway network for the process of interest are needed. Such a process integrates homologous data and provides testable predictions and information about unexplored pathway mechanism in other organisms. In the present work, we studied eugenol to vanillin bioconversion pathway in *Pseudomonas fluorescens* as a reference pathway and successfully reconstructed in *Aspergillus niger* on the basis of participating enzyme's protein sequences synteny map. However, vanillin (4- hydroxy-3- methoxybenzaldehyde) is one of the high value aided aromatic product of the flavor, food, cosmetic and pharmaceutical industry. Moreover, flavors and fragrances find wide applicability in food, cosmetic and pharmaceutical industries and represent a worldwide market of about 18 billion US dollars per year. Abundantly the flavors in the world market are chemically synthesized which is environmentally unfriendly and the production process too lacks the substrate specificity resulting in the formation of undesirable reaction mixtures, hazardous to health and increase the downstream cost. Less than 55 of the flavors are extracted from the plants and are natural products **[9].** However, the biotechnological means using naturally originated product as substrate for the production of flavors in high quantities are also considered as natural and is an emerging approach for the flavor production **[9].** On the other hand, production of natural vanillin from the orchid *Vanilla planifolia* is very low and unable to fulfill the demand of the world market of 12,000 tons per year. The microbial conversion of various phenyl propanoids such as ferulic acid, isoeugenol, eugenol, coniferyl aldehyde *etc*. to produce vanillin at a cost effective rate and free from health hazardous chemicals, has been an area of great interest **[10-13].** The bioconversion of eugenol to vanillin has been studied in various bacterial systems including different species of *Pseudomonas, Rhodococcus, Corynebacterium* and recombinant strains of *Amycolatopsis* so as to establish the pathways in the prokaryotic system. The drawback lies within the eukaryotic fungal system where no single fungus has been reported since now to be able to biotransform eugenol to vanillin **[14-17].** A combination of two fungal systems *A. niger* and *Pycnoporus cinnabarinus* were reported to convert ferulic acid to vanillin by catalysis in two different steps **[18, 19].** The establishment of the bioconversion pathway in fungal systems would lead to hypothesize the metabolic fate of eugenol in eukaryotic systems.

Aspergillus genome indicates the presence of 13,071 predicted genes with average gene length of 1,384 bp spread across 8 chromosomes. Notably, a number of hypothetical/unknown proteins are annotated in *A. niger* **[20].** This fact emphasizes the need to identify gene function by new approaches. In this work, we focus on eugenol to vanillin bioconversion pathway in *A. niger*. Currently, there are gaps in eugenol to vanillin bioconversion pathway of *A. niger* because of the insufficient annotation. We try to resolve such missing enzymes by using the protein local region sequence similarity in terms of conserved domain and motifs. Later these predicted domain and motifs were analyzed for their localization in orthologous sequences through KEGG and COG databases **[21].** Results showed conserved protein domains and motifs in both the organisms on the basis of pathway specific enzyme's functional domain and motif sequences & structural similarity. These predicted proteins motifs are evolutionarily conserved and form a basis for the identification of missing or unexplored enzymes, which ultimately helped in reconstruction of eugenol to vanillin bioconversion pathway. However, most of the works have been done in prokaryotes

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)

only and the details of this pathway are not fully understood in any eukaryotic system. In *P. fluorescens*, the pathway has been more extensively studied and so far following pathway enzymes are reported: (1a) Eugenol hydroxylase cytochrome C subunit (EhyA), (1b) Eugenol hydroxylase flavoprotein subunit (EhyB), (2) Coniferyl alcohol dehydrogenase (CalA), (3) Coniferyl aldehyde dehydrogenase (CalB), (4) Feruloyl CoA synthase (Fcs) and (5) Enoyl CoA hydratase/aldolase (Ech) **(Table 1 in supplementary material).** The reconstruction of biosynthetic pathway in *A. niger* indicates that functionally unexplored hypothetical proteins may have important role in the bioconversion of eugenol to vanillin. We also predicted the functional role of these hypothetical/unknown proteins as an enzyme for the development of high value aided aromatic product of the flavor, food, cosmetic and pharmaceutical industry.

Methodology:

Biological sequence retrieval:

Relevant information on the pathway specific genes and proteins of *P. fluorescens* involved in the bioconversion of eugenol to vanillin was studied through reported literature and later biological sequences were retrieved through nucleotide and protein sequence databases of both GenBank (NCBI) (www.ncbi.nlm.nih.gov/) & UniProtKB/Swiss-Prot (Expasy-EBI) (http://expasy.org/uniprot/) web servers.

Selection of datasets for pathway reconstruction

Two types of datasets were used for pathway construction: (1) **Known or Reference data set** - which includes protein sequences of five enzymes such as EhyA/EhyB, CalA, CalB, Fcs and Ech, participating in the vanillin biosynthetic pathway of model organism *i.e*., *P. fluorescens* with conserved protein domains & motifs, (2) **Predictive or Target data set** - which includes potential homologous protein sequences such as Cytochrome-c CYC_ASPNG or hypothetical protein (gi|2829474|sp|P56205.1 predicted as EhyA), hypothetical protein An09g01380 (gi|145241618|ref|XP_001393455.1 predicted as EhyB), hypothetical protein An15g01840 (gi|145250419|ref|XP_001396723.1 predicted as CalA), hypothetical protein An01g09260 (gi|145230075|ref|XP_001389346.1 predicted as CalB), hypothetical protein An14g05630 (gi|145249694|ref|XP_001401186.1 predicted as Fcs) and hypothetical protein An02g02820 (gi|145231906|ref|XP_001399422.1 predicted as Ech) predicted to be the pathway's enzymes of target organism *i.e*., *A. niger* with conserved protein domains & motifs **(Table 1 in supplementary material).**

Sequence similarity and homology search

We first identified the number and details of enzymes participating in eugenol to vanillin *bioconversion* through literatures. Then, each of the all-against-all proteins in the list were checked for sequence homology. Here homologous clusters are computationally defined clusters of similar proteins having ≥ 50% similarity and E-value threshold ≤1.00E-63 with similar conserved protein domain and motif local regions. Later this information was utilized to reconstruct the pathway reaction network in *A. niger*. Calculation of the statistical significance of matches was performed through Protein BLAST program (www.ncbi.nlm.nih.gov/BLAST/) at NCBI, USA webserver. Pairwise sequence alignment based studies indicates functional and evolutionary relationships between both the organisms as revealed by phylogenetic analysis and so helped in the identification of other family members which are not yet identified.

Detection of orthologs based on conserved protein domain, motif & superfamily

Homologous proteins in *A. niger* were searched for conserved protein regions such as protein motifs, domains and superfamilies, having related functions to that of the *P. fluorescens* **(Table 1 in supplementary material).** For conserved protein functional domain identification Conserved Domain search tool (CDD-Search at NCBI) was used (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), while InterPro-EBI (www.ebi.ac.uk/interpro/) and Pfam-Sanger (http://pfam.sanger.ac.uk/) databases were used for conserved protein motifs identification. Cluster of Orthologous Group (COG) database (www.ncbi.nih.gov/COG/) was used for identification of orthologous cluster group. We used *P. fluorescens* gene entries from the KEGG-GENES database (www.genome.jp/kegg/) **[22, 23]** and performed further annotation of motifs conservation to these genes in *A. niger*, by using the information about the gene ontology assignments through GenBank database at NCBI.

Experimental validation

Since protein sequence similarity or homology between reference and predicted pathway genes/proteins was lower therefore to confirm the theoretical predictions, localization of participating hidden and unexplored enzymes was further verified through chromatographic & mass spectroscopic techniques. The bioconversion of vanillin was screened on UV-TLC plates and later confirmed through GC and GC-MS techniques.

Discussion

The proteins of *P. fluorescens* that are known to be involved in eugenol to vanillin biosynthetic pathway are shown in **Table 1 (see supplementary material)**. Conserved common protein domain and motifs local regions in protein sequences of both the organisms are shown in **Table 1 (see supplementary material).** We first describe how the different large scale datasets are analyzed and combined, in order to inferred initial pathway network assembly in reference organism *i.e*., *P. fluorescens*. Then, we used local region sequence synteny mapping approach to analyze the pathway enzymes homolog in target organism *i.e*., *A. niger*. Finally, we hypothesized the most potential pathway enzymes based on the conserved protein motifs & experimental observations. Bibliographic analysis of published abstracts and papers were used to verify the different proteins/genes involved in *bioconversion* process **(Figure 1 & 2).** We have used the genes known to be involved in eugenol to vanillin biosynthetic pathway in *P. fluorescens* to test this assumption in *A. niger* as an alternative organism for production of vanillin at industrial level.

Pathway reconstruction using Phylogenetic profiling

It is not possible to infer probable reaction network structure for the pathway just from the co-occurrence of genes. Additional data is necessary in order to translate the group of proteins that are identified to be involved in the biological process of interest into a structured network. Phylogenetic profiling of the genes assists in adding some structure to the network **[24, 25].** We searched for motif patterns of co-occurrence of sets of orthologs in other species of *A. niger*. The assumption behind the use of this approach is that, if the conserved pattern or motif present in two or more proteins of different species of *Aspergillus* is very close, this is an indication of co-evolution among the proteins. Such co-evolution can be taken as an indication that they are likely to be involved in the same cellular process **[26, 27].** Results from phylogenetic co-evolution analysis of the different genes involved in eugenol bioconversion are explained in Methodology section. This additional information assists in inferring some form of common motifs to the network structure of the studied pathway **(Table 1 in supplementary material)**. For example, in *P. fluorescens*, conserved motifs *viz*. PS51007, IPR003088 & PF00034 of EhyA enzyme showed 100% homologues coincide with CYC_ASPNG gene of *A. niger* with 58% amino acid similarity and 38% identity. Similarly enzyme of *P. fluorescens* EhyB showed common evolutionary conserved motif *viz*., Pfam01565 and Pfam02913 in *A. niger* with 50% amino acid sequence similarity and 34% identity. This suggests that predicted hypothetical proteins are more likely to act directly on these reactions as participating unexplored enzymes, which are not yet functionally annotated.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)

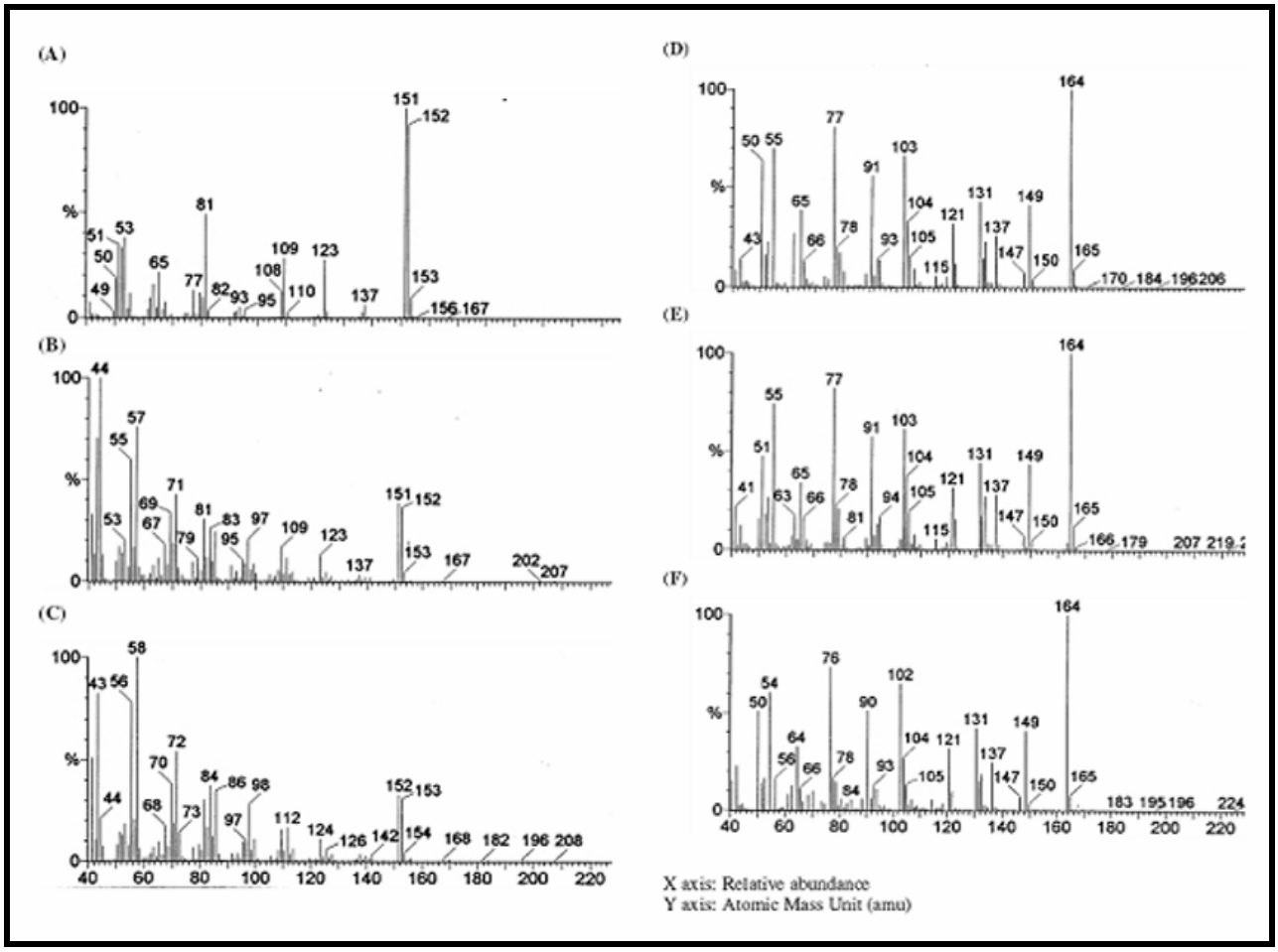

Figure 1: Mass spectra (MS) of standard vanillin (A), *Aspergillus* mediated biotransformed extracellular vanillin (B), intracellular vanillin (C), standard eugenol (D), eugenol extracellular (E) and eugenol intracellular (F).

Experimental validation through TLC and GC & GC-MS

Of five *P. fluorescens* genes, we annotated all five as putative enzymes of the eugenol to vanillin bioconversion pathway in *A. niger* with the information on common conserved protein motifs & domain. We predicted enzymatic reactions for all the participating enzymes on the basis of functionally conserved protein motif and domain homology and co-evolution. Later, we mapped them on target organism based on KEGG metabolic pathway database **(Table 1 in supplementary material).** We hypothesized cases where the reaction steps of missing enzymes can be catalyzed by other enzymes. Besides, to validate the above pathway, species of *Aspergillus* were screened experimentally in our lab for bioconversion capability from eugenol to vanillin. Two of the *Aspergillus species* namely, *A. niger* and *A. flavus* were found capable for bioconversion of eugenol to vanillin. The bioconversion of vanillin was screened on UV- TLC plates and later on confirmed through GC and GC-MS techniques **(Figure 1).** Experimental observations of UV- TLC and GC & GC-MS showed the likely roles of proteins that have been analyzed and also predicted through homology search and conserved domain and motif search tools. This eliminates some of the alternative network structures and assists in refining the remaining, by suggesting experiments that can further differentiate between them. As shown in this paper, bioinformatics tools like Interpro, ProDom, Pfam, COG, KEGG, BLAST etc. were successfully used along with experimental data through chromatographic and spectroscopic techniques so that to assist in the predictive analysis of the missing metabolic pathways. Another useful aspect of our approach is that it can be used in reassessing the interpretation of experimental chromatographic data. It is our goal to refine predictions based on updated sequence data and to apply this method to other organisms, thus assisting in the understanding of how the pathway of vanillin bioconversion has evolved. In principle, the combination of integrated approach presented here could be used in a flexible way to analyze similar problems in other biological system **(Figure 2).**

In our present work, we used a combination of experimental and theoretical methods to reconstruct the topology of the eugenol to vanillin bioconversion pathway in *A. niger* with the help of *P. fluorescens* reported data. The pathway enzymes were identified and a network of interactions between them was predicted using literature mining and computational analysis. Although bioinformatics tools do provide a synteny of the pathway network structure, human analysis remains necessary for curating all the relevant information. In fact, human curation of the pathway chain is a critical step for the derivation of possible alternative reaction schemes for the pathway. At this point, theoretical predictions are needed to validate experimentally in wet lab.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)

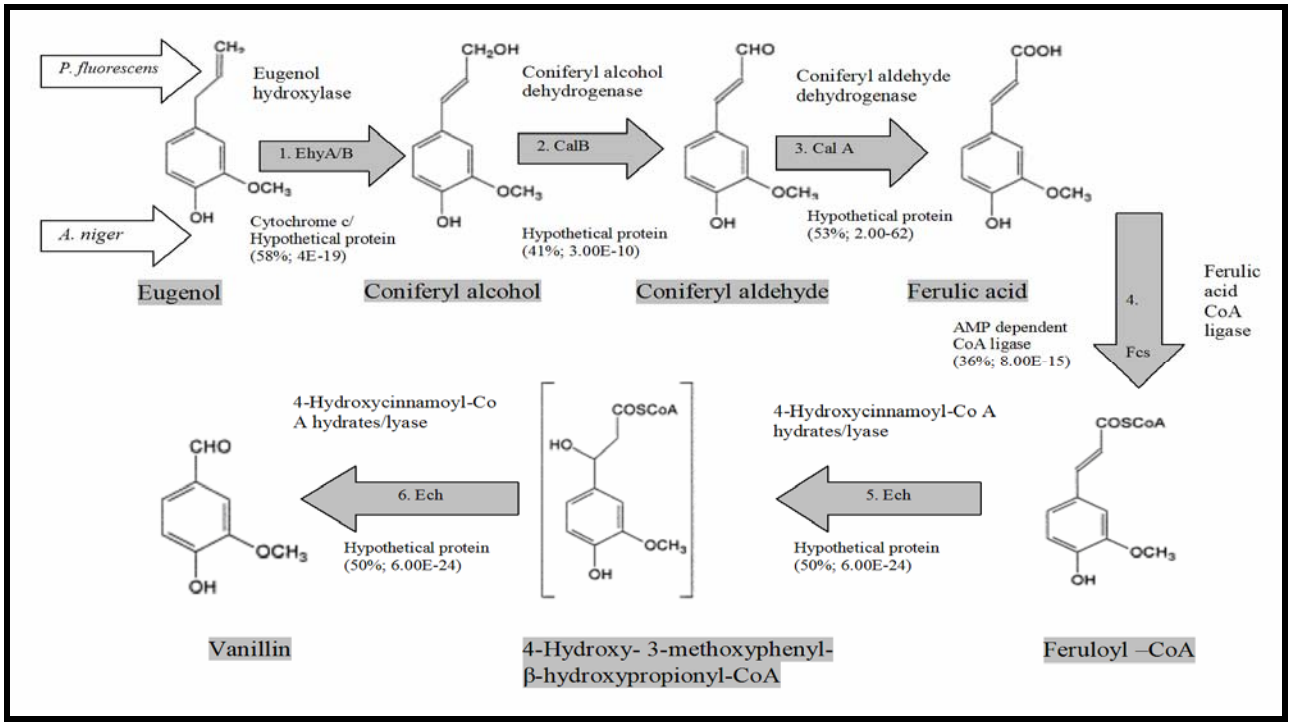

Figure 2: Comparison of reference vanillin biosynthetic pathway of *P. fluorescens* with reconstructed pathway of *A. niger*. Values in parentheses showed protein sequence similarity and BLAST E-value.

Conclusion:

The eugenol to vanillin bioconversion pathway of *P. fluorescens* was reconstructed by using a flexible computational methodology that combines sequence analysis, literature search and bioinformatics applications. The roles of different *P. fluorescens* proteins in vanillin bioconversion were successfully predicted in *A. niger*. Some predictions are cross checked through published experimental data. Other predictions need further experimental work to validate it. The methodology used here is flexible and can be applicable in other systems. This methodology could be a step forward in integrating different types of data to obtain systemic knowledge about novel pathways and to clarify how current prediction of known pathways would work, thus generating rational hypothesis for testing. Besides, to validate the above pathway, different species of *Aspergillus* were screened experimentally in our lab for bioconversion capability from eugenol to vanillin. Two of the *Aspergillus species* namely, *A. niger* and *A. flavus* were found capable of vanillin bioconversion. Moreover, our hypothesis of vanillin biosynthetic pathway was later successfully confirmed through UV - TLC and GC & GC-MS data, which showed sign of vanillin synthesis in *A. niger*. Thus, vanillin production along with the formation of intermediates like ferulic acid strongly supports the significance of bioinformatics approaches in the reconstruction of unexplored metabolic pathway. We found mainly hypothetical proteins of *A. niger* which are in the same gene groups as real enzymes on the basis of conserved protein domain and motifs. There are so many missing enzymes in the biosynthetic pathways of eukaryotes, using such strategies we can find more hypothetical proteins that may replace missing enzymes information of important pathways. Validation of the results by using additional information, such as microarray and proteomic data analysis could be our future works.

Acknowledgement:

We acknowledge Council of Scientific & Industrial Research, New Delhi Network Project'09 (CSIR-NWP09) for financial support at Central Institute of Medicinal & Aromatic Plants, Lucknow, India.

References:

[1] C Francke ., *et al*., *Trends Microbiol* (2005), **13:**550-558. [PMID: 16169729]
[2] P D Karp ., *Genome Biol* (2004), **5**:401. [PMID: 15287973]
[3] P D Karp., *et al*., *Bioinformatics* (2002), **18**:S225-232.
[4] P D Karp., *et al*., *Pac Symp Biocomput* (2004), **9**:190-201.
[5] T Ideker., *Adv Exp Med Biol* (2004), **547**:21-30. [PMID: 15230090]
[6] T Ideker., *et al*., *Ann Biomed Eng* (2006), **34**:257-264. [PMID: 16474915]
[7] R Alves., *et al*., *Nat Biotechnol* (2006), **24**:667-672. [PMID: 16763599]
[8] R Alves., Sorribas A., *BMC Systems Biol* (2007), **1**:10.
[9] E J Vandamme., Soetaert W., *J Chem Tech Biotechnol* (2002), **77**:1323-1332.
[10] J Rabenhorst., *Appl Microbiol Biotechnol* (1996), **46**:470-474
[11] H Furukawa., *et al*., *J Biosci Bioeng* (2003), **96**:401-403. [PMID: 16233545]
[12] M Yamada., *et al*., *Appl Microbiol Biotechnol* (2007), **73**:1025-1030. [PMID: 16944125]
[13] M Yamada., *et al*., *Biotechnol Lett* (2008), **30**:665-670. [PMID: 18040605]
[14] R Plaggenborg., *et al*., *Appl Microbiol Biotechnol* (2006), **72**:745-755. [PMID: 16421716]
[15] D Hua., *et al*., *J Biotechnol* (2007), **130**:463-470 [PMID: 17583367]
[16] R C Kasana., *et al*., *Curr Microbiol* (2007), **54**:457-461. [PMID: 17487530]
[17] S Tsujiyama., Ueno M, *Bioscience Biotechnol Biochem* (2008), **72**:212-215.
[18] B Falcnnier., *et al*., *J Biotechnol* (1994), **37**:123-132.
[19] Rao Sr, G A Ravishankar., *J Sci Food Agric* (2000), **80**:289-304.
[20] H Priefert., *et al*., *Arch Microbiol* (1999), **172**:354-363.
[21] F Nikitin., *et al*., *Genome Inform* (2004), **15**:266-75. [PMID: 15706512]

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)

[22] H Ogata., *et al*., *Biosystems* (1998), **47**:119-128. [PMID: 9715755]
[23] M Nakao., *et al*., *Genome Inform Ser Workshop Genome Inform* (1999), **10**:94-103. [PMID: 11072346]
[24] M Pellegrini., *et al*., *Proc Natl Acad Sci USA* (1999), **96**:4285-4288. [PMID: 10200254]
[25] H Li., *et al*., *Nat Biotechnol* (2005), **23**:253-260. [PMID: 15696156]
[26] I V Singh *et al*., *Online J Bioinformatics* (2004), **5**: 32-48.
[27] F Khan. *et al*., *J Integrative Bioinformatics* (2008), **5**(1): 86.

Edited by P. Kangueane
Srivastava *et al.,* Bioinformation, 4 (7) 320-325, (2010)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)

Supplementary material

Table 1: Details of predicted conserved protein domains & motifs found in protein sequences of putative enzymes involved in vanillin bioconversion in both *P. fluorescens* and *A. niger*.

S. No.	Reference Enzymes of Vanillin Biosynthetic Pathway	Reference Pathway protein in *P. fluorescens* with protein ID & Length	Predicted Pathway protein in *A. niger* with protein ID, Length & BLAST score	Conserved protein domain & motif	Conserved domain/motif Function
1.	Eugenol hydroxylase A (EhyA)	Eugenol hydroxylase cytochrome C EMB CP000076 111 aa	CYC-A (Cytochrome c) P56205 A61490 111 aa (87.4 score, 58% similarity, 38% identity, 4E-19 E-value)	PS51007 IPR003088 PF00034	Cytochrome-C motif
	Eugenol hydroxylase B (Ehy B)	Eugenol hydroxylase B CAB64355 AJ243941 517 aa	Hypothetical protein (An09g01380) XP_001393455 XM_001393418 552 aa (283 score, 50% similarity, 34% identity, 4.00E-77 E-value)	PFAM01565 PFAM 02913 COG0277	FAD_binding_4, FAD binding domain FAD-oxidase C, FAD linked oxidases, C-terminal domain GlcD, FAD/FMN-containing dehydrogenases
2.	Coniferyl alcohol dehydrogenase (Cal B)	Coniferyl alcohol dehydrogenase BAD11007 AB162132 255 aa	Hypothetical protein (An15g01840) XP_001396723 XM_001396686 280 aa (63.2 score 41% similarity, 22% identity, 3.00E-10 E-value)	CL09931 IPR002198 IPR002347 IPR016040	AdoHcyase IPR002198,short chain dehydrogenase/ reductase Glucose/ribitol dehydrogenase NAD(P)-binding
3.	Coniferyl aldehyde dehydrogenase (CalA)	Coniferyl aldehyde dehydrogenase YP_262923 NC_004129 476 aa	Hypothetical protein (An01g09260) XP_0013896346 XM_001389309 500 aa (238 score 53% similarity, 36% identity, 2.00-62 E-value)	CL00545 PFAM00171 IPR015590 IPR016162 IPR012394 IPR016161	LuxC, Acyl-CoA reductase (LuxC) Aldedh, Aldehyde dehydrogenase family Aldehyde dehydrogenase, N-terminal Aldehyde dehydrogenase NAD(P)-dependent Aldehyde/histidinol dehydrogenase
4.	Feruloyl CoA synthase (Fcs)	Feruloyl CoA synthase AAZ23792 DQ119298 589 aa	AMP dependent CoA ligase XP_001401186 XM_001401149 513 aa (80.1 score 36% similarity, 24% identity, 8.00E-15 E-value)	CL100401	AMP-binding, AMP-binding enzyme
5.	Enoyl CoA hydratase/ aldolase (Ech)	Enoyl CoA hydratase AAZ23790 119298 276 aa	Hypothetical protein XP_001399422 XM_001399385 294 aa (108 score 50% similarity, 34% identity, 6.00E-24 E-value)	CL109483	Enoyl Co A hydratase/ isomerase family

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 320-325 (2010)