Comparative Phylogenetics Approach for Discovering Alternative Source of Taxol

S Kushwaha, P Chauhan, M Shakya

Citation

S Kushwaha, P Chauhan, M Shakya. *Comparative Phylogenetics Approach for Discovering Alternative Source of Taxol*. The Internet Journal of Bioengineering. 2008 Volume 3 Number 2.

Abstract

Bioprospecting is one of the prominent areas of research of commercial and valuable compounds, as it provides alternative sources of these compounds. One of the most persuasive methods for bioprospecting is through molecular phylogenetics analysis. Enzymes involved in the biosynthetic pathway of these compounds are considered as a base for bioprospecting. In this paper, an attempt has been made to find alternative sources for valuable compound taxol by comparative phylogenetics analysis using enzymes physiochemical data and sequence data. Dendogram was generated through physiochemical data whereas phylogenetic tree was generated through sequence data. Consequently after comparison four different organisms, Fungal BT2, OzoniumT2, Abies grandis and Ginkgo
biloba, have been observed to be related to taxus plants both sequencially and physiochemically.

INTRODUCTION

Exploration of useful application, method, or product in nature is called bioprospecting (Biodiversity Prospecting). Bioprospecting is one of the prominent areas of research of commercial important compounds, which is not only providing alternative source for these compounds but also improving the quality of products and their cost effective extraction. One of the most persuasive methods for bioprospecting is through molecular phylogenetics analysis [$_{2}$]. Phylogenetic analysis provides new horizons for this type of study and shows their worthwhile by giving answers of their origin, development and other characteristics that forms the base for bioprospecting. Here, an attempt has been made to find alternative sources for valuable compound taxol. Taxol (paclitaxel) is one of the natural diterpenoid alkaloids firstly isolated from the bark of the yew (Taxus brevifolia) [$_{10}$]. Because it can kill tumor cells by enhancing the assembly of microtubules and inhibiting their depolymerisation, taxol has been well established and approved by FDA (the Food and Drug Administration) as a very important effective chemotherapeutic agent against a wide range of tumors since 1992 [$_{11}$]. It is most commonly used against the ovarian, breast and lung cancers, but currently also used in certain aspects of heart disease treatment and is showing promise in alzheimer's therapy. Until now, Taxol used in cancer chemotherapy and scientific research is isolated from yew tree or semisynthesized from its precursors such as baccatin III and 10-deacetylbaccatin III which are all isolated from this natural plant [$_{13}$]. Taxol production using microorganisms such as fungi (28) and bacteria (29) have also been proposed [$_{10}$]. However, this natural resource is being threatened day by day due to the destructive collection of Taxus bark for Taxol. In order to protect Taxus in the world and lighten the pressure of Taxol sourcing, various approaches to obtain Taxol are essentially required.

In the past, morphological data was used for inferring phylogenies. However, the abundance of DNA/Protein sequence data currently available from a variety of organisms has led to phylogenetics analysis [$_{8}$]. For bioprospecting, Enzymes involved in the biosynthetic pathway of the compounds act as mapping units. Questions concerning enzyme function and performance remain unanswered with molecular data. However these questions can be answered by analysing physiochemical data of these enzymes. Physiochemical properties and their similarity level can, best explore qualitative measurement of these enzymes. Studies in this direction focusing on individual pathways [$_{7}$] or on the entire metabolic repertoire [$_{6}$] have been attempted. In the present work, an attempt has been made to find alternate source for taxol by analyzing taxol biosynthesis pathway.

TAXOL BIOSYNTHESIS IN TAXUS

Except for a few undefined steps, the complete taxol biosynthetic pathway was elucidated and many genes encoding certain enzymes, which regulate taxol biosynthesis pathway, have been cloned and characterized (Table-1). Enzymes involved in the biosynthesis of taxol belong to diterpenoid biosynthesis pathways[1].

Figure 1

Table 1: List of 13reported enzymes of taxol biosynthetic pathways

Enzyme	c-DNA corresponding to the enzymes			Reference
	Gene bank Ac. No.	CDS(bp)	Enzyme(kda)	
Taxadiene synthase	AY364469	2,586	98.3	Wildung al.,1996
GGPS	AF081514	1,182	42.6	Hefner et al.,1998
TAT	AF190130	1,317	49	Walker et al.,2000a
TBT	AF297618	1,320	50	Walker et al.,2000b
DBAT	AF193765	1,320	49	Walker et al.,2000c
Taxane 10-beta hydroxylase	AF318211	1,494	56.7	Schoendorf et al.,2001
Taxane 13-alpha hydroxylase	AY056019	1,458	54.7	Jennewein et al.,2001
BAPT	AY082804	1,335	50	Walker et al.,2002a
DBTNBT	AF466397	1,323	49	Walker et al., 2000b
Taxane 2-alpha hydroxylase	AY518383	1,488	55	Chau et al., 2004a
Taxane 7-beta hydroxylase	AY307951	1,503	56.3	Chau et al., 2004b
Taxane 5-alpha hydroxylase	AY289209	1,509	56.8	Jennewein et al., 2004
PAM	AY582743	2,094	76.5	Walker et al., 2004

In the present work, comparative phylogenetic analysis using protein sequence and physiochemical data of all the l3 enzymes involved in the biosynthetic pathway of taxol is performed to determine the alternative source for taxol. Enzymes of the Source that gave same results with physiochemical data and with sequence data were proposed to be the most prominent alternate source for taxol biosynthesis.

MATERIALS AND METHODS

PERFORMED THE BLAST SEARCH FOR TAXOL BIOSYNTHETIC METABOLIC ENZYMES

Nucleotide and protein sequences were retrieved from the Genebank accession Numbers, which were then subjected to a BLASTp [9] search against the non-redundant database with the e-value inclusion threshold, set to 0.005. The search was restricted to plants through an option available in the BLAST program, which allows the user to select a particular organism [3]. In this way potential homolog of enzymes were selected.

PERFORMED DATA PREPARATION AND CHARACTERIZATION FOR PHYLOGENETICS ANALYSIS

For phylogenetic analysis, enzymes sequence data and physiochemical data was collected, processed and analyzed.

PHYSIOCHEMICAL DATA

Enzymes physiochemical property such as No. of AAs, No. of Atoms, Molecular weight, Iso-electric point, Positive charge residues(Asp + Glu), Negatively charge residues(Arg + Lys), Aliphatic index, Instability index[12], Hydropathocity [14], and Extinction coefficient were obtained from the protein sequence of each enzyme when subjected to ExPASy tool, ProtParam.

SEQUENCE BASED DATA

In order to prepare data for phylogenetic tree analysis, entire protein sequence of each enzyme was subjected to site based alignment. Constant sites(C), Variable sites (V) and Singleton sites(S) were analyzed with the help of MEGA4.0 software [4]. Then overall similarity of the alignments was calculated by the Feature similarity score (FSS).

Figure 2

$$Feature\ similarity\ score\,(FSS) = \frac{Constant\ site + Singleton\ site}{Total\ sites}$$

PHYLOGENETICS TREE AND DENDOGRAM GENERATION

DENDOGRAM GENERATION FROM PHYSIOCHEMICAL DATA.

Steps used for dendogram generation through physiochemical parameters are:

Calculation and collection of physiochemical properties for each enzyme and their homologous by ExPASy tool, ProtParam.

Choose distance measurement and linkage method for the physiochemical data, for clustering.

Generate dendogram through statistical software Minitab.

TREE GENERATION FROM SEQUENCE BASED DATA.

For phylogenetic tree construction, one of the most important tasks is method selection, which depends upon the nature of data [4]. Steps adopted for method selection [15] are:

1. Build alignment and check family of sequences.
2. If data has high similarity (high FSS) i.e. greater than 75%, use character based method (MP).
3. If data has moderate similarity (medium FSS) i.e. less than 75% and more than 50%, use distance based method (NJ, UPGMA).

4. If data has lower similarity (low FSS) i.e. less than 50%, use ML method

* After observation of data, Cut off (%) is planned.

RESULTS AND DISCUSSION

BLAST Similarity search of enzymes involved in the pathway provides information about other sources that contains the enzymes and thus can be an alternative source for the taxol biosynthesis.

Results for physiochemical data: Physiochemical data of GGPS and its homologous which is calculated by ExPASy tool, ProtParam is shown in table-2. GGPS is one of the 13 known enzymes involved in taxol biosynthetic pathway.

Figure 3

Table 2: Physiochemical data of GGPS enzyme and their homologs.

Name of Potein enzymes	N	M	MW	PI	NG	PS	I I	EC	AI	H
geranyl\geranyl diphosphate synthase [Taxus canadensis]	393	5990	42571.9	5.58	52	43	41.44	19535	90.18	-0.055
geranyl\geranyl diphosphate synthase [Abies grandis]	383	5922	42069.3	6.61	47	45	43.09	24910	91.72	-0.079
geranyl\geranyl diphosphate synthase [Ginkgo biloba]	391	5976	42533.8	5.98	51	45	46.03	17015	87.62	-0.096
geranyl\geranyl diphosphate synthase 5 [Picea abies]	382	5889	41738.1	7.60	44	45	39.49	24910	93.74	-0.022
geranyl diphosphate synthase [Abies grandis]	383	5855	41592.7	6.08	49	42	46.38	23420	96.11	-0.057
geranyl diphosphate synthase [Abies grandis]	381	5799	41324.5	6.83	44	43	42.64	23880	86.12	-0.136
geranyl diphosphate synthase 2 [Picea sbies]	386	5889	42154.5	5.53	52	41	49.33	22390	100.08	-0.052
geranyl diphosphate synthase [Abies grandis]	387	5979	42463	6.8	47	46	47.11	17920	90.54	-0.035
predicted protein [Physcomitrella patens subsp.	288	4378	30938.5	5.57	40	33	35.04	12170	97.95	-0.053

(Abbreviations: N=No. AAs, M= No. of Atoms, MW= Molecular Weight , PI= Isoelectric Point, NG=(Asp + Glu) PS=(Arg + Lys), II= Instability Index, EC=Extinction Coefficient, AI= Aliphatic Index, H= Hydropathicity)

Similarly, physiochemical data was collected for other twelve enzymes and dendrogram was generated from these data for each enzyme by using statistical software Minitab.

Results for Sequence based data: Site based alignment was done to find similarity between all the 13 enzymes involved in the pathway.

Figure 4

Table 3: Characterization of all 13 known enzymes of taxol biosynthetic pathway through MEGA 4.0

S.N.	Name of Enzymes	Conserve /Constant sites	Variable sites	Singleton -sites	Total sites	FSS index (%)
1.	Taxadiene synthase	338	542	139	883	54.02
2.	GGPS	164	229	122	394	72.58
3.	TAT	188	261	45	451	51.66
4.	TBT	167	283	62	451	50.77
5.	DBAT	270	173	14	443	64.10
6.	Taxane 10-beta hydroxylase	1	498	4	500	1.00
7.	Taxane 13-alpha hydroxylase	248	249	73	504	63.69
8.	BAPT	190	260	50	450	53.33
9.	DBTNBT	155	292	73	449	50.77
10.	Taxane 2-alpha hydroxylase	159	346	35	514	37.74
11.	Taxane 7-beta hydroxylase	159	346	38	514	38.32
12.	Taxane 5-alpha hydroxylase	159	346	38	514	38.32
13.	PAM	279	460	90	753	49.00

Out of 13 enzymes, enzymes with higher FSS are zero, moderate FSS are five and low FSS are seven, subsequently distance based and maximum liklihood tree are constructed by MEGA and phylip software (refer 2.3.2). The phylogenetic analysis reveals that the trees are unrooted. The main observations are as follows:

Fungal sp. BT2 contains five enzymes (Enzymes-3, 4, 5, 8 and 9) of taxol biosynthesis.

Ozonium sp. BT2 comprises of five enzymes(Enzymes-6,7,10,11 and 12) involved in taxol biosynthesis

Cladosporium cladosporioides(4, 5), Abies grandis(1, 2) and Ginkgo biloba(2,13) possess two-two enzymes each of the taxol biosynthetic pathway.

GGPS is found in Abies grandis and Ginkgo biloba. Abies grandis and ginko biloba which are distantly related to taxus plants.

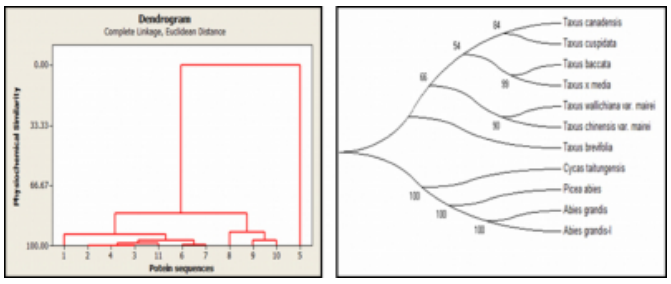
Transferases are found in cuspidata and Taxus x media species of taxus.

Hydroxylases are found in cuspidata and chinensis species of taxus.

Three type of groups were identified from comparative result of phylogenetics tree and physiochemical dendogram, less supporting (Enzymes-2 and 3),moderately supporting(Enzymes-10,11,and 12) and well supporting (Enzymes-1,4,5,6,7,8,9 and 13) shown in table-4. One phylogenetic tree and dendogram belonging to each group is shown in (figure-1, 2 and 3).

Figure 5

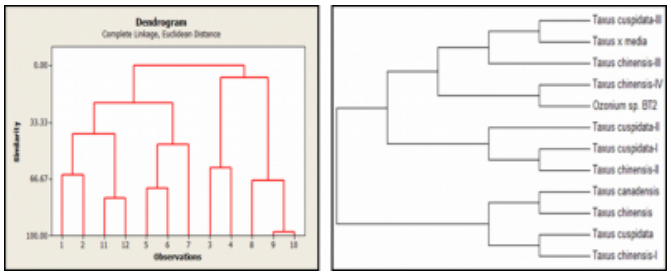
Figure 1: shows phylogenetic trees of Enzymes-1 from molecular data and physiochemical data



(1-Taxus canadensis, 2-Taxus cuspidata,3-Taxus baccata, 4-Taxus x media,5- Taxus brevifolia,6-Taxus wallichiana var. mairei,7- Taxus chinensis var. mairei,8-Abies grandis,9-Picea abies,10-Abies grandis-I,11-Cycas taitungensis)

Figure 6

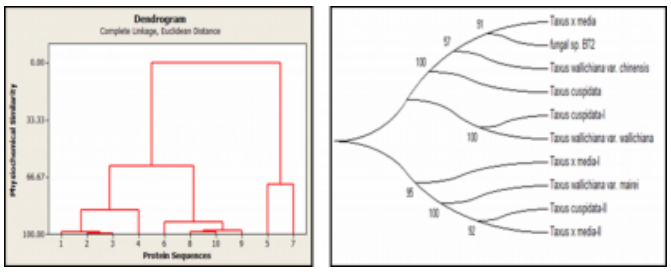
Figure 2: shows phylogenetic trees of Enzymes-10 from molecular data and physiochemical data



(1-Taxus canadensis, 2- Taxus chinensis, 3- Taxus cuspidata, 4- Taxus chinensis-I, 5-Taxus cuspidata-I 6- Taxus chinensis-II, 7-Taxus cuspidata-II, 8- Taxus chinensis-III, 9-Taxus cuspidata-III, 10- Taxus x-media, 11-Taxus chinensisIV, 12- Ozonium sp. BT2)

Figure 7

Figure 3: shows phylogenetic trees of Enzymes-3 from molecular data and physiochemical data



(1-Taxus cuspidata, 2- Taxus wallichiana var. chinensis,3-Taxus x media, 4- fungal sp. BT2,5- Taxus cuspidata-I, 6-Taxus x media-I, 7- Taxus wallichiana var. wallichiana,8-Taxus cuspidata-II,9- Taxus wallichiana var. mairei, 10-Taxus x media-II)

Figure 8

Table 4: Comparative result of sequence and physiochemical data based on phylogeny as well as quality of taxol enzymes with other plants.

Taxol Enzymes	Comparative results of phylogenetic trees
Taxadiene synthase	Sequence data phylogeney is well supported by the physiochemical data dendogram Physiochemical data level is enzyme-1 is matched (97.16 %)with *Abies grandis*, and *Picea abies*
GGPS	Sequence data phylogeney is less supported by the physiochemical data dendogram Physiochemical data level i enzyme-2 is matched (94.64 %)with *Abies grandis-III*, and *Ginkgo biloba*
TAT	Sequence data phylogeney is less supported by the physiochemical data dendogram Physiochemical data level is enzyme-3 is matched (85.99 %)with *fungal sp. BT2*
TBT	Sequence data phylogeney is well supported by the physiochemical data dendogram Physiochemical data level is enzyme-4 is matched (85.99 %)with *fungal sp. BT2*

CONCLUSION

In the present work, a comparative phylogenetic analysis of 13 enzymes involved in the biosynthetic pathway of taxol was conducted, consecutively to determine its alternate source. Feature similarity score (FSS) calculated for molecular data, whereas distance measurement and linkage method was studied for dendogram generation. Three type of groups were identified from comparative results of phylogenetics tree and physiochemical dendogram, among them two are less supporting (Enzymes-2 and 3), three are moderately supporting (Enzymes-10, 11and 12) and eight are well supported (Enzymes-1, 4, 5, 6,7,8,9 and 13). Four different organisms, Fungal BT2, Ozonium T2, Abies grandis and Ginkgo biloba, have been observed to be related to taxus plants both sequentially and physiochemicaly. Data provides highly authentic information about the alternative source, as it considers not only sequence similarity but also physiochemical relatedness

FUTURE DIRECTIONS

The present model for alternative sources can be further extended with some modifications, if necessary for analysis of other varieties of valuable compound. The work used here is in great demand, as is not only providing alternative sources for these valuable compounds but also provides useful information about the quality of products. This work will restrict the search area of the scientists working for bioprospecting. This knowledge will contribute positively to bioprospecting for new sources of valuable compounds.

ACKNOWLEDGEMENTS

We are grateful to Department of Bioinformatics, MANIT, Bhopal, India for support and cooperation.

References

1. Guo B. H., Kai G. Y., Jin H. B. and Tang K. X., (2006). Taxol synthesis, African Journal of Biotechnology. 5:015-020.
2. William L. S. and Ward C. W., (2006). Venom Evolution Widespread in Fishes: A Phylogenetic Road Map for the Bioprospecting of Piscine Venoms, Journal of Heredity. 97:206-217.

3. Anishetty S., Pulimi M. and Pennathur G.,(2005). Potential drug targets in Mycobacterium tuberculosis through metabolic pathway analysis, Comput. Bio.Chem. 29: 368-378.
4. Hall B. G.,2005. Comparison of the Accuracies of Several Phylogenetic Methods Using Protein and DNA Sequences, Mol. Biol. Evol. 22:792-802
5. Kumar S., Tamura K. And Nei M.,(2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. 5:150-163.
6. Liao L., Kim S., and Tomb J. F.,(2002). Genome comparisons based on profiles of metabolic pathways. In Sixth International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2002), Crema, Italy.
7. Forst C. and Schulten K., (2001).Phylogenetic analysis of metabolic pathways. Journal of Molecular Evolution, 52:471-489.
8. J. Felsenstein,(1998).Phylogenies from molecular sequences: Inferences and reliability. Annual Rev. Genet., 22:521-565.
9. Altschul S. F., Thomas L.M., Alejandro A.S., Jinghui Z., Zheng Z.,Webb M. and David J.L.,(1997). Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.
10. Seki M. and Furusaki S.,(1996). An immobilized cell system for Taxol production. CHEMTECH. 26:41-45.
11. Kohler, J., Goldspiel, B.R. (1994). Evaluation of new drug paclitaxel (taxol). Pharmacotherpy,14:3-34.
12. Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 4,155-161.
13. Denis, I.N., (1988). Highly efficient Practical approach to natural Taxol. J.Am. chen. soc. 110: 5917-5919
14. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-132.
15. David Mount, Bioinformatics: sequence and genome analysis, second ed.,2004

Author Information

Sandeep K. Kushwaha
Department of Bioinformatics, MANIT

Pallavi Chauhan
Department of Bioinformatics, MANIT

Madhvi Shakya
Department of Bioinformatics, MANIT