A Module tree of Wnt Signal Transduction Pathways

Losiana Nayak
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
losiana_t@isical.ac.in

Nitai P. Bhattacharyya
Biomedical Genomics Centre
5 Suburban Hospital Road
Kolkata, India
nitai_sinp@yahoo.com

Rajat K. De
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
rajat@isical.ac.in

***Abstract*—Wnt signal transduction pathway (Wnt STP) is a crucial intracellular pathway mainly due to its participation in important biological functions, *i.e.*, embryonic development, and stem-cell management among others as well as in human pathology, mainly cancer. For these very reasons, Wnt STP is one of the highest researched signal transduction pathways. Study and analysis of its origin, expansion and gradual development to the present state as seen in human beings is one facet of this multi-pronged research. Development of the Wnt signal transduction pathway in cellular environment among various species is not clear till date. A phylogenetic tree obtained from Wnt STPs of multiple species is one of the ways to handle this problem. In this respect, we propose a new idea of constructing a phylogenetic tree from modules of Wnt STPs of diverse species. We term it as the 'Module Tree'. A module is nothing but a self-sufficient minimally-dependent subset of the original Wnt STP. Authenticity of the module tree is tested by comparing it with the two reference trees. It performs better than an alternative phylogenetic tree constructed from pathway topology of Wnt STPs.**

***Index Terms*—Module Tree; Modularization algorithm; Wnt Signal Transduction Pathway; Pathway development**

I. INTRODUCTION

A module is a subset of a biochemical network, which tends to be self-sufficient and have minimal dependency on the rest part of the network. The justification for dividing a network into a number of modules lies in the fact that the complexity of each module is much less than that of the entire pathway and is an easier means of studying the entire network by parts. Thus analyzing all the modules generated from a pathway separately, we can have a better operational view of the whole pathway [1], [2]. Modules can also be compared among two different species and their dissimilarity can be used as a measure of their distance. These distances among a set of different species can be utilized to construct a phylogenetic tree namely the "Module Tree".

Phylogenetic trees can also be generated from signal transduction pathway by their size [3], [4], similarity of the nucleotide sequences [5], amino acid sequences [6], enzyme sequences [7], and protein structural classification [8]. Some other ways of alignment are based on functional similarity of the enzymes [9] and proteins [10], enzyme hierarchy and gene ontology [11], chemical structures or compound similarity [12]. Presence of common topological structures, *i.e.*, graphlets or sub-pathways [13], and presence or absence of pathways in an entire pathway repertoire [14] can also be utilized for pathway comparison and subsequently for phylogenetic tree construction.

Here, we have considered Wnt STPs of 48 species (Table I) ranging from placozoans to humans. Total number of genes discovered in the species-specific Wnt STPs, total number of modules created from a pathway, and the type and source of 18S rRNA sequence used [G: GenBank accession number (Complete sequence), G-P: GenBank accession number (Partial Sequence), G-S: GenBank accession number of predicted 18S rRNA sequence as given in SILVA database [15], and S-E: Sequence taken from Stage and EickBush, 2007 [16] are included in this table.

Wnt molecules are secreted cysteine-rich, lipid-modified glycoproteins. They bind to Frizzled seven-transmembrane-spanning receptors (FZDs) along with co-receptor LRPs (Lipoprotein Receptor-related Proteins) and initiate a cascade of reactions, which altogether form the Wnt STP [17]. The pathway is involved in crucial cellular functions, *i.e.*, regulation of fate determination, proliferation, differentiation, migration and apoptosis of cells [18], [19]. In matured organisms, Wnts are implicated in maintaining stem cell like fates in the intestinal epithelium [20], skin [21], and hematopoietic cells [22]. A number of studies has been dedicated to analyze the specific details of Wnt STPs in a few model organisms [23]–[25]. On the contrary, only a few investigations have been initiated to understand how this pathway itself has developed [26]. A proper phylogenetic tree can facilitate better understanding about Wnt STP development.

In this article we have created two alternative phylogenetic trees, *i.e.*, the module tree and the pathway tree, to study development of Wnt STPs. The trees have been created by considering modules and pathway topology (whole pathway) of species specific Wnt STPs respectively. Three datasets corresponding to 48, 29 and 12 species have been considered, the later two sets being subsets of the former one according to completeness of 18S rRNA sequence data and completeness of the Wnt STP itself respectively. These phylogenetic trees represent development of the Wnt

978-1-4673-6799-8/15/$31.00

Sl. No.	KEGG code	Binomial nomenclature	Common name	No. of genes	No. of modules	18S rRNA sequence id
1	aag	*A. aegypti*	Yellow fever mosquito	34	6	U65375 [G]
2	aga	*A. gambiae*	Mosquito	31	6	AM157179 [G]
3	ame	*A. mellifera*	Honey bee	38	7	AY703484 [G-P]
4	aml	*A. melanoleuca*	Giant Panda	59	8	GL196163 [G]
5	api	*A. pisum*	Pea aphid	32	6	U27819 [G]
6	bfo	*B. floridae*	Florida lancelet	45	7	M97571 [G]
7	bmy	*B. malayi*	Filaria	34	7	AAQA01003643 [G-S]
8	bta	*B. taurus*	Cow	59	8	NR_036642 [G]
9	cbr	*C. briggsae*	-	23	3	FJ380929 [G]
10	cel	*C. elegans*	Nematode	23	3	EU196001 [G-P]
11	cfa	*C. familiaris*	Dog	58	8	AAEX02007663 [G-S]
12	cin	*C. intestinalis*	Sea squirt	42	7	AB013017 [G-P]
13	cqu	*C. quinquefasciatus*	Southern house mosquito	35	7	AAWU01013261 [G-S]
14	dan	*D. ananassae*	-	37	7	XR_046314 [G]
15	der	*D. erecta*	-	37	7	XR_046906 [G]
16	dgr	*D. grimshawi*	-	37	7	[S-E]
17	dme	*D. melanogaster*	Fruit fly	37	7	M21017 [G]
18	dmo	*D. mojavensis*	-	37	7	XR_047783 [G]
19	dpe	*D. persimilis*	-	32	5	XR_046906 [G]
20	dpo	*D. pseudoobscura pseudoobscura*	-	32	6	XR_053284 [G]
21	dre	*D. rerio*	Zebrafish	59	8	AC139725 [G-S]
22	dse	*D. sechellia*	-	37	7	XR_048770 [G]
23	dsi	*D. simulans*	-	22	5	AY037174 [G]
24	dvi	*D. virilis*	-	38	7	XR_049279 [G]
25	dwi	*D. willistoni*	-	37	7	XR_049811 [G]
26	dya	*D. yakuba*	-	36	6	XR_050457 [G]
27	ecb	*E. caballus*	Horse	56	8	AJ311673 [G-P]
28	gga	*G. gallus*	Chicken	54	8	M59389 [G]
29	hmg	*H. magnipapillata*	-	31	6	HQ392522 [G-P]
30	hsa	*H. sapiens*	Human	60	8	X03205 [G]
31	isc	*I. scapularis*	Black-legged tick	30	6	ABJB010180167 [G-S]
32	mcc	*M. mulatta*	Rhesus Monkey	59	8	FJ436026 [G-P]
33	mdo	*M. domestica*	Opossum	55	7	AJ311676 [G-P]
34	mmu	*M. musculus*	Mouse	60	8	X00686 [G]
35	nve	*N. vectensis*	Sea anemone	33	6	AF254382 [G]
36	nvi	*N. vitripennis*	Jewel wasp	38	7	GQ410677 [G-P]
37	oaa	*O. anatinus*	Platypus	47	7	AJ311679 [G-P]
38	phu	*P. humanus corporis*	Human body louse	37	7	FJ267399 [G-P]
39	ptr	*P. troglodytes*	Chimpanzee	58	8	AADA01268803 [G-S]
40	rno	*R. norvegicus*	Rat	60	8	X01117 [G]
41	smm	*S. mansoni*	-	25	5	U65657 [G]
42	spu	*S. purpuratus*	Purple sea urchin	41	6	L28055 [G]
43	ssc	*S. scrofa*	Pig	21	4	AY265350 [G]
44	tad	*T. adhaerens*	-	23	6	Z22783 [G]
45	tca	*T. castaneum*	Red flour beetle	37	7	HM156711 [G-P]
46	tgu	*T. guttata*	Zebra finch	47	7	ABQF01063677 [G-S]
47	xla	*X. laevis*	African clawed frog	51	8	X04025 [G]
48	xtr	*X. tropicalis*	Western clawed frog	57	8	AAMC01103672 [G-S]

TABLE I
LIST OF SPECIES AND 18S rRNA REFERENCE IDS

STP obtained at module and pathway level. They were compared with the NCBI taxonomy and 18S rRNA trees for assessment of their quality in representing development of Wnt STP. The phylogenetic trees have been created with MEGA version 4.0.2 [27].

II. Data

Species-specific Wnt STPs in KEGG/Pathway database [28] has been taken as raw data (Table I). The pathway specific interactions were extracted from their corresponding KGML (KEGG Markup Language) files. The database uses a unique three letter code for each species along with their biological and common names (wherever applicable), *i.e.*, 'hsa' for *H. sapiens* (human). These three letter codes have been used extensively in this manuscript.

A. 18S rRNA Sequence Data

18S rRNA is a component of small eukaryotic ribosomal subunit (40S). 18S rRNA data is widely used in molecular analysis to reconstruct the evolutionary history and ancient divergences of organisms due to its slow evolutionary rate. Here, most of the 18S rRNA sequences have been taken from GenBank [29] for construction of the 18S rRNA tree. With a simple search dialogue of "— [organism] AND 18S ribosomal RNA [keyword] NOT (partial)", the sequence of interest can be extracted easily. If complete sequences are not available, the "NOT (partial)" dialogue can be omitted and a search for partial sequences can be done.

We have found 28 complete and 11 partial 18S rRNA nucleotide sequences for which GenBank accession numbers are listed in Table I. Eight sequences have been taken from the SILVA comprehensive ribosomal RNA databases (http://www.arb-silva.de/). It is an on-line resource for quality checked and aligned ribosomal RNA sequence data, which is free for academic use [15]. It provides regularly updated datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) rRNA sequences for all three domains of life (Bacteria, Archaea and Eukarya). We have taken sequences from the SSU r106 database and their respective GenBank accession numbers are given in Table I. 18S rRNA sequence of *D. grimshawi* has been taken from Stage and Eickbush, 2007 [16] as it is not available in GenBank or SILVA.

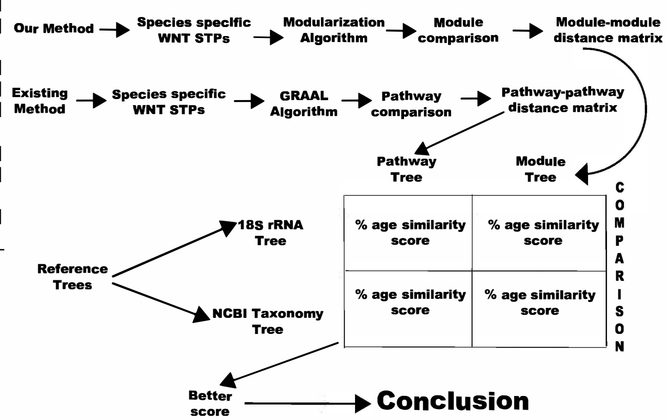


Fig. 1. Pipeline of the methodology

III. Methodology

Here we describe the proposed methodology (Figure 1) involved in creating phylogenetic trees from species specific Wnt STPs, taxonomy information from NCBI and 18S rRNA sequences. We have considered two different

```
PSEUDO-CODE FOR THE MODULARIZATION ALGORITHM
ENSURE Node Pool (universal set of connected nodes) is not empty
FOR Creating modules from a network DO
    REPEAT
       FOR Building a complete module DO
        Find start node (the node with maximum total degree)
        Exclude it from the node pool
        Include the central node's neighbors in the module
        REPEAT
            FOR Extending the module DO
            Check the total relation of added nodes
            IF all the relations are lying in the modules THEN
            The node is a permanent member of the module
            Exclude it from the node pool
            IF Number of relations lying outside the module is > c THEN
            Exclude it from the module
            Decrease its associated nodes' total degree by one
            IF Number of relations lying outside the module is <= c THEN
            Include further neighbors of the added nodes in the module
        UNTIL relations of each node present in the module are accounted for
    UNTIL Node pool is Empty
```

sets of factors to do system level development analysis of species-specific Wnt STPs. Our aim is to know the similarity in percentage of our constructed phylogenetic trees (constructed from pathway topology and modules) with respect to the standard evolutionary trees (18S rRNA and NCBI taxonomy tree).

A. Module tree method

The module tree has been generated solely based on one-to-one mapping of members present in modules of different species-specific Wnt STPs. A module can be defined as a subset of a STP that tends to be self-sufficient by maintaining minimal dependency on the remaining part of the STP. Modules have been created by the Modularization Algorithm developed earlier by the authors [1] based on a user defined factor c . The lowest possible c-value is always 1. The upper limit of c-value is the highest total degree of a node present in the considered network.

Number of modules found in a pathway depends on its size and complexity. Hence, varying number of modules can be found for different species-specific Wnt signal transduction pathways. Presence, absence or modification (increase/decrease due to addition/deletion of nodes) found in modules of a pair of species-specific pathways represent their distance. This approach is inspired by the NCE method described by Heymans and Singh, 2003 [9]. NCE method detects number of common enzymes between two pathways and tries to guess similarity based on that number. Here, rather than considering number of common enzymes, we are calculating number of common molecules present in the corresponding module of two species-specific Wnt signal transduction pathways. The score is then normalized by dividing it with the total number of non-redundant molecules present in both the species-specific modules. Hence, if pathway of species x has m_1 modules and pathway of species y has m_2 modules, we will get a similarity matrix of order $m_1 \times m_2$. Each element of the matrix represents similarity score between two different modules belonging to two different species (Equation 1). Let M_1 be a module of species x, *i.e.*, M_1 is the set of all the nodes in the module. Similarly, in a species y, M_2 is a set of nodes that constitute a module. The score of similarity $Sim(M_1, M_2)$ between module M_1 of species x and module M_2 of species y is defined as

$$Sim(M_1, M_2) = |M_1 \cap M_2| / |M_1 \cup M_2| \qquad (1)$$

Now the similarity $S_M(x, y)$ between species x and y is defined as

$$S_M(x, y) = \sum_{i=1,j=1}^{m_1,m_2} Sim(M_i, M_j)/(m_1 \times m_2) \qquad (2)$$

Distance score $D_M(x, y)$ is defined as

$$D_M(x, y) = 1 - S_M(x, y) \qquad (3)$$

These distance scores found among 48 different species are utilized for creation of the module tree. Our purpose in creating such a tree is to test its novelty in presenting the pathway's development.

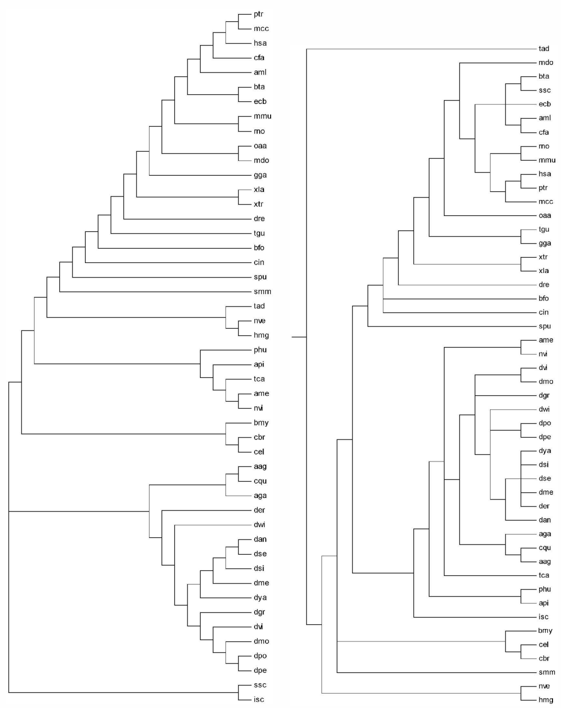
B. Pathway tree method

We have generated the pathway tree based on topological distances among species-specific pathways. Topological distance $D_P(x, y)$ between two pathways of species x and y has been defined as $D_P(x, y) = 1 - S_P(x, y)$ where $S_P(x, y)$ has been the topological similarity between the two species-specific pathways. $S_P(x, y)$ has been calculated by the GRAph ALigner algorithm (GRAAL) developed by Kuchaiev et al. [3] and implemented in the GraphCrunch2 software [30]. $S_P(x, y)$ is nothing but Edge Correctness (EC) value between a pair of species-specific Wnt STPs. Edge correctness is the percentage of edges in the first graph that are aligned to edges in the second graph. High edge correctness means the pair of networks considered share similar topologies.

The algorithm uses topological information based on graphlets in order to perform network alignment. Given two networks, GRAAL finds an embedding of the smaller network into the larger one such that every node in the smaller network is aligned to exactly one node in the larger one. It aims at exposing as much topological similarity between the networks as possible. It is a seed-and-extend algorithm that greedily aligns nodes based on their signature similarities while traversing both networks simultaneously in a breadth-first manner.

C. 18s rRNA tree method

Standard 18S rRNA sequences (Table I) have been used to create the 18S rRNA tree as shown in Figure 2(a). The evolutionary history has been inferred using the Neighbor-Joining method [31]. The optimal tree with the sum of branch length = 2.18407422 has been considered. The evolutionary distances have been computed in the units of the number of base substitutions per site using the Maximum Composite Likelihood method [32]. All positions containing gaps and missing data are eliminated only in pairwise sequence comparisons (Pairwise deletion option). There are a total of 2687 positions in the final dataset.



(a) The 18S rRNA tree

(b) The NCBI taxonomy tree

Fig. 2. The Reference Trees

D. NCBI taxonomy tree method

The NCBI taxonomy tree (Figure 2(b)) has been created with the help of NCBI taxonomy database[1] [33]. Newick format of the tree has been saved as a text tree after adding organism names in the "Taxonomy Common tree" page.

E. Tree comparison method

Nye et al. [34] have developed an algorithm that pairs up each branch in one phylogenetic tree with a matching branch in the second one, and finds the optimum one-to-one map between branches in the two trees in terms of a topological score. They have developed an applet[2] in Java that enables one to explore the corresponding mapping between the phylogenetic trees interactively, and clearly highlights similar/ different parts of the trees, both in terms of topology and branch length. Here, we have considered topology mainly.

[1]http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi
[2]http://www.mas.ncl.ac.uk/∼ntmwn/phylo_comparison/pairwise.html

Let us now describe the algorithm that compares two phylogenetic trees created from the same set of species. Given two phylogenetic trees T_1 and T_2 that share the same set of leaves L, the algorithm firstly assigns a score $s(i, j)$ to every pair of edges (i, j) with $i \in T_1$ and $j \in T_2$. Then the algorithm pairs up branches in the two trees to optimize the overall score. This is equivalent to finding a bijection (*i.e.*, a one-to-one and onto correspondence) $f : T_1 \rightarrow T_2$ between the branches of the trees that maximizes the quantity $\sum_{i \in T_1} S(i, f(i))$.

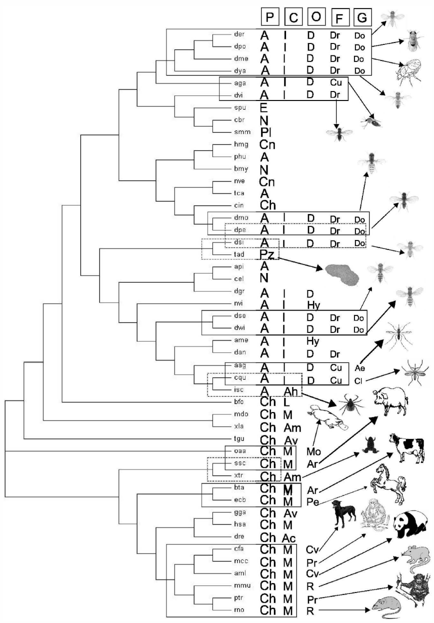
Notation	Rank	Name	Notation	Rank	Name
P	Phylum	-	R	Order	Rodentia
A	Phylum	Arthropoda	Ar	Order	Artiodactyla
Cn	Phylum	Cnidaria	Pe	Order	Perissodactyla
N	Phylum	Nematoda	Pr	Order	Primates
Ch	Phylum	Chordata	Aa	Order	Anura
E	Phylum	Echinodermata	Di	Order	Didelmorphia
Pl	Phylum	Platyhelminthes	Cv	Order	Carnivora
Pz	Phylum	Placozoa	Rh	Order	Rhabditida
C	Class	-	F	Family	-
I	Class	Insecta	Dr	Family	Drosophilidae
Ah	Class	Arachnida	Mu	Family	Muridae
Av	Class	Aves	Pi	Family	Pipidae
Am	Class	Amphibia	Cd	Family	Canidae
Ac	Class	Actinopterygii	U	Family	Ursidae
M	Class	Mammalia	Cu	Family	Culicidae
As	Class	Ascidiacea	Ra	Family	Rhabditidae
L	Class	Leptocardii	G	Genus	-
S	Class	Secernentea	Do	Genus	Drosophila
O	Order	-	Ms	Genus	Mus
D	Order	Diptera	Rt	Genus	Rattus
Hy	Order	Hymenoptera	X	Genus	Xenopus
Mo	Order	Monotremata	Ae	Genus	Aedes
Co	Order	Coleoptera	Cl	Genus	Culex
Ph	Order	Phthiraptera	Al	Genus	Anophelinae
He	Order	Hemiptera	Cr	Genus	Caenorhabditis

TABLE II
LIST OF NOTATIONS USED IN FIGURE 3(A) AND FIGURE 3(B)

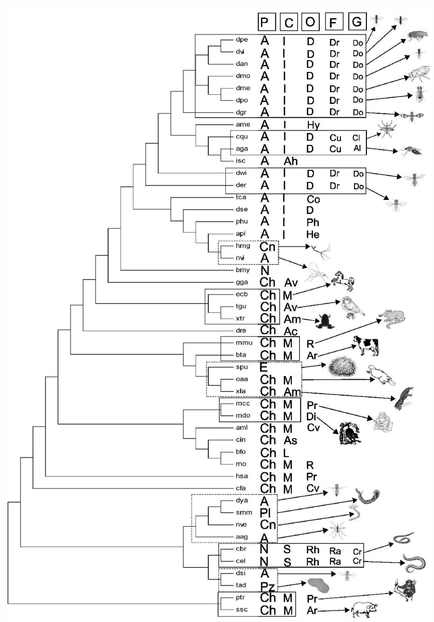
IV. Results and Discussions

In this section, we describe ranks and positional significance of the species in the pathway tree (Figure 3(a)) and the module tree (Figure 3(b)). The species with similar taxonomic ranks coming under the same clade have been marked by continuous rectangles. The species coming under different clades despite having similar taxonomy have been marked by dotted rectangles. While discussing positional significance of species, we have furnished the similarities in terms of taxonomic ranks (phylum, class and others) to the lowest possible taxonomic rank. Throughout this manuscript we have used some common notations while analyzing the phylogenetic trees. The notations are listed alphabetically in Table II.

We have compared Wnt STPs of 48 different species as provided in Table I for creating a module tree (Figure 3(a)) and a pathway tree (Figure 3(b)). The considered species belong to seven different phyla, most of which (21) belong to the phylum Arthropoda followed by Chordata (19), Nematoda (3), Cnidaria (2), and single species from phyla Echinodermata, Placozoa and Platyhelminthes. As expected, some species have been placed closely in the



(a) The Module Tree



(b) The Pathway Tree

Fig. 3. The Constructed Trees

tree following their taxonomic ranks. On the other hand, we have also found some deviations.

A. *Finding a Better Tree*

Closely related species come under a single clade in both the module and pathway trees. In this sense, they have shown preservation of contemporary notions regarding development of species. But, they have shown numerous deviations also, when evolutionarily distant species came under a single clade or species having similar taxonomy got included in different clades. In order to put a universal measure to their quality, and to determine the better tree that represents development of Wnt STP, we have followed the concept of alternative phylogenetic tree comparison [34]. A brief description of this method of comparison is furnished in Section III-E.

We have compared the module and pathway trees with the reference trees (NCBI taxonomy tree and 18S rRNA tree). The reference trees represent phylogeny of living organisms from multiple point-of-views while the phylogenetic trees derived from species-specific Wnt STPs solely represent development of Wnt STP over the taken set of species. So a huge gap can be noticed among the two sets of phylogenetic trees in terms of similarity percentage. Still, a more similar tree is better than a less similar tree for development analysis.

The module tree has showed 42.4% topological similarity with the NCBI taxonomy tree and 42.2% similarity with the 18S rRNA tree followed by the pathway tree (38.2% and 37.1% respectively) as given in Table III. Hence, the module tree has outperformed the pathway tree in representing Wnt STP development. However, among 48 18S rRNA sequences, 11 are partial and 8 are predicted. Did the incomplete sequences influence the 18S rRNA tree (e.g. two unrelated species are clustered together because of lack of the same part of 18s rRNA sequences)? To avoid such a notion, we have repeated our protocol with a dataset of 29 species (aag, aga, aml, api, bfo, bta, cbr, dan, der, dgr, dme, dmo, dpe, dpo, dse, dsi, dvi, dwi, dya, gga, hsa, mmu, nve, rno, smm, spu, ssc, tad and xla), for which complete 18S rRNA sequences are available as given in Table I. The module tree has also outperformed the pathway tree, for these species.

	48 species		29 species		12 species	
	NCBI Taxonomy tree	18S rRNA tree	NCBI Taxonomy tree	18S rRNA tree	NCBI Taxonomy tree	18S rRNA tree
Module tree	42.4	42.2	48.9	39	45.4	55.4
Pathway tree	38.2	37.1	38.2	36.2	36.7	37.3

TABLE III
SIMILARITIES AMONG TREES FOR 48, 29 & 12 SPECIES IN PERCENTAGE

Some of the pathways in our dataset comprising 48 species-specific pathways are partially known as the underlying graph structures of pathways (KEGG) are highly incomplete. Some model organisms are better studied than others. This could cause a bias and the pathway-based graph may be more prone to this bias. To avoid such a bias, and in order to strengthen our results, we have considered Wnt STP of a smaller and more complete pathway dataset of 12 species (aml, bta, cfa, dre, ecb, hsa, mcc, mdo, mmu, ptr, rno and xtr). These pathways have varying number (55-60) of nodes. For this dataset too, the module tree has showed maximum similarity with the NCBI taxonomy tree (45.4%) and the 18S rRNA tree (55.4%) as given in Table III.

V. Conclusion

This article emphasized on development of Wnt STP over various species. Here, we have created two alternate phylogenetic trees, *i.e.*, the pathway and module trees from three datasets of species-specific pathways (comprising 48, 29 and 12 species), and compared them with two reference trees (the NCBI taxonomy tree and the 18S rRNA tree). The module tree is found to be more similar to the reference trees than the pathway tree. Hence the module tree is a better candidate to represent Wnt STP development. The increased performance of the module tree is due to consideration of local similarities, which probably we ignore in a global scenario. This concept of taking modules/sub-units/sub-sets of pathways (local information) in to construction of a phylogenetic tree rather than considering the whole pathways (global information) can be extended to other fields of phylogenetic analysis.

VI. Acknowledgments

L. Nayak acknowledges University Grants Commission, India for a UGC Post-Doctoral Fellowship (No. F.15-1/2013-14/PDFWM-2013-14-GE-ORI-19068(SA-II)).

References

[1] L. Nayak and R. K. De, "An algorithm for modularization of MAPK and calcium signaling pathways: Comparative analysis among different species," *Journal of biomedical informatics*, vol. 40, no. 6, pp. 726–749, 2007.

[2] ——, "Developmental trend derived from modules of Wnt signaling pathways," in *Pattern Recognition and Machine Intelligence*. Springer, 2011, pp. 400–405.

[3] O. Kuchaiev, T. Milenković, V. Memišević, W. Hayes, and N. Pržulj, "Topological network alignment uncovers biological function and phylogeny," *Journal of the Royal Society Interface*, no. rsif.2010.0063, 2010.

[4] R. Y. Pinter, O. Rokhlenko, E. Yeger-Lotem, and M. Ziv-Ukelson, "Alignment of metabolic pathways," *Bioinformatics*, vol. 21, no. 16, pp. 3401–3408, 2005.

[5] C. V. Forst and K. Schulten, "Phylogenetic analysis of metabolic pathways," *Journal of Molecular Evolution*, vol. 52, no. 6, pp. 471–489, 2001.

[6] B. P. Kelley, B. Yuan, F. Lewitter, R. Sharan, B. R. Stockwell, and T. Ideker, "PathBLAST: a tool for alignment of protein interaction networks," *Nucleic acids research*, vol. 32, no. suppl 2, pp. W83–W88, 2004.

[7] M. A. Ovacik and I. P. Androulakis, "Enzyme sequence similarity improves the reaction alignment method for cross-species pathway comparison," *Toxicology and applied pharmacology*, vol. 271, no. 3, pp. 363–371, 2013.

[8] S. Hariharaputran, T. Töpel, T. Oberwahrenbrock, and R. Hofestädt, "Alignment of linear biochemical pathways using protein structural classification," *Nature Precedings*, 2008.

[9] M. Heymans and A. K. Singh, "Deriving phylogenetic trees from the similarity analysis of metabolic pathways," *Bioinformatics*, vol. 19, no. suppl 1, pp. i138–i146, 2003.

[10] W. Ali and C. M. Deane, "Functionally guided alignment of protein interaction networks for module detection," *Bioinformatics*, vol. 25, no. 23, pp. 3166–3173, 2009.

[11] J. C. Clemente, K. Satou, and G. Valiente, "Reconstruction of phylogenetic relationships from metabolic pathways based on the enzyme hierarchy and the gene ontology," *GENOME INFORMATICS SERIES*, vol. 16, no. 2, p. 45, 2005.

[12] Y. Tohsato and Y. Nishimura, "Metabolic pathway alignment based on similarity between chemical structures," *Information and Media Technologies*, vol. 3, no. 1, pp. 191–200, 2008.

[13] F. Ay, M. Kellis, and T. Kahveci, "SubMAP: aligning metabolic pathways with subnetwork mappings," *Journal of Computational Biology*, vol. 18, no. 3, pp. 219–235, 2011.

[14] L. Liao, S. Kim, and J.-F. Tomb, "Genome comparisons based on profiles of metabolic pathways," in *Proceedings of the 6th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES'02).* Citeseer, 2002, pp. 469–476.

[15] E. Pruesse, C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glckner, "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB," *Nucleic acids research*, vol. 35, no. 21, pp. 7188–7196, 2007.

[16] D. E. Stage and T. H. Eickbush, "Sequence variation within the rRNA gene loci of 12 Drosophila species," *Genome Res*, vol. 17, no. 12, pp. 1888–1897, 2007.

[17] K. M. Cadigan and Y. I. Liu, "Wnt signaling: complexity at the surface," *Journal of Cell Science*, vol. 119, no. 3, pp. 395–402, 2006.

[18] R. T. Moon and K. Shah, "Developmental biology: Signalling polarity," *Nature*, vol. 417, no. 6886, pp. 239–240, 2002.

[19] K. Willert, J. D. Brown, E. Danenberg, A. W. Duncan, I. L. Weissman, T. Reya, J. R. Yates, and R. Nusse, "Wnt proteins are lipid-modified and can act as stem cell growth factors," *Nature*, vol. 423, no. 6938, pp. 448–452, 2003.

[20] D. Pinto and H. Clevers, "Wnt, stem cells and cancer in the intestine," *Biology of the Cell*, vol. 97, no. 3, pp. 185–196, 2005.

[21] W. Lowry, C. Blanpain, J. Nowak, G. Guasch, L. Lewis, and E. Fuchs, "Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells," *Genes Dev*, vol. 19, no. 13, pp. 1596–1611, 2005.

[22] T. Reya, A. W. Duncan, L. Ailles, J. Domen, D. C. Scherer, K. Willert, L. Hintz, R. Nusse, and I. L. Weissman, "A role for Wnt signalling in self-renewal of haematopoietic stem cells," *Nature*, vol. 423, no. 6938, pp. 409–414, 2003.

[23] M.-A. Félix, "An inversion in the wiring of an intercellular signal: evolution of wnt signaling in the nematode vulva," *BioEssays*, vol. 27, no. 8, pp. 765–769, 2005.

[24] B. Hendriks and E. Reichmann, "Wnt signaling: a complex issue," *Biological research*, vol. 35, no. 2, pp. 277–286, 2002.

[25] R. Janssen, M. Le Gouar, M. Pechmann, F. Poulin, R. Bolognesi, E. E. Schwager, C. Hopfen, J. K. Colbourne, G. E. Budd, S. J. Brown, *et al.*, "Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation," *BMC evolutionary biology*, vol. 10, no. 1, p. 374, 2010.

[26] A. Pires-daSilva and R. J. Sommer, "The evolution of signalling pathways in animal development," *Nature Reviews Genetics*, vol. 4, no. 1, pp. 39–49, 2003.

[27] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular biology and evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.

[28] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic acids research*, vol. 28, no. 1, pp. 27–30, 2000.

[29] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic acids research*, p. gks1195, 2012.

[30] O. Kuchaiev, A. Stevanović, W. Hayes, and N. Pržulj, "GraphCrunch 2: software tool for network modeling, alignment and clustering," *BMC bioinformatics*, vol. 12, no. 1, p. 24, 2011.

[31] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular biology and evolution*, vol. 4, no. 4, pp. 406–425, 1987.

[32] K. Tamura, M. Nei, and S. Kumar, "Prospects for inferring very large phylogenies by using the neighbor-joining method," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 30, pp. 11 030–11035, 2004.

[33] S. Federhen, "The NCBI taxonomy database," *Nucleic acids research*, vol. 40, no. D1, pp. D136–D143, 2012.

[34] T. M. Nye, P. Lio, and W. R. Gilks, "A novel algorithm and web-based tool for comparing two alternative phylogenetic trees," *Bioinformatics*, vol. 22, no. 1, pp. 117–119, 2006.