Developmental Trend Derived from Modules of Wnt Signaling Pathways

Losiana Nayak and Rajat K. De

Machine Intelligence Unit, Indian Statistical Institute
203 B. T. Road, Kolkata - 700108, India
{losiana_t,rajat}@isical.ac.in
http://www.isical.ac.in/~{losiana_t,rajat}

Abstract. In this paper, we deal with the idea of creating a developmental trend from Wnt signaling pathways of different species. Wnt signaling pathway is involved in many crucial biological processes including from early embryonic development to stem cell management at later stages. The pathway varies in topology and size for each species that gets reflected in its modules. A comparison among species-specific pathways, taking into account the modules and pathway structure (in terms of nodes and edges) will throw light on crucial turning points in the development of Wnt signaling pathway. Hence, 31 species-specific Wnt signaling pathways have been modularized by the Modularization algorithm already developed by the authors. The modules were compared among themselves to find the trend of development. The trend established conserved modules among these pathways.

Keywords: Modularization algorithm, Evolution, Phylogenetic tree construction, Computational Phylogenetics.

1 Introduction

In biological terms, a signal transduction pathway, is a set of established genes and related factors, which operate in synchronous manner to create cascades of reactions, ultimately generating a response to stimuli *in vivo*. From graph theoretical point of view, these genes and related factors can be considered as nodes and the interactions among them as edges of a network. A pathway conceived in such a way is open to all kind of network analysis paradigms. Network comparison (by alignment) to uncover biological functions and phylogeny [1] is one of them. Networks derived from biological pathways (gene regulatory, metabolic, signal transduction, protein-protein interaction networks) can be aligned by size of the network, sequence similarity of the genes/proteins, functional similarity of the enzymes/proteins and presence of common topological structures (graphlets) among others. One or all of these factors are considered while creating a tree from a set of biological networks. In addition to such factors, we add another factor named 'modules' [2]. Modules throw light on operational sophistication of a network. It is dependent on two other parameters, *viz.*, size (nodes) and topology (interactions). A tree generated by taking these three linked parameters can shed enough light on subtle changes of the network among the species,

S.O. Kuznetsov et al. (Eds.): PReMI 2011, LNCS 6744, pp. 400–405, 2011.

if not all. Here, these ideas were implemented for creating a tree from a set of species-specific Wnt signaling pathways.

Table 1. The list of species taken from KEGG/ PATHWAY database. The database uses a unique three letter code, viz., 'hsa' for *H. sapiens* (human) for each species along with their biological and common names (wherever applicable).

Sl. No.	Species Name	Common Name	KEGG Code
01	*H. sapiens*	Human	hsa
02	*M. musculus*	Mouse	mmu
03	*R. norvegicus*	Rat	rno
04	*B. taurus*	Cow	bta
05	*C. familiaris*	Dog	cfa
06	*P. troglodytes*	Chimpanzee	ptr
07	*M. mulatta*	Rhesus Monkey	mcc
08	*M. domestica*	Opossum	mdo
09	*G. gallus*	Chicken	gga
10	*D. rerio*	Zebrafish	dre
11	*X. laevis*	African clawed frog	xla
12	*S. purpuratus*	Purple sea urchin	spu
13	*X. tropicalis*	Western clawed frog	xtr
14	*D. melanogaster*	Fruitfly	dme
15	*E. caballus*	Horse	ecb
16	*N. vectensis*	Sea anemone	nve
17	*A. mellifera*	Honey bee	ame
18	*D. pseudoobscura pseudoobscura*	-	dpo
19	*T. castaneum*	Red flour beetle	tca
20	*A. aegypti*	Yellow fever mosquito	aag
21	*O. anatinus*	Platypus	oaa
22	*C. elegans*	Nematode	cel
23	*A. gambiae*	Mosquito	aga
24	*S. scrofa*	Pig	ssc
25	*B. floridae*	Florida lancelet	bfo
26	*C. intestinalis*	Sea squirt	cin
27	*D. ananassae*	-	dan
28	*B. malayi*	Filaria	bmy
29	*A. pisum*	Pea aphid	api
30	*T. adhaerens*	-	tad
31	*C. briggsae*	-	cbr

Wnt molecules are secreted cysteine-rich, lipid-modified glycoproteins. They bind to Frizzled seven-transmembrane-span receptors (FZDs) along with co-receptor LRPs (Lipoprotein Receptor-related Proteins) and initiate the downstream steps. These protein-protein and protein-DNA interactions altogether are known as Wnt signaling pathway [3]. Wnt signaling pathway is involved in regulation of cell fate determination, proliferation, differentiation, migration and apoptosis [4]. It enables cells to influence behavior of their neighboring cells during development [5]. In matured organisms, Wnts are implicated in maintaining stem cell-like fates in the intestinal epithelium [6], skin [7] and hematopoietic cells [8].

In this article, we have created modules from Wnt signaling pathways of 31 different species. By comparing these modules, a distance score is established between each pair of species and such scores were utilized in creating a tree. The tree reflects the course of development of Wnt signaling pathway among the taken set of species along with detection of conserved modules.

2 Methodology

Species-specific Wnt signaling pathways of KEGG/ Pathway database [9] were taken as data. The database is maintained by the Kanehisa Laboratories, Bioinformatics Center, Kyoto University and the Human Genome Center, University

of Tokyo. It is a collection of manually drawn pathway maps, whose XML data files along with KGML and PNG diagrams are publicly accessible. Detail information of these species is given in Table 1. The database uses a unique three letter code for each species along with their biological and common names (wherever applicable), *viz.*, 'hsa' for *H. sapiens* (human). These three letter codes are used extensively in this manuscript.

The species-specific pathways were modularized by the algorithm developed by Nayak et al., 2007 [2]. The algorithm creates a set of modules, for each value of the user defined parameter c. The parameter decides whether a node belongs to a particular module or not. A node gets excluded from a module, if it has more than c relations that lie outside the module boundary. In general, a range of c-value is fixed for running the modularization algorithm. The lowest possible c-value is the lowest total degree of a node, while the upper limit of c-value is the highest total degree of a node found in a considered network. The task is to choose the ideal c-value that can create best partitions. *Modus operandi* of the algorithm can be followed by going through the pseudocode (Algorithm 1). Implementation of the algorithm is done in C. Wnt signaling pathways of the above-mentioned species were subjected to modularization for c=3 (as meaningful modules have been found from Wnt signaling pathway of hsa for the same c-value). We are getting 2 to 8 modules for each species that vary in their size (number of nodes present in the module) as shown in Table 2. It gives module details (number of connected nodes, relations and modules) of all the considered pathways.

Algorithm 1. [Pseudocode for 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/central 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 All the relations of each node present in the module are accounted for
  until Node pool is Empty
```

The tree of development was created by taking into account distance between three factors, *i.e.*, the number of connected nodes, the number of relations and the number of modules. For example, distance between hsa and spu is $(|60 - 39|) + (|70 - 45|) + (|8 - 6|)/3 = 16$. Average distance (if the mod value is found to be non-zero) between these three parameters of different species were used to create a distance matrix. A tree (Figure 1) was constructed using this matrix by Phylip package (Version 3.6) [10].

3 Results and Discussion

In this section, we have compared modules of 31 different species (aag, aga, ame, api, bfo, bmy, bta, cbr, cel, cin, cfa, dan, dme, dpo, dre, ecb, gga, hsa, mcc, mdo, mmu, nve, oaa, ptr, rno, ssc, spu, tad, tca, xla and xtr). Module details are given in Table 2.

Table 2. Module information of species-specific Wnt signaling pathways. [sp.: three lettered species code, n: number of connected nodes in a species-specific pathway; r: number of relations present the connected component of a species-specific pathway; t: total number of modules created from a species-specific pathway]. The modules have been created for c=3. Each module's size in terms of nodes is given with it in parentheses. The table throws light on the developmental trend of Wnt signaling pathways among the taken set of species.

sp.	n	r	t	WNT	(DVL)1	Axin	β-catenin	TCF	p53	(DVL)2	PLC
hsa	60	70	8	WNT [8]	(DVL)1 [7]	Axin [4]	β-catenin [8]	TCF [14]	p53 [2]	(DVL)2 [10]	PLC [7]
mmu	60	70	8	WNT [8]	(DVL)1 [7]	Axin [4]	β-catenin [8]	TCF [14]	p53 [2]	(DVL)2 [10]	PLC [7]
rno	59	69	8	WNT [7]	(DVL)1 [7]	Axin [4]	β-catenin [8]	TCF [14]	p53 [2]	(DVL)2 [10]	PLC [7]
bta	58	68	8	WNT [7]	(DVL)1 [6]	Axin [4]	β-catenin [8]	TCF [14]	P53 [2]	(DVL)2 [10]	PLC [7]
cfa	58	68	8	WNT [8]	(DVL)1 [7]	Axin [4]	β-catenin [7]	TCF [13]	p53 [2]	(DVL)2 [10]	PLC [7]
ptr	58	67	8	WNT [8]	(DVL)1 [7]	Axin [4]	β-catenin [8]	TCF [13]	p53 [2]	(DVL)2[10]	PLC [6]
mcc	55	63	8	WNT [7]	(DVL)1 [6]	Axin [4]	β-catenin [8]	TCF [13]	p53 [2]	(DVL)2 [8]	PLC [7]
mdo	54	64	7	WNT [8]	(DVL)1 [7]	Axin[2]	β-catenin [9]	TCF [11]	-	(DVL)2 [10]	PLC [7]
gga	54	63	8	WNT [7]	(DVL)1 [6]	Axin [3]	β-catenin [8]	TCF [11]	p53 [2]	(DVL)2 [10]	PLC [7]
dre	52	60	7	WNT [8]	-	Axin [4]	β-catenin [7]	TCF [13]	p53 [2]	(DVL)2 [11]	PLC [7]
xla	43	45	6	WNT [7]	-	-	β-catenin [8]	TCF [11]	p53 [2]	(DVL)2 [8]	PLC [7]
spu	39	45	6	-	(DVL)1 [7]	Axin [2]	β-catenin [5]	TCF [10]	-	(DVL)2 [9]	PLC [6]
xtr	37	36	6	WNT [3]	-	-	β-catenin [7]	TCF [6]	p53 [2]	(DVL)2 [12]	PLC [7]
dme	36	42	7	WNT [6]	(DVL)1 [5]	Axin [2]	β-catenin [6]	TAK1 [2]	-	(DVL)2 [9]	PLC [6]
ecb	36	38	7	(Frizzled)1 [5]	(DVL)1 [3]	-	β-catenin [9]	TAK1 [2]	p53 [2]	(DVL)2 [8]	PLC [7]
nve	32	33	6	(Frizzled)1 [5]	-	Axin [5]	β-catenin [6]	TAK1 [2]	-	(DVL)2 [7]	PLC [7]
ame	30	32	5	-	(DVL)1 [4]	-	β-catenin [9]	TAK1 [2]	-	(DVL)2 [8]	PLC [7]
dpo	28	30	4	(Frizzled)1 [8]	-	-	β-catenin [7]	-	-	(DVL)2 [8]	PLC [5]
tca	26	27	4	(Frizzled)1 [7]	-	-	β-catenin [7]	-	-	(DVL)2 [6]	PLC [6]
aag	24	22	4	(Frizzled)1 [4]	-	-	β-catenin [4]	-	-	(DVL)2 [10]	PLC [6]
oaa	22	22	4	WNT [2]	-	-	β-catenin [7]	-	-	(DVL)2 [8]	PLC [5]
cel	22	20	3	-	-	-	β-catenin [10]	-	-	RhoA [6]	PLC [6]
aga	20	18	3	(Frizzled)1 [11]	-	-	β-catenin [4]	-	-	-	PLC [5]
ssc	19	16	4	FRP [2]	-	-	β-catenin [5]	TCF [7]	-	-	PLC [5]
bfo	18	16	3	-	-	-	β-catenin [9]	-	-	(DVL)2 [5]	PLC [4]
cin	17	14	3	-	(DVL)1 [7]	-	-	-	-	(DVL)2 [5]	PLC [5]
dan	16	12	4	-	(DVL)1 [2]	-	β-catenin [4]	-	-	(DVL)2 [5]	PLC [5]
bmy	13	11	3	-	(DVL)1 [4]	-	-	-	-	(DVL)2 [5]	PLC [4]
api	13	10	3	-	(DVL)1 [4]	-	-	-	-	(DVL)2 [5]	PLC [4]
tad	6	4	2	-	-	-	-	-	-	Rac [2]	PLC [4]
cbr	4	3	1	-	(DVL)1 [4]	-	-	-	-	-	-

Comparison of the modules brought forward their functional conservation among the taken species. Modules Wnt and $\beta - catenin$ were found to be conserved in 9 species (hsa, mmu, rno, bta, cfa, ptr, mcc, mdo and gga). Module TCF was found to be conserved in 5 species (hsa, mmu, rno, bta and cfa). Module $Tp53$ was observed in altogether 12 species (hsa, mmu, rno, bta, cfa, ptr, mcc, gga, dre, xla, xtr and ecb) and it was conserved by size and topology in all these species. Module $(DVL)2$ was conserved in 11 species (hsa, mmu, rno, bta, cfa, ptr, mdo, gga, dre, spu and dme) and module PLC turned out to be the most conserved module, found in a maximum number of 17 species (hsa, mmu, rno, bta, cfa, ptr, mcc, mdo, gga, dre, xla, spu, xtr, dme, ecb, nve and ame).

Figure 1 provides an incidental peek of Wnt signaling pathway development in different species. The tree is analyzed by considering taxonomy of the taken

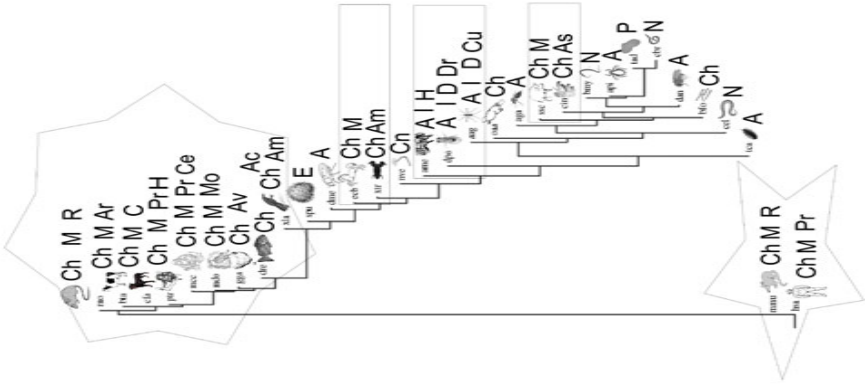

Fig. 1. Tree created from 31 species-specific Wnt signaling pathways. The species, belonging to the same Phylum, Class, Order and Family, which show proximity among them, are marked separately. [A: Phylum Arthropoda, Ch: Phylum Chordata, Cn: Phylum Cnidaria, E: Phylum Echinodermata, N: Phylum Nematoda, P: Phylum Placozoa, Ac: Class Actinopterygii, Am: Class Amphibia, As: Class Ascidiacea, Av: Class Aves, I: Class Insecta, M: Class Mammalia, Ar: Order Artiodactyla, C: Order Carnivora, D: Order Diptera, H: Order Hymenoptera, Mo: Order Monotremata, Pr: Order Primates, R: Order Rodentia, Ce: Family Cercopithecidae, Cu: Family Culicidae, Dr: Family Drosophilidae, H: Family Hominidae].

species. Many species of Phylum Chordata [(hsa, mmu, rno, bta cfa, ptr, mcc, mdo, gga, dre and xla), (ecb and xtr) and (ssc and cin)] and some species of Phylum Arthropoda (ame, dpo and aag) are found to be placed closer to each other. Further closeness among the Chordates is found as we proceed towards Class specifications. All of them belong to Class Mammalia with exception of five species [gga: Class Aves, dre: Class Actinopterygii, (xla and xtr): Class Amphibia and cin: Class Ascidiacea]. These species show gradual divergence. All the three Arthropods belong to Class Insecta. Among the closely placed Mammalians, three species (hsa, ptr and mcc) belong to Order Primates and two species (mmu and rno) to Order Rodentia. Two species of Class Insecta (dpo and aag) belong to Order Diptera. These species are found in proximity of each other in the tree, indicating towards a further level of conservation in development. Further specification in terms of Family throws some light on variation in development among these species. The species, belonging to the same Phylum, Class, Order and Family, which show proximity in the tree, are marked separately (Figure 1).

4 Conclusions

This article emphasizes on deriving the developmental trend from modules of species-specific Wnt signaling pathways. Here, we have done a comparative study among 31 species-specific set of modules. Module *PLC* is the most conserved module, found in a maximum number of 17 species. A developmental trend was

established among the modules (created by the modularization algorithm) of these 31 species. This study established that hsa and mmu have the most developed Wnt signaling pathway, followed by several species like rno, bta, mcc among others. The species having the least developed Wnt signaling pathway is cbr. It possesses a single rudimentary module $(DVL)1$. The developmental tree displays conservation at Class level and gradual divergence as we proceed towards the lower ranks, in accordance to basic principle of evolution. But, quite a number of discrepancies were also found that defy the general notion of evolution. They may turn out to be environmental influence on development of the pathway, if prodded further. In short, this work displays conservation of Wnt signaling pathway at Phylum level that gradually decreases as the Phyla diversify into various Classes, Orders and Families, with some exceptions that possibly reflect effect of other factors on pathway development.

Acknowledgement. Losiana Nayak is grateful to the CSIR, India, for providing her a Senior Research Fellowship [No. 9/93(102)08].

References

1. Kuchaiev, O., Milenkovic, T., Memisevic, V., Hayes, W., Przulj, N.: Topological network alignment uncovers biological function and phylogeny. J. R. Soc. Interface 7, 1341–1354 (2010)
2. Nayak, L., De, R.K.: An algorithm for modularization of MAPK and calcium signaling pathways: Comparative analysis among different species. J. Biomed. Inform. 40, 726–749 (2007)
3. Cadigan, K.M., Liu, Y.I.: Wnt signaling: complexity at the surface. J. Cell Sci. 119, 395–402 (2006)
4. Willert, K., Brown, J.D., Danenberg, E., Duncan, A.W., Weissman, I.L., et al.: Wnt proteins are lipid-modified and can act as stem cell growth factors. Nature 423, 448–452 (2003)
5. Logan, C.Y., Nusse, R.: The Wnt signaling pathway in development and disease. Annu. Rev. Cell Dev. Biol. 20, 781–810 (2004)
6. Pinto, D., Clevers, H.: Wnt, stem cells and cancer in the intestine. Biol. Cell 97, 185–196 (2005)
7. Lowry, W.E., Blanpain, C., Nowak, J.A., Guasch, G., Lewis, L., et al.: Defining the impact of β-catenin/tcf transactivation on epithelial stem cells. Genes and Dev. 19, 1596–1611 (2005)
8. Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C.: A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature 423, 409–414 (2003)
9. Kanehisa, M., Goto, S.: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30 (2000)
10. Felsenstein, J.: PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166 (1989)