

Shyamasree Ghosh<sup>1</sup>

# Application of Computational Methods in Planaria Research: A Current Update

<sup>1</sup> School of Biological Sciences, National Institute of Science Education and Research (NISER), Bhubaneswar, Odisha 752050, India; and Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India, E-mail: sree.s@niser.ac.in

## Abstract:

Planaria is a member of the Phylum Platyhelminthes including flatworms. Planarians possess the unique ability of regeneration from adult stem cells or neoblasts and finds importance as a model organism for regeneration and developmental studies. Although research is being actively carried out globally through conventional methods to understand the process of regeneration from neoblasts, biology of development, neurobiology and immunology of Planaria, there are many thought provoking questions related to stem cell plasticity, and uniqueness of regenerative potential in Planarians amongst other members of Phylum Platyhelminthes. The complexity of receptors and signalling mechanisms, immune system network, biology of repair, responses to injury are yet to be understood in Planaria. Genomic and transcriptomic studies have generated a vast repository of data, but their availability and analysis is a challenging task. Data mining, computational approaches of gene curation, bioinformatics tools for analysis of transcriptomic data, designing of databases, application of algorithms in deciphering changes of morphology by RNA interference (RNAi) approaches, understanding regeneration experiments is a new venture in Planaria research that is helping researchers across the globe in understanding the biology. We highlight the applications of Hidden Markov models (HMMs) in designing of computational tools and their applications in Planaria decoding their complex biology.

**Keywords:** Datamining, Computational Biology, Bioinformatics, Planaria, regeneration

**DOI:** 10.1515/jib-2017-0007

**Received:** February 27, 2017; **Revised:** April 4, 2017; **Accepted:** April 13, 2017

## 1 Introduction

Phylum Platyhelminthes belongs to the primitive bilateria group including diverse flatworms with features of triploblastic unsegmented, acoelomate body, fresh water, terrestrial, free living or parasitic forms. Body shows the absence of circulatory system, use of cilia instead of muscles for locomotion and the presence of protonephridia for excretion. They include four classes of flatworms including Classes Trematoda, Monogenea and Cestoidea which show entirely parasitic existence. The fourth class Turbellaria are mostly aquatic, marine, free living with few pelagic forms.

*Planaria* is a member of order Tricladida under Class Turbellaria (Table 1) with a small sized body ranging from microscopic to more than 60 cm in length. Turbellarians are characterised by the presence of gland cells located either in the epidermis or completely submerged below the muscles, which provide for adhesion, mucous and other secretions. They are largely carnivorous and prey on invertebrates like protozoans, rotifers, insect larvae, small crustaceans, snails and annelids. Nutrients enter into tissues by diffusion from the central gut. They excrete nitrogen in the form of ammonia that diffuses out of the body and use protonephridia to release water and other waste metabolites as they lack a circulatory system. The Turbellarian nerve cord, subepidermal ring like brain and a nerve net constitute the nervous system. Amongst the Planarian members *Schmidtea mediterranea* (*S. med*) and *Dugesia japonica* (*Dj*) are the most studied. In *Dugesia*, the peripheral nerve net is joined by longitudinal cords at regular intervals by transverse cords, thereby showing a highly organised ladder like orientation with hierarchic order and control. The large fresh water Planarians show asexual reproduction by transverse fission and show regeneration. Of the two types of regeneration observed in the Turbellarians, Architomy and Paratomy, the former includes transverse fission and thereafter generation of new organs, while in the later, the split occurs perpendicular to the antero-posterior (AP) axis and regeneration of new organs start even before complete separation. In the freshwater Planarians, Architomy (Figure 1) is observed and it finds great importance in biology due to the immense regenerative potential of its totipotent adult stem cells called neoblasts. The

Shyamasree Ghosh is the corresponding author.

 ©2017, Shyamasree Ghosh, published by De Gruyter.

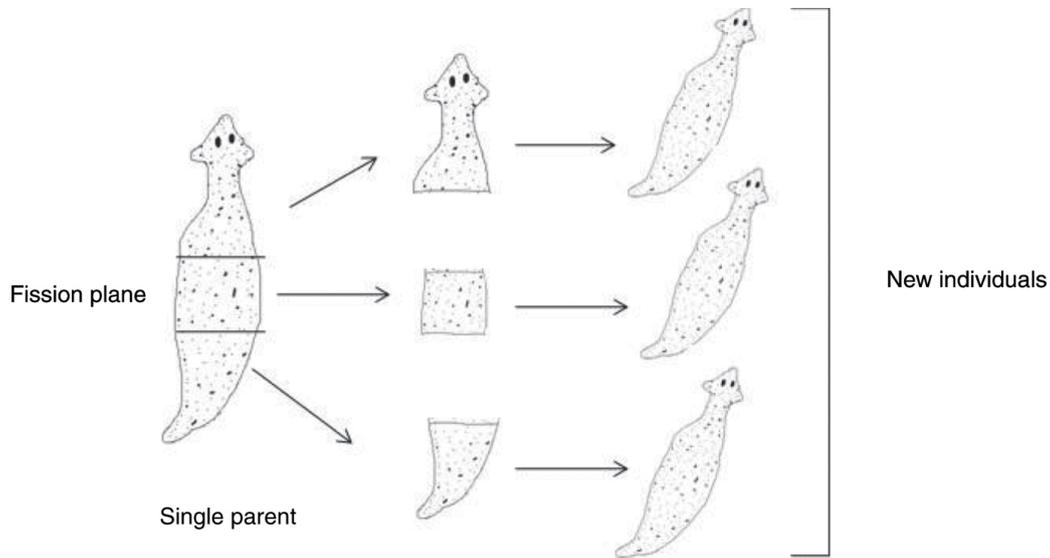
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.

uniqueness of the adult stem cell in *Planaria* arises from its hallmark property to differentiate into different cell lineages [2]. Molecules and pathways in the Planarian regeneration are discussed in Table 2. But the intricate mechanisms controlling the plasticity of the stem cells and homeostasis are not yet known. Although decades of study through morphological and molecular phylogenetic evidences were conducted, yet the evolutionary relationships amongst the members of the phylum were poorly understood. Only recently by the application of high Bayesian posterior probabilities analysis of transcriptomic data, the evolutionary relationship amongst the members of flatworms has been deciphered [3].

**Table 1:** Systematic position of *Planaria* sp., *Schmidtea* sp., *Dugesia* sp.

Kingdom: Animalia			
Order	Tricladida	Tricladida	Tricladida
Family	Planariidae	Dugesiidae	Dugesiidae
Genus	<i>Planaria</i>	<i>Schmidtea</i>	<i>Dugesia</i>

See Ref. [1].



**Figure 1:** Architomy in *Planaria*: a mode of asexual reproduction by which a single individual regenerates into new individuals.

**Table 2:** Molecules and pathways associated with Planarian regeneration.

Molecules/pathway	Planarian physiology	References
The WNT/ $\beta$ -Catenin pathway regulates the Planarian anterioposterior (AP)axis. Bone morphogenetic proteins (BMP)/Decapentaplegic (Dpp) proteins homologues are essential in Dorsoventral (DV) partitioning. WntA is important for the patterning of the posterior brain	Regeneration and homeostasis	[4]
The planarian BMP/anti-dorsalizing morphogenetic protein (ADMP) circuit is regulated by canonical antagonists of the noggin family and novel noggin-like genes (nlg).	DV axis regeneration	[5]
Gap Junction (GJ) proteins are essential regulators of neoblast biology Innexin proteins modulate neoblast responses	Neoblast Biology	[6]
Three major Piwi proteins: SMEDWI-1, 2 and -3 in <i>Schmidtea mediterranea</i> and DJPIWI A, B and C in <i>Dugesia Japonica</i> important in neoblast function		[7]

Hidden Markov Models (HMMs) is a type of dynamic Bayesian network, a statistical model that finds tremendous application in pattern recognition. They find wide scale application in computational biology in

modelling of data from proteins, Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) sequences [8]. Information on the protein family or domain can be inferred at. While profile-HMMs is helpful in analysis and modelling DNA and protein sequences, pair-HMM enables pairwise analysis of DNA and protein sequences but context sensitive HMMs finds application in analysis of RNA sequences [9]. Multiple Sequence Alignment (MSA) is another bioinformatics tool that enables alignment of different protein or nucleic acid sequences similar in length from other members of the group, closely or distantly related species, and infers at the homology and the evolutionary relationships amongst them. Both HMM and MSA have been applied in designing of computational tools, identification of different receptors and other important signalling molecules in *Planaria*. They also find application in designing of computational tools that is helping our understanding of regeneration experiments, morphology and manipulation experiments in *Planaria*. In this article we discuss the different computational approaches, databases, data mining tools, analysis tools for transcriptomic data, gene curation strategies employed in understanding of the biology of the model organism *Planaria* (Table 3).

**Table 3:** Applications of computational methods in *Planaria* research.

Insights into <i>Planaria</i> biology	Computational methods	References
Anatomy, habitat, location, distribution, reproductive strategies for <i>Planaria torva</i> (Müller, 1774), <i>Polycelis tenuis</i> (Ijima, 1884), <i>Dendrocoelum lacteum</i> (Müller, 1774), <i>Polycelis nigra</i> (Müller, 1774), <i>Schmidtea mediterranea</i> S2F2 sexual strain, <i>Schmidtea polychroa</i> and transcript for <i>Schmidtea mediterranea</i> (Benazzi <i>et al.</i> 1975).	Data mining tool PlanMine	[10]
Intermediates in signalling and their possible conserved role in body shape generation	Mining of genome and transcriptome data	[11]
Identification of receptors	Receptor mining and gene curation approaches	[12]
Identification of morphology and regeneration.	Image Plane high-throughput image analysis and information extraction software based on self-parameterizing adaptive thresholding algorithm and automated pipeline	[13]
Regulatory networks	Simulation approach	[14]
Graphical representation of regeneration, manipulation, morphology experiments	Manually curated database and software tool "Planform"	[15], [16]
Information on Planarian genome, including predicted and annotated genes, Expressed sequence tags (ESTs), protein homologies, gene expression patterns and RNA interference (RNAi) phenotypes.	The <i>Schmidtea mediterranea</i> Genome Database (SmedGD) designed to integrate in a single web-accessible portal all available information	[17]

Despite conventional methods very limited information exists globally on a highly complex organism *Planaria*. This study particularly finds importance as many questions still remain unknown in *Planaria* biology. Questions like how and why the system of regeneration, is so unique to *Planaria* amongst its near relatives, the mechanism of plasticity of stem cells, homeostasis of regeneration, unique characteristics of these adult stem cells in *Planaria* are thought provoking.

## 2 Problems in *Planaria* Research that are Not Well Understood

The current day research in *Planaria* revolves around the several questions related to the plasticity of stem cells, the molecular mechanisms, signalling mechanisms involved in stem cells and their regulation and homeostasis events, designing networks and generating models to understand regeneration experiments. Despite studies with morphological and molecular phylogenetic approaches, the diversity within the Planarians and the members of Platyhelminthes phylum are poorly understood. Besides the immune system vital to *Planaria* and its responses to repair after injury is less understood. Although modern day transcriptomic studies have

generated huge data, problems related to their accessibility and appropriate analysis calls in for designing of suitable tools that will help in our understanding of the biology of Planaria.

### 3 Data Mining Tools and Computational Approaches and their Current Applications in Planaria Research

Although only a few computational strategies and data mining approaches are currently applied to research in Planaria, they have been instrumental in addressing some of the problems in Planaria biology. Text mining tools InterMine and PlanMine designed with automated pipelines, are enabling effective easy access, mining and analysis of Planarian transcriptome sequence data and functions to understand their diversity and interspecies relationships [10]. PlanMine is an integrated query-able database which is used by entering identifiers or keywords for contigs, to search databases like InterPro, SMART, PROSITE, Treefam and gene ontology [10]. InterPro enables protein sequence analysis, predicting domain and categorising into families and predicting domains. Pfam enables grouping of protein families, by applying multiple sequence alignment (MSA) and Hidden Markov Models (HMMs). Simple Modular Architecture Research Tool (SMART) database enables protein domain analysis. PROSITE is the biological database for proteins, domain, families. TreeFam is a gene family database and gene ontology (GO) is a database to predict ontology.

PlanMine uses automated pipelines integrating basic local alignment search tool (BLAST) for homology search utilising heuristic algorithm for comparing nucleotide sequences, BLASTX to search protein database NCBI RefSeq using translated nucleotide query, InterProScan suite to search for protein domains and EMBOSS getorf tool for open reading frames (ORF). It uses Reciprocal BLASTP to predict orthologues i.e. genes in different species with common ancestor [10]. Information on the anatomy, habitat, location, distribution and reproductive strategies for the Planarians including *Planaria torva* (Müller, 1774), *Polycelis tenuis* (Ijima, 1884), *Dendrocoelum lacteum* (Müller, 1774), *Polycelis nigra* (Müller, 1774), *Schmidtea mediterranea* S2F2 sexual strain, *Schmidtea polychroa* and transcript information of *Schmidtea mediterranea* can be mined from this database [10].

Genome and transcriptome data mining approaches have enabled understanding of the molecules involved in maintaining left-right asymmetry, namely signalling molecules, tumor growth factor- $\beta$  (TGF- $\beta$ ), Nodal and its downstream effector Paired-like homeodomain transcription factor 2 (also known as pituitary homeobox 2, Pitx) are conserved across all organisms that show spiral cleavage at the third division, grouped under Spiralia including the Planarians [11].

Despite studies on the immune responses and injury induced repair, the immune system in Planarians remains poorly understood. Studies encompassing strategies of comparative genomic analysis and data mining has made it possible to decipher that Planarian immune system houses homologs of the innate immune system common to all Bilateria which are activated as a result of injury thereafter leading to its repair [18]. Receptor mining and gene curation approaches has enabled the identification of the G protein-coupled receptors (GPCRs), comprising families of Rhodopsin, Glutamate, Adhesion, Secretin and Frizzled which play role in eukaryotic signal transduction, in Planarian *Schmidtea mediterranea*. With a transmembrane focussed strategy using Hidden Markov Models (HMMs) it has been possible to predict the occurrence of GPCRs in the proteomes of *S. mansoni* and *S. mediterranea*, extracting information from the GPCR Database (GPCRDB). The mining strategy and formal schemes of analytical processes is discussed in detail [12]. Such strategies has also made possible the identification of novel receptors, including a large and highly-diverged Platyhelminthes specific Rhodopsin subfamily (PROF1), a Planarian-specific Adhesion-like family (PARF1), and a typical Glutamate-like receptor [12].

With the help of ImagePlane high-throughput image analysis and information extraction tool based on novel self-parameterizing adaptive thresholding algorithm and automated pipeline, written in Python, identification of morphology of Planaria through observation of wild phenotype and those with morphological alterations mediated by RNAi induced methods has been possible. We refer the reader to the multiresolution algorithm capable of segmenting the image into anterior-posterior and or left-right quadrants *ab initio* for detailed discussion in the cited article [13].

Perhaps the most important challenge in computational biology is the development of a tool in a quantitative approach to infer the regulatory networks from experimental data on morphology of Planaria. An *in vitro* simulation approach aided by a machine learning module is used to build regulatory networks from experimental data [14]. Virtual worms and simulated experiments are tested by network model to understand pattern formation and algorithm is designed to compare the resultant phenotypes from the simulation studies with real data from planarian database. We refer the readers to the cited article [14] for a detailed discussion on the analytical processes and algorithms. In another study the pattern formation during development is studied [15] from experimental phenotypic morphological data to determine the regulatory networks based

on published results in Planarian regeneration [14], [15]. Approaches for understanding Planarian morphology from regeneration and manipulation experiments is being tested through a manually curated database and software tool termed as “Planform” [16]. PlanformDB provides information on experiments on Planarian morphologies, regeneration and manipulations. It enables a graphical interface with interactive graphs that enable easy to query and user friendly study of Planaria biology [15], [16]. The *Schmidtea mediterranea* Genome Database (SmedGD) is designed by integrating data associated with the Planarian genome, including predicted and annotated genes, expressed sequence tags (ESTs), protein and gene expression patterns and RNA interference (RNAi) phenotypes in a single web portal and is a powerful tool to enable the study of Planaria biology [17]. SmedGD provides data structures compatible with other model organism databases and thus is instrumental in the study of developmental and evolutionary biology, comparative genomics, stem cell biology and regeneration and comparative genomics with other databases like Flybase and Wormbase [17].

The study of regulatory networks in the development and regeneration of Planaria, is a complex dynamic process that is self-regulated and has been modelled for identifying the underlying control mechanisms responsible for the dynamic regulation in regeneration and growth [14].

## 4 Discussion

Computational tools have undoubtedly helped immensely in documenting, analysing and recording of high throughput data in biological sciences. Application of computational tools [2], [3], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18] in the understanding of the Planaria biology is but a very recent approach developed in the last decade (Figure 1). Even at its infancy, the strategies immensely promise at better understanding of Planaria biology analysing the experimental data. The era of transcriptomics has generated huge data in all sphere of biology and also in Planaria biology, however, their analysis through efficient computational tools and resources are being helpful in deciphering the complexity in Planaria. Still there are many questions in Planaria related to its entire network of functioning of the stem cells, immune system and neurobiology which are not yet understood in a holistic approach in Planaria. The challenge for either experimental or computational biology still remains at targeting these vital questions to understand completely the Planaria biology. Scope remains in designing of the biochemical pathways in Planarians to understand the role in regeneration. The future of this article remains at the designing of an integrated approach of experimental and computational resources towards the understanding of complex biology of Planaria. We propose a unique model, to understand the genes and proteins involved in the regeneration pathway comprising the signalling cascades (Table 2) in the Planarian stem cell or neoblast and their structural-functional correlation with biochemical pathways involved and their comparison with the normal somatic cells, through both experimental and computational approaches, so that the information is not only restricted to the genes and proteins involved in the regeneration but also bears direct correlation of the signalling pathways downstream controlling the Planarian regeneration. Such biomolecules recorded from neoblast cells when compared with the somatic cells would help in the understanding of the uniqueness of pathways in Planarian stem cell. We also propose to incorporate the information on homology of such proteins reported from other known organisms so as to decode the uniqueness of Planarian regeneration biology in structure and function in a systems biology approach.

## Acknowledgement

The author acknowledges the School of Biological Sciences, National Institute of Science Education and Research (NISER), Bhubaneswar, Odisha 752050, India; Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India for the support.

**Conflict of interest statement:** Authors state no conflict of interest. All authors have read the journal’s Publication ethics and publication malpractice statement available at the journal’s website and hereby confirm that they comply with all its parts applicable to the present scientific work.

## References

- [1] Ronald S, Masaharu K, Marta R, Jaume B. [A new higher classification of planarian flatworms](#) (Platyhelminthes, Tricladida). *J Nat Hist.* 2009;43:1763–1777.
- [2] Jochen CR. [Stem cell systems and regeneration in planaria](#). *Dev Genes Evol.* 2013;223:67–84.
- [3] Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, et al. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. *Curr Biol.* 2015;25:1347–53.
- [4] Adell T, Cebrià F, Saló E. Gradients in Planarian regeneration and homeostasis. *Cold Spring Harb Perspect Biol.* 2010;2:a000505.
- [5] Molina MD, Saló E, Cebrià F. [Organizing the DV axis during planarian regeneration](#). *Commun Integr Biol.* 2011;4:498–500.
- [6] Peiris TH, Oviedo NJ. Gap junction proteins: master regulators of the planarian stem cell response to tissue maintenance and injury. *Biochim Biophys Acta.* 2013;1828:109–17.
- [7] Resch AM, Palakodeti D. [Small RNA pathways in Schmidtea mediterranea](#). *Int J Dev Biol.* 2012;56:67–74.
- [8] Eddy SR. [What is a hidden Markov model?](#). *Nature Biotechnol.* 2004;22:1315–6.
- [9] Yoon BJ. [Hidden Markov models and their applications in biological sequence analysis](#). *Curr Genom.* 2009;10:402–15.
- [10] Brandl H, Moon H, Vila-Farré M, Liu SY, Henry I, Rink JC. Plan Mine—a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res.* 2016;44:D764–73.
- [11] Grande C, Martín-Durán JM, Kenny NJ, Truchado-García M, Hejnl A. Evolution, divergence and loss of the Nodal signalling pathway: new data and a synthesis across the Bilateria. *Int J Dev Biol.* 2014;58:521–32.
- [12] Zamanian M, Kimber MJ, McVeigh P, Carlson SA, Maule AG, Day TA. The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. *BMC Genom.* 2011;12:596.
- [13] Flygare S, Campbell M, Ross RM, Moore B, Yandell M. [ImagePlane: an automated image analysis pipeline for high-throughput screens using the planarian Schmidtea mediterranea](#). *J Comput Biol.* 2013;20:583–92.
- [14] Lobo D, Levin M. Inferring regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. *PLoS Comput Biol.* 2015;e100429511(6).
- [15] Lobo D, Malone TJ, Levin M. Towards a bioinformatics of patterning: a computational approach to understanding regulative morphogenesis. *Biol Open.* 2013;2:156–69.
- [16] Lobo D, Malone TJ, Levin M. Planform: an application and database of graph-encoded planarian regenerative experiments. *Bioinformatics.* 2013;29:1098–100.
- [17] Robb SM, Ross E, Sánchez Alvarado A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* 2008;36(Database issue):D599–606.
- [18] Peiris TH, Hoyer KK, Oviedo NJ. Innate immune system and tissue regeneration in planarians: an area ripe for exploration. *Semin Immunol.* 2014;26:295–302.