

# Evolutionary dynamics of the *Wnt* gene family: implications for lophotrochozoans\*

LIU Jun (刘俊)<sup>1, 2, 4</sup>, XU Fei (许飞)<sup>1, 3, 4, \*\*</sup>, JI Peng (纪鹏)<sup>1, 2, 4</sup>, LI Li (李莉)<sup>1, 4, 5</sup>,  
ZHANG Guofan (张国范)<sup>1, 3, 4, \*\*</sup>

<sup>1</sup> Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266000, China

<sup>4</sup> National & Local Joint Engineering Laboratory of Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

<sup>5</sup> Laboratory for Marine Fisheries and Aquaculture, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266000, China

Received Apr. 27, 2017; accepted in principle Jun. 15, 2017; accepted for publication Jul. 19, 2017

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2018

**Abstract** Genes encoding Wnt ligands, which have important roles in cell communication and organ development, are restricted to multicellular animals. We systematically studied *Wnt* genes from eumetazoan genomes, with emphasis on the poorly studied superphylum Lophotrochozoa (four annelids, seven mollusks, eight platyhelminths, one bdelloid rotifer, and one brachiopod species). Between 3 and 39 *Wnt* loci were identified in each genome, and the protostome-specific loss of *Wnt3* genes was confirmed. We identified gastropod-specific loss of *Wnt8*, refining the previously proposed mollusk-specific loss. Some duplicated *Wnt* genes belonging to a same subfamily or closely related subfamilies showed tandem distribution in the lophotrochozoan genomes, indicating tandem duplication events during Wnt family evolution. Members of the conserved *Wnt10-Wnt6-Wnt1-Wnt9* cluster showed highly correlated expression patterns over time in two assayed lophotrochozoans, the oyster *Crassostrea gigas* and the brachiopod *Lingula anatina*, reflecting the possible similar function of the clustered *Wnt* genes.

**Keyword:** phylogeny; gene cluster; time-course expression; tissue distribution

## 1 INTRODUCTION

Genes encoding Wnt ligands are restricted to multicellular animals and have important roles during the complicated developmental processes characteristic of these species (Nusse and Varmus, 1992; Logan and Nusse, 2004). Much has been learnt about *Wnt* genes in metazoan model animals (Gordon et al., 2005; Marikawa, 2006; Garriock et al., 2007; Lavery et al., 2008; Simons and Mlodzik, 2008; Fagotto, 2014), particularly in the Deuterostomia and Ecdysozoa, since the first report of *Wnt1* as a protooncogene in mouse in 1982 (Nusse and Varmus, 1982); however, these genes have not been well studied in non-model animals, and functional studies

in the Lophotrochozoa, the third main clade of Bilateria, are limited. In addition to some early reported genomes and Wnt complements, including that of the sponge *Amphimedon queenslandica* (Porifera) (Adamska et al., 2010; Srivastava et al., 2010), the cnidarian, *Nematostella vectensis* (Radiata)

\* Supported by the National Natural Science Foundation of China (Nos. 31402285, 31530079), the Scientific and Technological Innovation Project financially supported by the Qingdao National Laboratory for Marine Science and Technology (No. 2015ASKJ02), and the earmarked fund for Modern Agro-Industry Technology Research System (No. CARS-48)

\*\* Corresponding authors: xufei@qdio.ac.cn; gfzhang@qdio.ac.cn  
LIU Jun and XU Fei contributed equally to this work.

(Kusserow et al., 2005; Putnam et al., 2007), the echinoderm, *Strongylocentrotus purpuratus* (Croce et al., 2006; Sodergren et al., 2006), the chordate, *Danio rerio* (Garriock et al., 2007; Howe et al., 2013), the nematode, *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998; Korswagen, 2002), and the arthropods *Drosophila melanogaster* (Adams et al., 2000; Croce et al., 2006) and *Daphnia pulex* (Janssen et al., 2010; Colbourne et al., 2011), the recently published lophotrochozoan *Wnt* complements (including those of the Platyhelminthes species, *Schistosoma mansoni* (Berriman et al., 2009; Riddiford and Olson, 2011), *Schmidtea mediterranea* (Riddiford and Olson, 2011; Robb et al., 2015), *Echinococcus granulosus*, *Echinococcus multilocularis*, and *Hymenolepis microstoma* (Riddiford and Olson, 2011; Tsai et al., 2013); the annelids, *Platynereis dumerilii* (Janssen et al., 2010), *Capitella teleta*, and *Helobdella robusta*; and the mollusc, *Lottia gigantea* (Cho et al., 2010; Simakov et al., 2013) have greatly improved our understanding of *Wnt* gene evolutionary dynamics across the Eumetazoa. The rapid growth in the generation and analysis of genome data in recent years presents an opportunity to investigate the evolution of *Wnt* genes by thorough analyses of their distribution and function across a broad range of lophotrochozoans. These include species for which *Wnt* complement studies are lacking or only part of the *Wnt* genes has been identified: the platyhelminths *Taenia solium* (Tsai et al., 2013), *Schistosoma japonicum* (Zhou et al., 2009), and *Schistosoma haematobium* (Young et al., 2012); the bdelloid rotifer, *Adineta vaga* (Flot et al., 2013); the brachiopod, *Lingula anatina* (Luo et al., 2015); the annelid *Spirobranchus (Pomatoceros) lamarcki* (Kenny et al., 2015); and the quickly accumulating genomes of molluscs including *Crassostrea gigas* (Zhang et al., 2012), *Pinctada fucata* (Takeuchi et al., 2016), *Patella vulgata* (Kenny et al., 2015), *Octopus bimaculoides* (Albertin et al., 2015), *Biomphalaria glabrata*, and *Aplysia californica*.

Here, we report the identification of *Wnt* genes based on genome data from 20 animals in the superphylum Lophotrochozoa, and we describe the evolutionary dynamics of *Wnt* gene families through comparisons among closely related species. We characterized the functional correlation of *Wnt* genes in Lophotrochozoans by comparing their time-course expression patterns. Finally, we verified our hypotheses by studying the dynamic expression patterns of *Wnt* genes in the Pacific oyster, *C. gigas*.

## 2 MATERIAL AND METHOD

### 2.1 Animal materials and ethics statement

All of the Pacific oyster (*C. gigas*) specimens used in this study were collected from Qingdao, Shandong, China, and acclimatized in seawater at 22°C for 7 days before use. No specific permissions were required for any of the experimental processes used, and all experiments were conducted according to local and national regulations. Fresh oyster tissues for RNA isolation were isolated, those from two animals mixed, and then frozen immediately and stored in liquid nitrogen. Three biological replicates were conducted for RNA isolation, i.e. a total of six animals were used in the study.

### 2.2 Collection of genome assemblies and annotations

All genome assemblies and annotations were collected from the NCBI Eukaryotic Genome Annotation project (Pruitt et al., 2012) (accession numbers are listed in Table S1), except for the *P. fucata* genome from the genome database ([http://marinegenomics.oist.jp/pearl/viewer/download?project\\_id=36](http://marinegenomics.oist.jp/pearl/viewer/download?project_id=36)), *P. vulgata* and *S. (Pomatoceros) lamarcki* genomes from (<https://ora.ox.ac.uk/objects/uuid:6471e7d4-dd34-4eb3-883f-a5c438f11741>). Additionally, we included *Wnt* genes from the annelid *P. dumerilii*, whose *Wnt* complement has been well studied (Pruitt et al., 2014), although no genome data are publically available. For species for which genomic-wide protein sets have been deposited in NCBI, the whole protein sets were downloaded in GenPept format. *Wnt* protein and domain sequences were retrieved by searching the protein database with conserved *Wnt* family domain accessions (CDD numbers 278536, 302926, and 128408). To confirm the completeness of *Wnt* complement or to identify all possible *Wnt* gene loci in the genome assembly, additional similarity searches against genomes were performed using TBLASTN with the domain sequences of all known *Wnt* proteins as queries. *Wnt* loci for species with un-annotated genomes were also identified with TBLASTN. The open reading frame (ORF) around a hit region was predicted with the NCBI ORFfinder service, and identified *Wnt* domains were assessed against the PFAM database (<http://pfam.sanger.ac.uk/search>) with an e-value threshold of 1e-10.

### 2.3 Sequence alignment and phylogenetic analyses

A multiple alignment of the *Wnt* domain sequences was produced with MAFFT 7.221 (Katoh and Standley,

2013) using the L-INS-I algorithm. The alignment was trimmed with TrimAl, with parameters -gt 0.9 -st 0.001 -cons 40 (Capella-Gutiérrez et al., 2009). Then, molecular phylogenetic trees were constructed to support *Wnt* gene classification using three different phylogenetic methods: FastTree (Price et al., 2010), RAxML (Stamatakis, 2014), and MrBayes (Huelsenbeck and Ronquist, 2001). FastTree allows constructing approximately-maximum-likelihood phylogenetic trees quickly. Therefore, it was used to test alignment-trimming results and to preliminarily name *Wnt* genes. The resulting alignment belonging to the Wnt domain was used to perform maximum-likelihood (ML) phylogenetic analyses with the program RAxML, using the evolutionary model LG+Gamma+Invariant; 1 000 replicates were performed to obtain bootstrap support (BS) values to evaluate the nodal support. To confirm and compare the nodal support, Bayesian inference (BI) phylogenetic analysis was conducted, using MrBayes. Eight Markov chains were run for  $1 \times 10^6$  generations, with sampling performed every 100 generations. We used the values, as well as congruence of statistical support (BS from RAxML and posterior marginal probabilities from MrBayes) as indicators of the reliability of different subfamilies. Two trees were constructed: a large tree containing all 360 *Wnt* genes identified from the 28 species, and a small tree containing well supported groups of *Wnt* genes from well annotated genomes (from 16 species, Table S1). Some known divergent *Wnt* genes (discussed later) from the chosen genomes were excluded during construction of the small tree. To further classify the divergent protein sequences mentioned above, e.g. Cel\_wnt9\_NP\_505154, Aqu\_wnt16b\_NP\_001266204, Aqu\_wnt11\_XP\_003388134, we defined a “standard” *Wnt* gene set by extracting representatives of Radiata, Lophotrochozoa, Ecdysozoa, and Deuterostomia from each well-supported *Wnt* subfamily, and then constructed a phylogenetic tree by adding them one by one (Fig.S1).

## 2.4 Criteria for inference of evolutionary relationships

Orthology groups were defined from phylogenetic trees following the criteria adopted by other similar analyses (Gyoja, 2014) with some modifications. In brief, if genes from two or more organisms formed a single clade with a bootstrap value >50% by ML or a posterior probability >90% by BI, they were considered to constitute an orthologous subfamily most solidly. Members within the subfamily were then named according to the subfamily name. Where two or more

members of a subfamily were identified in one species, they were named “a” (or “aa”, “ab”), “b,” “c,” etc. For example, the two *D. rerio* members from the *Wnt6* subfamily are named *Wnt6a* and *Wnt6b*, while the four *Wnt7* subfamily members are named *Wnt7aa*, *Wnt7ab*, *Wnt7ba*, and *Wnt7bb*, to reflect the degree of relatedness (e.g., *Wnt7aa* and *Wnt7ab* are closer paralogs pair). Gene losses are revealed where a lineage is expected to have a gene based on the inferred ancestral gene subfamily complement, but the gene is not detected. However, this can also reflect missing data, and the genomes analyzed here vary in assembly or annotation quality. Accordingly, gene loss was modeled using Dollo parsimony, which allows multiple, independent losses (Riddiford and Olson, 2011), and we only inferred a loss where a gene was absent from two or more sister lineages.

## 2.5 Time-course microarray or RNAseq data collection and analysis

Time-course expression patterns of *Wnt* genes were determined for six species, as follows: microarray data from 61 developmental stages of *D. rerio* were derived from a previous report (Domazet-Lošo and Tautz, 2010); FPKM (fragments per kilobase of transcript per million mapped reads) values for *D. melanogaster* and *C. elegans* *Wnt* genes were retrieved from the modENCODE project (Li et al., 2014); and RNAseq data for the different developmental stages of three other species (*N. vectensis*, *L. anatina*, and *C. gigas*) were downloaded from the NCBI SRA database and analyzed with Hisat2, StringTie, and Ballgown (Pertea et al., 2016), using NCBI annotation information. Details of these data are provided in Table S2. Correlation coefficient for the Wnt expression matrix was calculated by correlation function of the R software (Ihaka and Gentleman, 1996). The probability values of individual correlations were determined by t-test followed by Holm’s adjustment for multiple comparisons.

## 2.6 RNA isolation, cDNA synthesis, and quantitative reverse transcription (qRT-)PCR

RNA isolation and cDNA synthesis from different oyster tissues were conducted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and a PrimeScript RT reagent kit (TaKaRa, Shiga, Japan), respectively, according to the manufacturers’ instructions. qRT-PCR of the *C. gigas* *Wnt* genes was conducted as previously described, with three technical replicates in each of three biological replicates (Qu et al., 2014).

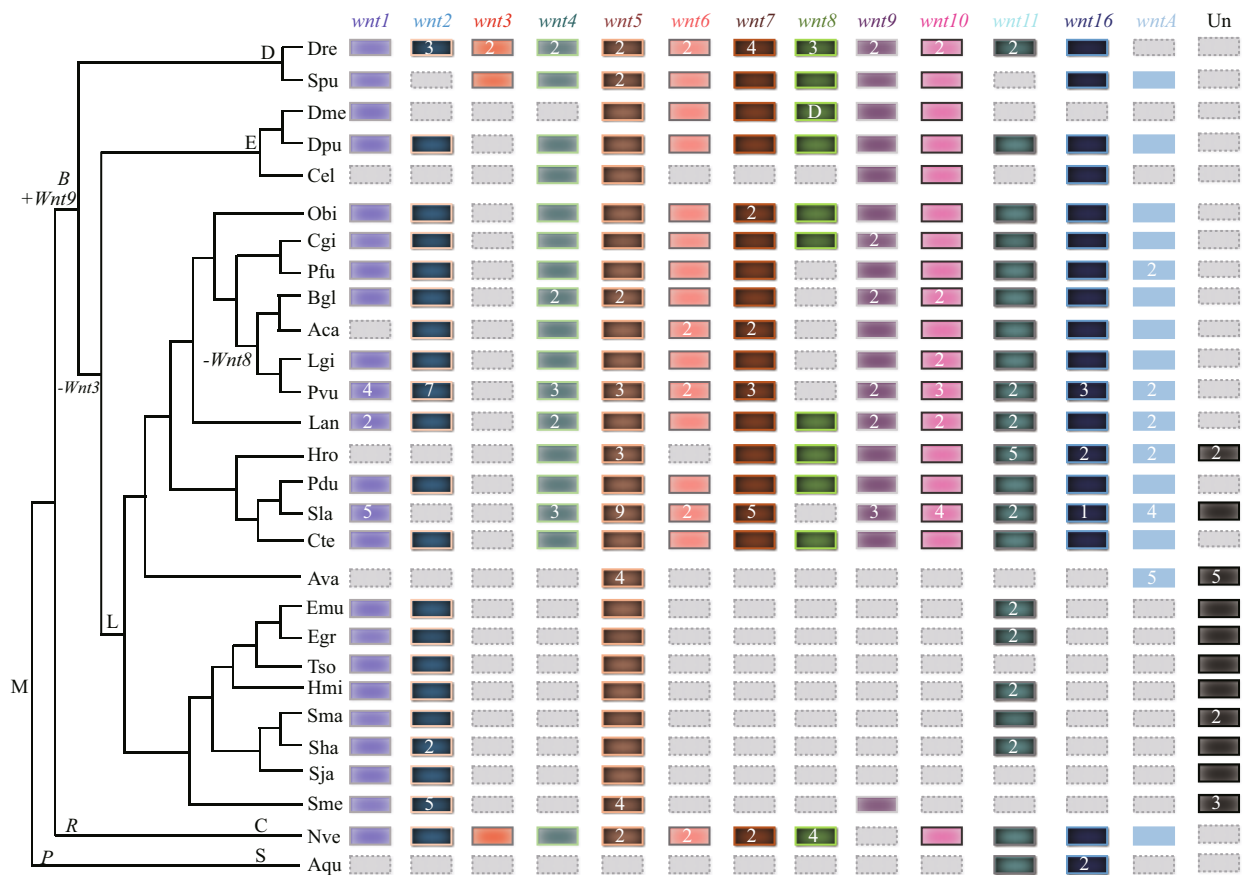


Fig.1 *Wnt* genes in typical Eumetazoan phyla

The *Wnt* subfamilies (*Wnt1–11*, *Wnt16*, and *WntA*) are represented by colored boxes, and unclassified genes (Un) are shown in black. Each box stands for a *Wnt* gene. Grey boxes with dashed outlines indicate the possible loss of particular *Wnt* subfamily and the numbers in the box indicate the number of paralogs in each species. The origin of subfamilies is indicated by a “+” sign below the respective branch, while loss is with a “–” sign. The phylogenetic relationship between the selected taxa based on NCBI Taxonomy and previous reports (Croce et al., 2006, Riddiford and Olson, 2011) is provided on the left. Species included *Amphimedon queenslandica* (Aqu) [Porifera (P), Spongia (S)], *Nematostella vectensis* (Nve) [Radiata (R), Cnidaria (C)]; *Strongylocentrotus purpuratus* (Spu) and *Danio rerio* (Dre) Deuterostomia (D)); *Caenorhabditis elegans* (Cel), *Drosophila melanogaster* (Dme) and *Daphnia pulex* (Dpu) [Ecdysozoa (E)]; *Schmidtea mediterranea* (Sme), *Schistosoma japonicum* (Sja), *Schistosoma haematobium* (Sha), *Schistosoma mansoni* (Sma), *Hymenolepis microstoma* (Hmi), *Taenia solium* (Tso), *Echinococcus granulosus* (Egr), *Echinococcus multilocularis* (Emu), *Adineta vaga* (Ava), *Capitella teleta* (Cte), *Spirobranchus (Pomatoceros) lamarki* (Sla), *Platynereis dumerilii* (Pdu), *Helobdella robusta* (Hro), *Lingula anatina* (Lan), *Patella vulgata* (Pvu), *Lottia gigantea* (Lgi), *Aplysia californica* (Aca), *Biomphalaria glabrata* (Bgl), *Pinctada fucata* (Pfu), *Crassostrea gigas* (Cgi), and *Octopus bimaculoides* (Obi) [Lophotrochozoa (L)].

The *EF-1α* gene was used as an internal control (Du et al., 2013) and the  $2^{-\Delta\Delta C_t}$  method was used to calculate the expression level of target genes (Livak and Schmittgen, 2001). All primers used for qRT-PCR are listed in Table S3.

3 RESULT AND DISCUSSION

3.1 Identification of *Wnt* genes from selected Eumetazoa species

We screened 20 lophotrochozoan genomes, 12 of which had never been assayed for *Wnt* complement at the genome-wide level, together with the latest assemblies of seven other metazoan genomes. All thirteen known *Wnt* gene subfamilies, i.e., *WntA*, *Wnt1–11*, and *Wnt16*, were identified (Fig.1). As

previously reported, some flatworm *Wnt* proteins are highly divergent (Riddiford and Olson, 2011) and thus, failed to be grouped within the thirteen known subfamilies. Similar results have been reported for *H. robusta* (Cho et al., 2010), in which some *Wnt* genes show greater structural variability. For example, the *H. robusta Wnt2* has only 13 cysteines and *Wnt7* has 27, while most *Wnt* proteins have normal cysteine counts between 23–24 (Cho et al., 2010). Three *Amphimedon Wnt* genes (Srivastava et al., 2010; Holstein, 2012) as well as *C. elegans mom-2* and *egl-20* (Prud'homme et al., 2002; Janssen et al., 2010) have also proven difficult to classify.

All assayed flatworm *Wnt* genes clustered into five groups with strong support (Figs.S2, S3). One group was within the previously defined subfamily, *Wnt5*



(Riddiford and Olson, 2011). Another group was in subfamily *Wnt11*, with strong support. Among the other three groups, they also showed close relationships (with strong support) with the neighboring previously defined subfamilies, we are careful to name them after the corresponding subfamily because these groups were isolated from the known subfamilies, whose members are broadly from Cnidaria and Bilateria. At the same time, our results for these three groups seemed to conflict with a previous classification (Riddiford and Olson, 2011). As previous studies have made good classification on two of these groups through expression patterns and other characters (Gurley et al., 2010; Koziol et al., 2016), we thus kept the current classification. We named these groups “*Wnt1*,” “*Wnt2*,” and “*Unclassified*” accordingly. The *Wnt1* group of flatworm is a sister of the *Wnt10* subfamily in our tree (BS support 29%, Fig.S2, and BI support 0.44, Fig.S3); the *Wnt2* is a separate group in both ML and Bayesian trees; the *Unclassified* group contains members of the annelid *H. robusta*, together with genes from the bdelloid rotifer *A. vaga*. As in the previous report, *Wnt4* is a closely related subfamily (BS support 25%, Fig.S2, and BI support 0.64, Fig.S3). The flatworm *Wnt* genes indicate the problem of family classification through phylogeny analysis with the sequences, especially for highly divergent genes. More evolutionary and functional details are needed before correctly classifying such genes. The three sponge *Wnt* genes have been previously annotated; one was suggested to belong to the *Wnt11* subfamily, another to *Wnt16*, and one remained unclassified (Adamska et al., 2010). Our results based on the phylogenetic tree to which these genes were added one by one support the classification of one *Wnt11* (Fig.S1) and two *Wnt16* (Fig.1 and Fig.S1) genes. One-by-one phylogenetic analysis of *C. elegans mom-2* classified it into the subfamilies *Wnt9*, as previously reported (Prud'homme et al., 2002).

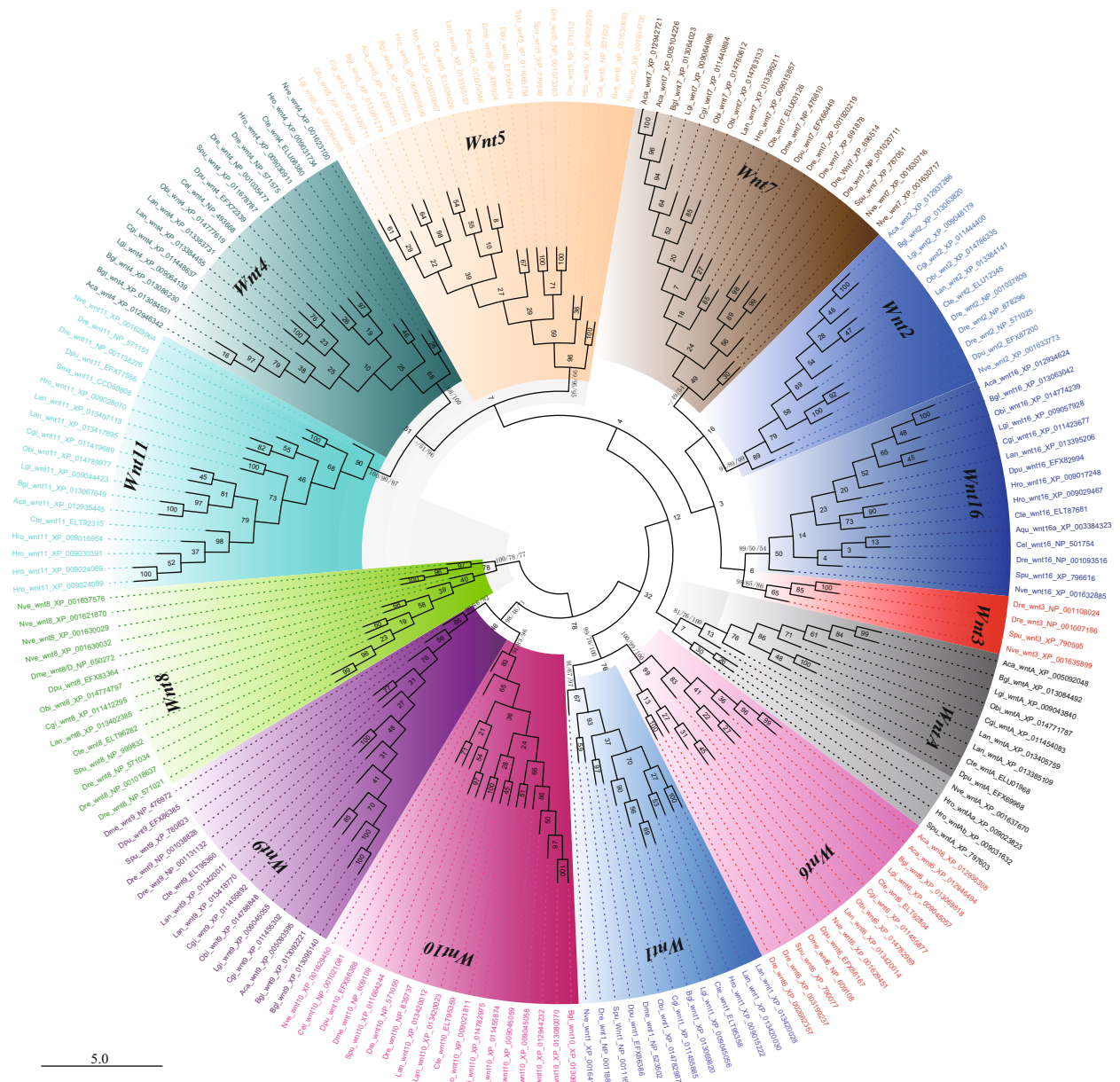
The other *Wnt* genes were classified into known subfamilies with high support. All reported *Wnt* genes were identified in the genome of each species, except for *D. rerio*, for which 5 *Wnt2* and 3 *Wnt7* genes were previously reported (Garriock et al., 2007), while we identified 3 and 4, respectively (Table S4). Our gene counts correspond with the genome annotation data in the Zebrafish Model Organism Database (Howe et al., 2013). The previous report may have assayed only genes with a full *Wnt* domain, while we used all possible *Wnt* loci identified across the whole genome.

Most of the gene products from the additional loci are characterized by a partial *Wnt* domain, which may be because of genome misassemblies or pseudogenes. In *N. vectensis*, *S. purpuratus*, *L. gigantea* and *H. robusta*, we identified four, one, one, and four additional possible *Wnt* loci with partial *Wnt* domain, respectively. *Wnt* counts for other, unreported species were between 4 and 39 (Fig.1, Tables S1, S4). The genome sequences of the mollusc *P. vulgata* and the annelid *S. lamarcki* were of poor quality. There were only contigs, with N50 only 3 160 bp and 1 939 bp, respectively. We identified 34 and 39 possible *Wnt* loci in the genome assemblies of *P. vulgata* and *S. lamarcki*, respectively, which are obviously abnormal numbers when compared to other relative species. However, we included these genes in the analysis as they could provide information for verifying our gene loss hypothesis in certain lineages.

### 3.2 Phylogenetic analyses and repertoire of *Wnt* genes

We adopted a commonly used approach for constructing *Wnt* gene phylogenetic trees based on the well-annotated gene sequences of vertebrate (*D. rerio*), and ecdysozoan (*D. melanogaster* and *C. elegans*) species. As no outgroup can be used for *Wnt* genes (Schubert et al., 2000), an unrooted phylogenetic tree was produced. Based on the tree generated, *Wnt* genes were subdivided into orthologous *Wnt* subfamilies and named according to the orthologous subfamily name (Fig.2). As FastTree uses only the approximately-maximum-likelihood method to construct a tree within seconds, in this study, we mainly used RAxML and MrBayes, the topologies of which were largely the same (Figs.2, S2, S3).

As previously reported, *Wnt3* was identified only in species classified as Cnidaria and Deuterostomia (Prud'homme et al., 2002; Janssen et al., 2010; Riddiford and Olson, 2011), and *Wnt9* was only found in Bilateria (Croce and McClay, 2008; Lengfeld et al., 2009; Cho et al., 2010). It has been speculated that *Wnt8* has undergone lineage-specific loss in Mollusca (Cho et al., 2010; Riddiford and Olson, 2011); however, we identified *Wnt8* in two mollusks, *C. gigas* (Bivalvia, Mollusca) and *O. bimaculoides* (Cephalopoda, Mollusca), suggesting that the loss may have been specific to Gastropod mollusks. Previous reports used only the gastropods *P. vulgata* and *L. gigantea* as representatives of mollusks; thus, the conclusion that mollusks are characterized by a loss of *Wnt8* (Riddiford and Olson, 2011) on the basis



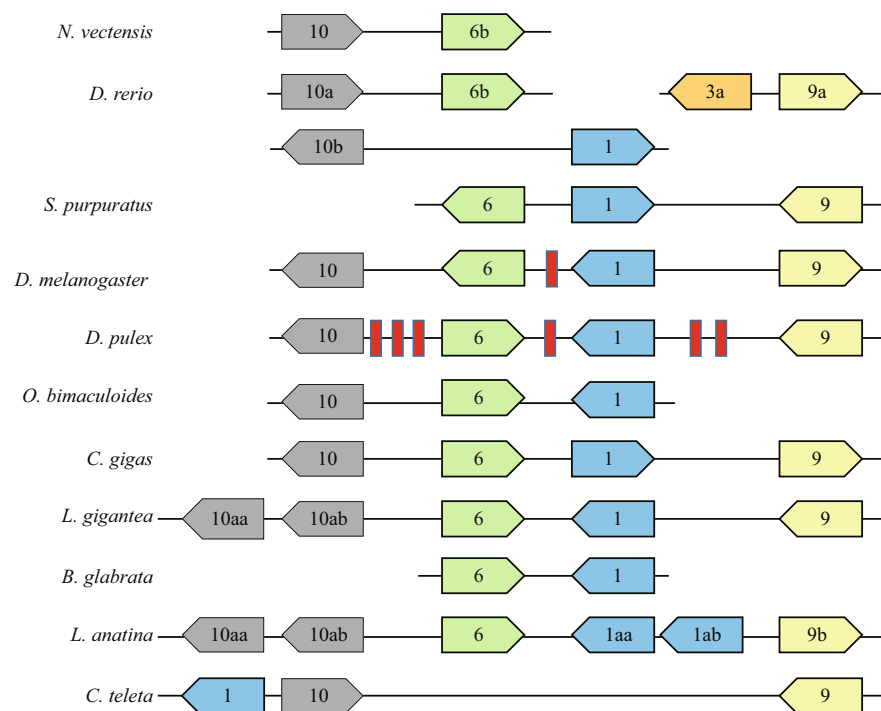
**Fig.2 Molecular phylogenetic analysis of lophotrochozoan *Wnt* genes**

RAxML tree constructed with a total of 193 *Wnt* genes from 16 complete Eumetazoan genomes (including 9 lophotrochozoans). Subfamilies are shown with different colors, and the support values from FastTree, RAxML, and MrBayes analyses (e.g. 100/90/97 for Wnt11) are shown at the basal node. FastTree analysis did not support the subfamilies *Wnt4* and *Wnt7*; therefore, the values are indicated as “–”. *WntA* is marked by dark gray and light gray to distinguish well supported groups and diverged members within a single clade. Species abbreviations are as defined in Fig.1.

of data lacking other classes of Mollusca has been refined by our study. Similarly, although we have shown the loss of *Wnt8* in four species from both the Patellogastropoda (*P. vulgata* and *L. gigantea*) and Heterobranchia (*A. californica* and *B. glabrata*) clades of Gastropoda, our claim requires further validation in other mollusks.

Based on genome analysis of the ancient sponges, *Wnt* genes appeared in the metazoan last common ancestor, followed by gene duplication in the period between demosponge-eumetazoan and cnidarian-

bilaterian last common ancestors (Adamska et al., 2010). However, the evolutionary relationships between subfamilies are difficult to infer through phylogenetic analysis because of insufficient resolution (Prud'homme et al., 2002). For example, the BS support for most sister groups is lower than 20% (Fig.2). Therefore, it is difficult to infer evolutionary trajectories since the origin of *Wnt* genes, although studies have suggested that the three *Amphimedon* *Wnt* genes belong to Wnt11 and Wnt16 subfamilies. Meanwhile, some sister groups showed



**Fig.3 Schematic representation of *Wnt10-Wnt6-Wnt1-Wnt9* clusters in different animals**

Arrows indicate the transcriptional orientations. Red bars indicate non-*Wnt* genes located among the cluster. Other *Wnt* linkages are described in Fig.S4 (based on Cho et al., 2010; Kusserow et al., 2005).

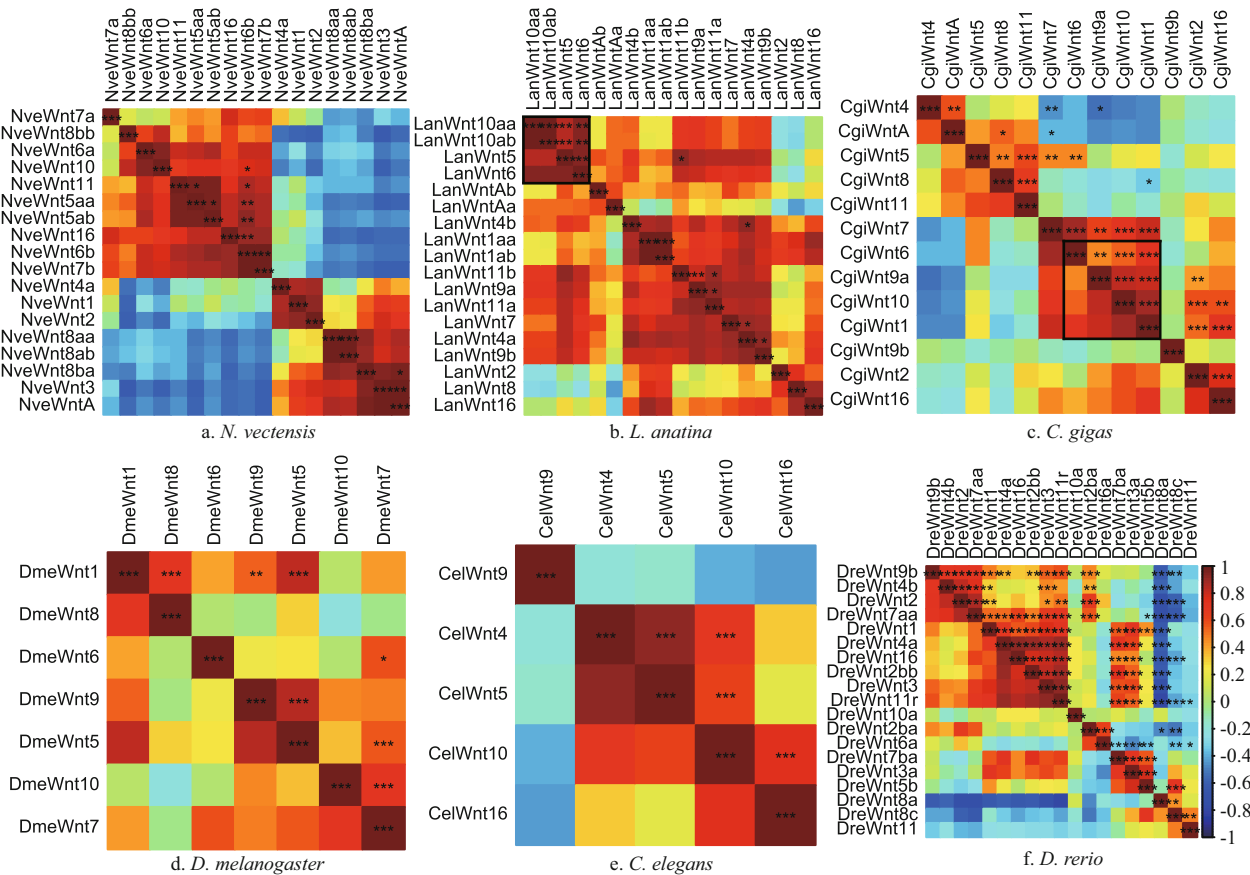
relatively strong BS support, including *Wnt4-Wnt11* (51%), *Wnt1-Wnt6* (76%), and *Wnt9-Wnt10* (46%), in accordance with a previous report (Cho et al., 2010). These sister subfamilies have been hypothesized to have arisen from specific gene duplications (Prud'homme et al., 2002; Cho et al., 2010). Linked genes are considered as evidence of ancestral arrangements when they are shared by more than one lineage. Moreover, the closer genes are located to each other in a genome, the more likely they are a pair evolved by tandem duplication. In this context, the *Wnt10-Wnt6-Wnt1-Wnt9* cluster, which widely exists in Metazoan genomes (Kusserow et al., 2005; Cho et al., 2010), should represent a *Wnt* tandem duplication from an ancestral gene. Such tandem duplication within a single *Wnt* subfamily has been reported in zebrafish *Wnt8* (Ramel et al., 2004). In this study, we identified as many as 8 potential *Wnt8* loci tandemly distributed in the scaffold NW\_001834285.1 of *N. vectensis*. However, the ORFs of six loci are interrupted by a stop codon and produce only a partial Wnt domain, suggesting these may be misassemblies or pseudogenes. Possible tandem-duplicated *Wnt* loci have been also identified in *Wnt7* of *N. vectensis*, *Wnt10* of *L. gigantea*, *Wnt1* and *Wnt10* of *L. anatina*, *Wnt13* of *S. mediterranea* and *S. haematobium* (Fig.S4).

### 3.3 Expression patterns of clustered *Wnt* genes

To investigate the functions of *Wnt* genes further, we explored their expression patterns by analyzing published expression data from six animals (*N. vectensis*, *D. rerio*, *D. melanogaster*, *C. elegans*, *L. anatina*, and *C. gigas*) during their ontogeny. The results indicated that few *Wnt* genes are expressed maternally (Table S2). In contrast, expression of the majority of *Wnt* genes began during embryogenesis, corresponding with their primary function in the control of cell-cell interactions during development and adult tissue homeostasis. Additionally, we calculated the correlation between expression patterns of different *Wnt* genes in six species (Fig.4). In this way, gene functional relationships could be inferred from the time-course expression patterns.

In oyster, the genes in the *Wnt10-Wnt6-Wnt1-Wnt9* cluster showed an obviously consistent expression pattern (Fig.4c), indicating the possible interaction between these adjacent genes. Although clusters in *Lingula* are more diverse because of the duplication of *Wnt10* and *Wnt1* (Fig.3), a high correlation was observed between *Wnt10* and *Wnt6* (Fig.4b).

Expression correlation was also indicated from the tissue expression pattern of oyster *Wnt* genes. We assayed the expression levels of the 13 oyster *Wnt*



**Fig.4** Correlation analysis of the time-course expression patterns of *Wnt* genes in six selected species

The Pearson correlation coefficient was calculated for expression levels of *Wnt* genes among species. Dark red blocks show high correlation between the gene pairs. Black boxes indicate the high correlation between clustered *Wnt* genes. Detailed expression data are listed in Table S2. \*\*\* indicates significance at adjusted  $P < 0.001$ , \*\* at adjusted  $P < 0.01$ , and \* at adjusted  $P < 0.05$ .

genes through real-time qPCR on seven main tissues. The results indicated that the linked *Wnt10*, *Wnt6*, *Wnt1*, *Wnt9a* all had high level expression gill and mantle (Fig.5). However, differences were also observed between these four genes. For example, *Wnt1* showed specific expression in gill and mantle, while *Wnt6* also expressed in gonad. Additionally, *Wnt9a* also showed high expressed in transparent adductor muscle and gonad respectively. In this context, the clustered *Wnt* genes may be coordinately regulated during certain development stages or in some tissues, but also have diverged biological functions in other biological processes.

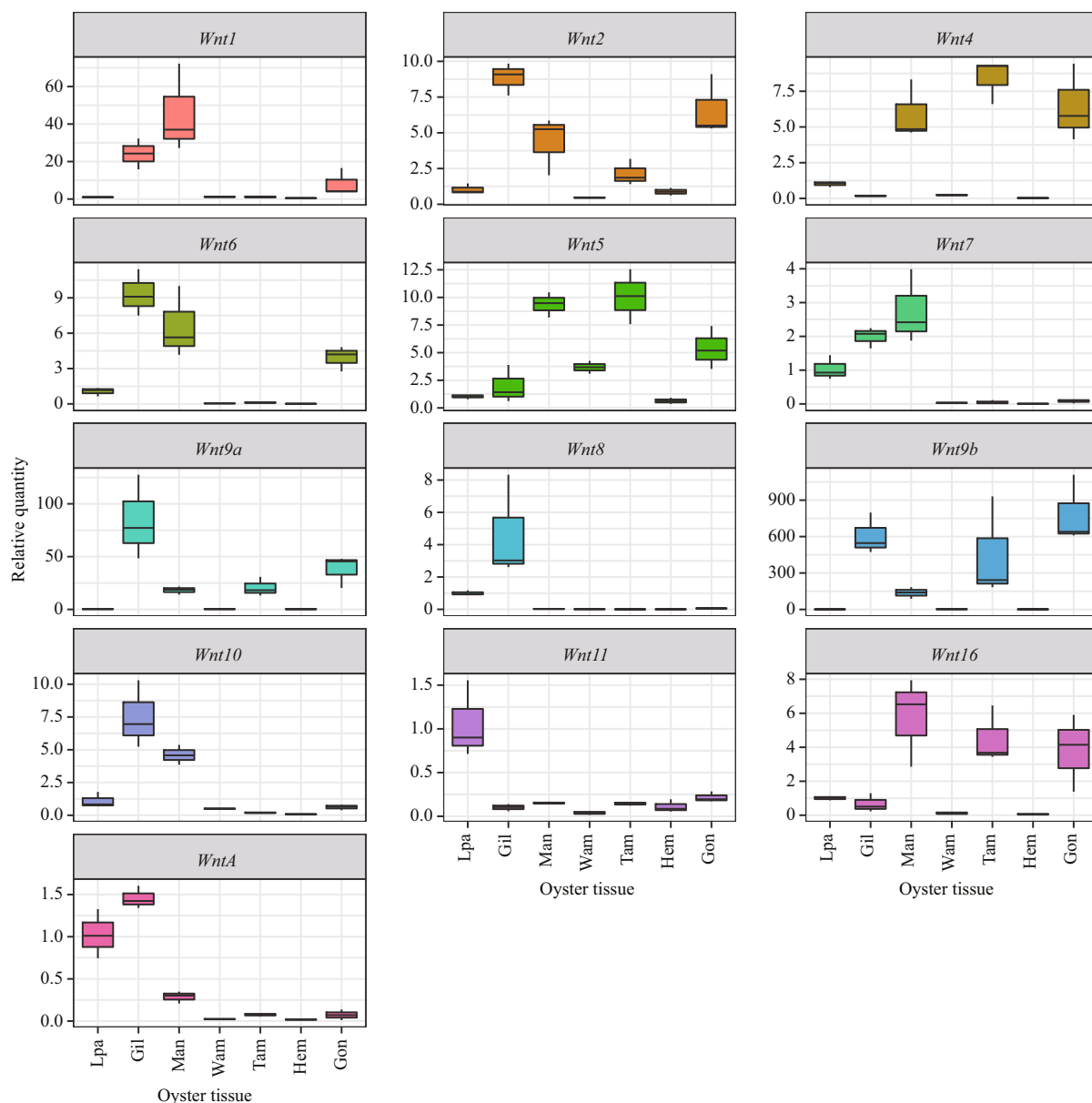
At the same time, no strong correlation between clustered genes was detected in fruit fly, *C. elegans*, and zebrafish (Fig.4d, e, f). If the coexpression of *Wnt10-Wnt6-Wnt1-Wnt9* genes in lophotrochozoans reflects their functional correlation and the members indeed arose by tandem duplication, the diverged expression pattern in other animals should reflect subfunctionalization of the duplicated genes as predicted by the duplication degeneration

complementation model (Force et al., 1999) or neofunctionalization.

#### 4 CONCLUSION

Members of the *Wnt* gene family are widely distributed in the Eumetazoa. *Wnt* genes arose rapidly during the early evolution of animals. Loss of members of some *Wnt* subfamilies occurred in certain clades later in evolution, including *Wnt3* from the Ecdysozoa and Lophotrochozoa, *Wnt8* from the Gastropoda, and the simultaneous loss of *Wnt2* and *Wnt11* from *D. melanogaster*, *C. elegans*, and *S. purpuratus*. A number of genes exhibited elevated diversification relative to the main subfamilies, and were therefore assigned to the temporary groups, Unclassified. Other than these highly dynamic subfamilies, *Wnt* genes tended to appear in clusters, suggesting the occurrence of multiple tandem duplication events during *Wnt* family evolution. A number of *Wnt* orthologs exhibited similar time-course expression patterns in different animals, indicating their possible similar functions. The





**Fig.5 Expression patterns of *Wnt* genes in oyster tissues**

The x-axis shows the 7 oyster tissues and y-axis indicates the relative expression levels compared to a reference sample (Lpa). Tissues include: Lpa: labial palps; Gil: gill; Man: mantle; Wam: white adductor muscle; Tam: transparent adductor muscle; Hem: hemolymph; Gon: gonad. The *EF-1a* gene was used as an internal control. Triplicate independent experiments were carried with three technical replicates.

correlation of expression for genes of the *Wnt10-Wnt6-Wnt1-Wnt9* cluster was identified in oyster, suggesting functional conservation of these genes.

## 5 DATA AVAILABILITY

The authors declare that all data supporting the findings of this study are available within the article and its supplementary files.

## 6 ACKNOWLEDGEMENT

We thank the High Performance Computing Center

(HPCC) at the Institute of Oceanology, Chinese Academy of Sciences for providing the supercomputer cluster to conduct gene expression analyses.

## References

- Adams M D, Celniker S E, Holt R A et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science*, **287**(5461): 2 185-2 195.
- Adamska M, Larroux C, Adamski M et al. 2010. Structure and expression of conserved Wnt pathway components in the demosponge *Amphimedon queenslandica*. *Evolution & Development*, **12**(5): 494-518, <https://doi.org/10.1111/>

- j.1525-142X.2010.00435.x.
- Albertin C B, Simakov O, Mitros T et al. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, **524**(7564): 220-224, <https://doi.org/10.1038/nature14668>.
- Berriman M, Haas B J, LoVerde P T et al. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature*, **460**(7253): 352-358.
- Capella-Gutiérrez S, Silla-Martínez J M, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1 972-1 973.
- Cho S J, Vallès Y, Giani Jr V C et al. 2010. Evolutionary dynamics of the *wnt* gene family: a lophotrochozoan perspective. *Molecular Biology and Evolution*, **27**(7): 1 645-1 658.
- Colbourne J K, Pfrender M E, Gilbert D et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**(6017): 555-561.
- Croce J C, McClay D R. 2008. Evolution of the Wnt pathways. In: Vincan E eds. *Wnt Signaling*. Methods in Molecular Biology. Humana Press, Totowa, NJ, USA. p.3-18.
- Croce J C, Wu S Y, Byrum C et al. 2006. A genome-wide survey of the evolutionarily conserved Wnt pathways in the sea urchin *Strongylocentrotus purpuratus*. *Developmental Biology*, **300**(1): 121-131.
- Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, **468**(7325): 815-818, <https://doi.org/10.1038/nature09632>.
- Du Y S, Zhang L L, Xu F et al. 2013. Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR. *Fish & Shellfish Immunology*, **34**(3): 939-945, <https://doi.org/10.1016/j.fsi.2012.12.007>.
- Fagotto F. 2014. Wnt signaling during early *Xenopus* development. In: Kloc M, Kubiak J Z eds. *Xenopus Development*. John Wiley & Sons, Inc., Oxford, UK.
- Flot J F, Hespeels B, Li X et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, **500**(7463): 453-457.
- Force A, Lynch M, Pickett F B et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**(4): 1 531-1 545.
- Garriock R J, Warkman A S, Meadows S M et al. 2007. Census of vertebrate Wnt genes: isolation and developmental expression of *Xenopus* Wnt2, Wnt3, Wnt9a, Wnt9b, Wnt10a, and Wnt16. *Developmental Dynamics*, **236**(5): 1 249-1 258.
- Gordon M D, Dionne M S, Schneider D S et al. 2005. WntD is a feedback inhibitor of Dorsal/NF- $\kappa$ B in *Drosophila* development and immunity. *Nature*, **437**(7059): 746-749, <https://doi.org/10.1038/nature04073>.
- Gurley K A, Elliott S A, Simakov O et al. 2010. Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Developmental Biology*, **347**(1): 24-39.
- Gyoja F. 2014. A genome-wide survey of bHLH transcription factors in the Placozoan *Trichoplax adhaerens* reveals the ancient repertoire of this gene family in metazoan. *Gene*, **542**(1): 29-37, <https://doi.org/10.1016/j.gene.2014.03.024>.
- Holstein T W. 2012. The evolution of the Wnt pathway. *Cold Spring Harbor Perspectives in Biology*, **4**(7): a007922, <https://doi.org/10.1101/cshperspect.a007922>.
- Howe D G, Bradford Y M, Conlin T et al. 2013. ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Research*, **41**(D1): D854-D860, <https://doi.org/10.1093/nar/gks938>.
- Huelsenbeck J P, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**(8): 754-755.
- Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**(3): 299-314.
- Janssen R, Le Gouar M, Pechmann M et al. 2010. Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation. *BMC Evolutionary Biology*, **10**: 374.
- Katoh K, Standley D M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4): 772-780.
- Kenny N J, Namigai E K O, Marlétaz F et al. 2015. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patellogastropoda) and *Spirobranchus (Pomatoceros) lamarcki* (Annelida, Serpulida). *Marine Genomics*, **24**: 139-146.
- Korswagen H C. 2002. Canonical and non-canonical Wnt signaling pathways in *Caenorhabditis elegans*: variations on a common signaling theme. *BioEssays*, **24**(9): 801-810.
- Kozioł U, Jarero F, Olson P D et al. 2016. Comparative analysis of Wnt expression identifies a highly conserved developmental transition in flatworms. *BMC Biology*, **14**: 10.
- Kusserow A, Pang K, Sturm C et al. 2005. Unexpected complexity of the *Wnt* gene family in a sea anemone. *Nature*, **433**(7022): 156-160, <https://doi.org/10.1038/nature03158>.
- Lavery D L, Davenport I R, Turnbull Y D et al. 2008. Wnt6 expression in epidermis and epithelial tissues during *Xenopus* organogenesis. *Developmental Dynamics*, **237**(3): 768-779, <https://doi.org/10.1002/dvdy.21440>.
- Lengfeld T, Watanabe H, Simakov O et al. 2009. Multiple Wnts are involved in *Hydra* organizer formation and regeneration. *Developmental Biology*, **330**(1): 186-199, <https://doi.org/10.1016/j.ydbio.2009.02.004>.
- Li J J, Huang H Y, Bickel P J et al. 2014. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Research*, **24**(7): 1 086-1 101.
- Livak K J, Schmittgen T D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods*, **25**(4): 402-408, <https://doi.org/10.1006/meth.2001.1228>.

- org/10.1006/meth.2001.1262.
- Logan C Y, Nusse R. 2004. The Wnt signaling pathway in development and disease. *Annual Review of Cell and Developmental Biology*, **20**(1): 781-810, <https://doi.org/10.1146/annurev.cellbio.20.010403.113126>.
- Luo Y J, Takeuchi T, Koyanagi R et al. 2015. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature Communications*, **6**: 8301.
- Marikawa Y. 2006. Wnt/ $\beta$ -catenin signaling and body plan formation in mouse embryos. *Seminars in Cell & Developmental Biology*, **17**(2): 175-184.
- Nusse R, Varmus H E. 1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*, **31**(1): 99-109.
- Nusse R, Varmus H E. 1992. *Wnt* genes. *Cell*, **69**(7): 1 073-1 087.
- Pertea M, Kim D, Pertea G M et al. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, **11**(9): 1 650-1 667.
- Price M N, Dehal A P S, Arkin A P. 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**(3): e9490.
- Prud'homme B, Lartillot N, Balavoine G et al. 2002. Phylogenetic analysis of the *Wnt* gene family. Insights from lophotrochozoan members. *Current Biology*, **12**(16): 1 395-1 400.
- Pruitt K D, Tatusova T, Brown G R et al. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, **40**(D1): D130-D135.
- Pruitt M M, Letcher E J, Chou H C et al. 2014. Expression of the *wnt* gene complement in a spiral-cleaving embryo and trochophore larva. *The International Journal of Developmental Biology*, **58**(6-8): 563-573.
- Putnam N H, Srivastava M, Hellsten U et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, **317**(5834): 86-94, <https://doi.org/10.1126/science.1139158>.
- Qu T, Huang B Y, Zhang L L et al. 2014. Identification and functional characterization of two executioner caspases in *Crassostrea gigas*. *PLoS One*, **9**(2): e89040, <https://doi.org/10.1371/journal.pone.0089040>.
- Ramel M C, Buckles G R, Lekven A C. 2004. Conservation of structure and functional divergence of duplicated Wnt8s in pufferfish. *Developmental Dynamics*, **231**(2): 441-448, <https://doi.org/10.1002/dvdy.20141>.
- Riddiford N, Olson P D. 2011. *Wnt* gene loss in flatworms. *Development Genes and Evolution*, **221**(4): 187-197, <https://doi.org/10.1007/s00427-011-0370-8>.
- Robb S M C, Gotting K, Ross E et al. 2015. SmedGD 2.0: the *Schmidtea mediterranea* genome database. *Genesis*, **53**(8): 535-546, <https://doi.org/10.1002/dvg.22872>.
- Schubert M, Holland L Z, Holland N D et al. 2000. A phylogenetic tree of the *Wnt* genes based on all available full-length sequences, including five from the cephalochordate amphioxus. *Molecular Biology and Evolution*, **17**(12): 1 896-1 903.
- Simakov O, Marletaz F, Cho S J et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493**(7433): 526-531, <https://doi.org/10.1038/nature11696>.
- Simons M, Mlodzik M. 2008. Planar cell polarity signaling: from fly development to human disease. *Annual Review of Genetics*, **42**(1): 517-540, <https://doi.org/10.1146/annurev.genet.42.110807.091432>.
- Sodergren E, Weinstock G M, Davidson E H et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*, **314**(5801): 941-952.
- Srivastava M, Simakov O, Chapman J et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature*, **466**(7307): 720-726, <https://doi.org/10.1038/nature09201>.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1 312-1 313.
- Takeuchi T, Koyanagi R, Gyoja F et al. 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters*, **2**: 3, <https://doi.org/10.1186/s40851-016-0039-2>.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**(5396): 2 012-2 018.
- Tsai I J, Zarowiecki M, Holroyd N et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, **496**(7443): 57-63, <https://doi.org/10.1038/nature12031>.
- Young N D, Jex A R, Li B et al. 2012. Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics*, **44**(2): 221-225, <https://doi.org/10.1038/ng.1065>.
- Zhang G F, Fang X D, Guo X M et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**(7418): 49-54, <https://doi.org/10.1038/nature11413>.
- Zhou Y, Zheng H J, Chen Y Y et al. 2009. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*, **460**(7253): 345-351, <https://doi.org/10.1038/nature08140>.

### Electronic supplementary material

Supplementary material (Supplementary Figs.S1–S4 and Tables S1–S5) is available in the online version of this article at <https://doi.org/10.1007/s00343-018-7132-7>.