

# SCIENTIFIC REPORTS



OPEN

## Evolution of the EGFR pathway in Metazoa and its diversification in the planarian *Schmidtea mediterranea*

Received: 20 April 2016

Accepted: 31 May 2016

Published: 21 June 2016

Sara Barberán<sup>1</sup>, José M. Martín-Durán<sup>2</sup> & Francesc Cebrià<sup>1</sup>

The EGFR pathway is an essential signaling system in animals, whose core components are the epidermal growth factors (EGF ligands) and their trans-membrane tyrosine kinase receptors (EGFRs). Despite extensive knowledge in classical model organisms, little is known of the composition and function of the EGFR pathway in most animal lineages. Here, we have performed an extensive search for the presence of EGFRs and EGF ligands in representative species of most major animal clades, with special focus on the planarian *Schmidtea mediterranea*. With the exception of placozoans and cnidarians, we found that the EGFR pathway is potentially present in all other analyzed animal groups, and has experienced frequent independent expansions. We further characterized the expression domains of the EGFR/EGF identified in *S. mediterranea*, revealing a wide variety of patterns and localization in almost all planarian tissues. Finally, functional experiments suggest an interaction between one of the previously described receptors, *Smed-egfr-5*, and the newly found ligand *Smed-egf-6*. Our findings provide the most comprehensive overview to date of the EGFR pathway, and indicate that the last common metazoan ancestor had an initial complement of one EGFR and one putative EGF ligand, which was often expanded or lost during animal evolution.

In multicellular organisms, communication between individual cells is essential for the regulation of complex biological processes, such as growth, differentiation, tissue renewal, and cell death. In animals, the receptor tyrosine kinases (RTKs) are well known mediators of intercellular signaling<sup>1,2</sup>. In particular, the epidermal growth factor receptor (EGFR) subfamily is the central constituent of a signaling pathway involved in multiple embryonic and adult processes<sup>3</sup>. For instance, it controls the fate and proliferation of multiple cell types in the invertebrate model systems *Caenorhabditis elegans* and *Drosophila melanogaster*<sup>4,5</sup>, and exerts a crucial function in the development and morphogenesis of many organs in vertebrate animals, such as the skin<sup>6</sup>, intestine<sup>7</sup>, the central nervous system<sup>8</sup> and mammary gland<sup>9</sup>, as well as in the progression of prevalent human diseases, such as cancers<sup>3</sup>.

Although constitutive and ligand-independent activations of the EGFR pathway have been described, especially in the context of several cancers<sup>10</sup>, EGFR signaling depends mainly on the binding of a peptide growth factor of the EGF-family to an EGFR. This interaction induces the phosphorylation of the cytoplasmic tyrosine kinase (TK) domain of the EGFR, and the subsequent activation of different intercellular transduction cascades that regulate downstream effector genes<sup>11</sup>. Despite slight differences, all identified EGFRs share conserved structural features that distinguish them from other RTKs. All consist of an extracellular ligand-binding domain containing two or more cysteine rich clusters, a trans-membrane region, and a cytoplasmic domain with tyrosine kinase activity<sup>12</sup>. The nematode *C. elegans* and the fruit fly *D. melanogaster* have one single receptor (*let-23* and *der*, respectively)<sup>13,14</sup>, but vertebrates have four EGFR classes: *egfr* (*ErbB-1/her1*), *ErbB-2 (her2)*, *ErbB-3 (her3)* and *ErbB-4 (her4)*<sup>15</sup>. With respect to the ligands, all known EGF proteins are synthesized as transmembrane precursors characterized by the presence of an EGF-like domain (in some cases multiple EGF-like domains) in their extracellular region<sup>16</sup>. Cleavage of this membrane precursor in the extracellular milieu releases the active

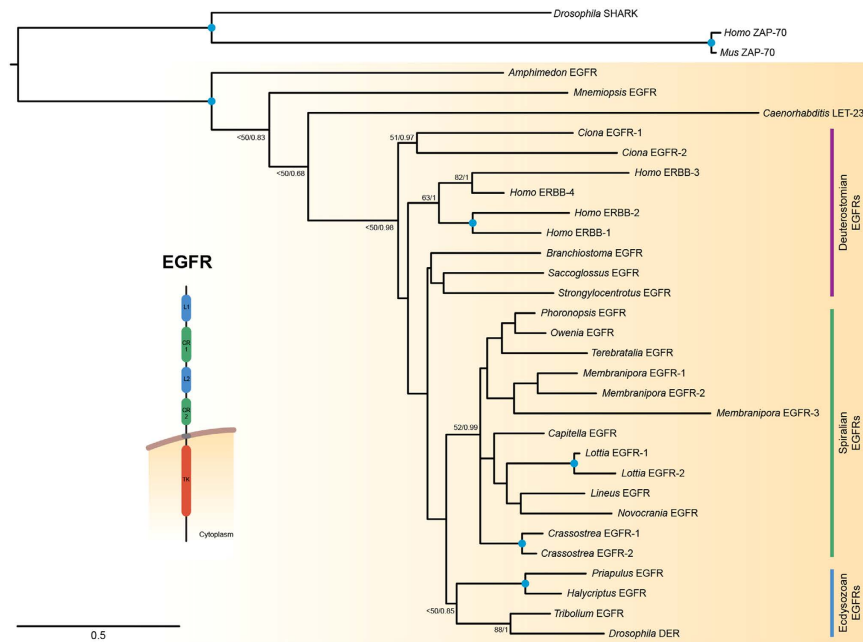
<sup>1</sup>Department of Genetics, Faculty of Biology, University of Barcelona and Institute of Biomedicine of the University of Barcelona (IBUB), Av. Diagonal 643, edifici Prevosti, planta 1, 08028 Barcelona, Catalunya, Spain. <sup>2</sup>Sars International Centre for Marine Molecular Biology, University of Bergen, Thormøhlensgate 55, 5008 Bergen, Norway. Correspondence and requests for materials should be addressed to J.M.M.-D. (email: chema.martin@uib.no) or F.C. (email: fcebrias@ub.edu)

Clade	Species	No. EGFR	TM <sup>1</sup>	SP	Cys rich	Inactive TK
Ctenophora	<i>M. leidyi</i>	1	1	Yes	3	No
Porifera	<i>A. queenslandica</i>	1	1	?	1	No
Placozoa	<i>T. adhaerens</i>	0	–	–	–	–
Cnidaria	<i>N. vectensis</i>	0	–	–	–	–
	<i>H. magnipapillata</i>	0	–	–	–	–
Xenacoelomorpha	<i>X. bocki</i>	2	1(2) <sup>2</sup>	Yes	2(1), 3(1)	No
	<i>M. stichopi</i>	3	1(3)	Yes	2(1), 3(1), 4(1)	No
	<i>I. pulchra</i>	5	1(4), ?(1)	Yes(3), ?(2)	1(2), 3(2), ?(1)	No
	<i>C. macropyga</i>	3	1(3)	Yes	4(2), 1(1)	No
Priapulida	<i>P. caudatus</i>	1	1	?	3	No
	<i>H. spinulosus</i>	1	1	Yes	3	No
Nematoda	<i>C. elegans</i>	1	1	Yes	3	No
Arthropoda	<i>T. castaneum</i>	1	1	Yes	2	No
	<i>D. melanogaster</i>	1	1	Yes	3	No
Gastrotricha	<i>L. squamata</i>	0	–	–	–	–
Platyhelminthes	<i>M. lignano</i>	8	1(6), ?(2)	Yes(5), ?(3)	1(1), 2(2), 3(1), 4(1), ?(3)	No
	<i>P. vittatus</i>	6	1(6)	Yes(5), ?(1)	2(2), 3(4)	No
	<i>S. mediterranea</i>	6	1(6)	Yes(5), ?(1)	3(4), 4(2)	No(5), Yes(1)
	<i>S. mansoni</i>	3	1(3)	Yes(1), ?(2)	3(1), 5(2)	No
	<i>E. multilocularis</i>	3	1(3)	Yes(1), ?(2)	3(2), 4(1)	No
Mollusca	<i>L. gigantea</i>	2	1(1), ?(1)	Yes(1), ?(1)	2(1), 3(1)	No(1), Yes(1)
	<i>C. gigas</i>	2	1(1), ?(1)	?(2)	?	No
Annelida	<i>O. fusiformis</i>	1	1	Yes	3	No
	<i>C. teleta</i>	1	1	?	3	No
	<i>H. robusta</i>	6	1(4), ?(2)	?(6)	2(1), 3(1), 4(2), ?(2)	No
Nemertea	<i>L. ruber</i>	1	1	Yes	3	No
Bryozoa	<i>M. membranacea</i>	3	1(2), 2(1)	Yes(2), ?(1)	3	No
Brachiopoda	<i>T. transversa</i>	1	1	Yes	3	No
	<i>N. anomala</i>	1	1	Yes	3	No
Phoronida	<i>P. harmeri</i>	1	?	?	?	No
Hemichordata	<i>S. kowalevskii</i>	1	1	?	4	No
Echinodermata	<i>S. purpuratus</i>	1	1	?	4	No
Chordata	<i>B. floridae</i>	1	?	?	?	No
	<i>C. intestinalis</i>	2	1	Yes(1), ?(1)	1(1), 3(1)	No
	<i>H. sapiens</i>	4	1	Yes	2	No(3), Yes(1)

**Table 1. Distribution and diversity of EGFRs in Metazoa.** <sup>1</sup>TM means trans-membrane domains and SP signal peptide. <sup>2</sup>in parenthesis, number of identified genes exhibiting that number of trans-membrane domains, signal peptide, cysteine-rich domains and inactive/active TK domains.

EGF ligand, which contains a conserved cysteine-containing motif that confers binding specificity<sup>17</sup>. While *C. elegans* has only one ligand (*lin-3*)<sup>18</sup>, *D. melanogaster* has four (*spitz*, *vein*, *gurken* and *keren*)<sup>4,19</sup>, and vertebrates have an expanded set of 11 EGF ligands: the *epidermal growth factor (egf)*, *amphiregulin (areg)*, *epigen*, *epiregulin (epr)*, *betacellulin (bct)*, *heparin-binding EGF-like growth factor (hb-egf)*, *transforming growth factor- $\alpha$  (TGF $\alpha$ )* and four neuregulins (*nrg1-4*)<sup>16,20,21</sup>. Notably, an additional extracellular immunoglobulin-like domain is present in the *D. melanogaster* ligand *vein*, and in the vertebrate ligands *nrg1* and *nrg2*, which also contain an intracellular neuregulin domain<sup>21,22</sup>.

In recent years, research on the planarian flatworm *Schmidtea mediterranea* (Platyhelminthes)<sup>23–25</sup> has clarified the role of the EGFR signaling pathway in the Spiralia (previously referred to as Lophotrochozoa<sup>26</sup>), which together with the Ecdysozoa (e.g. *C. elegans* and *D. melanogaster*) and the Deuterostomia (e.g. vertebrates) form the three major groups of bilaterally symmetrical animals<sup>26</sup>. The planarian *S. mediterranea* is a well-established model in regenerative research, due to its striking ability to regrow any missing body part after injury<sup>27</sup>. Previous studies have identified four EGFRs in *S. mediterranea* (*Smed-egfr-1*, *Smed-egfr-2*, *Smed-egfr-3* and *Smed-egfr-5*), but only one representative in the related parasitic flatworm species *Echinococcus multilocularis* (*EmER*) and *Schistosoma mansoni* (*SER*)<sup>23,25,28,29</sup>. Additionally, one putative EGF ligand (*Smed-epiregulin-1*) and a neuregulin-like ligand (*Smed-nrg-1*) have been identified<sup>30,31</sup>. Interestingly, the planarian EGFRs and ligand show a broad spectrum of expression patterns<sup>23,25,30</sup> and are functionally implicated in the regeneration and homeostasis of different organs and cell types, such as the pharynx, the excretory and nervous systems, and the eye pigment cells<sup>23,25</sup>. Despite this progress in a member of the Spiralia, the presence, composition, and role of the EGFR pathway in most animal lineages –apart from the classical developmental and biomedical model systems–, is virtually



**Figure 1. Maximum likelihood (ML) phylogenetic tree of representative EGFRs as obtained by RAxML.**

The tree was rooted with the EGFR-related tyrosine kinases SHARK and ZAP-70. The model of protein evolution used was LG + G + I. Nodal support was obtained by RAxML 1000 replicates (bootstrap value [BV]) and Bayesian posterior probabilities (PP). Both values are shown for key branches. A blue dot at the node indicates BV > 95% and PP > 0.95. Fast evolving sequences, such as those of the planarian *S. mediterranea*, were not included in this analysis (for the entire dataset, see Additional file 1: Figure S1). The archetypal domain architecture of an EGFR is shown on the left.

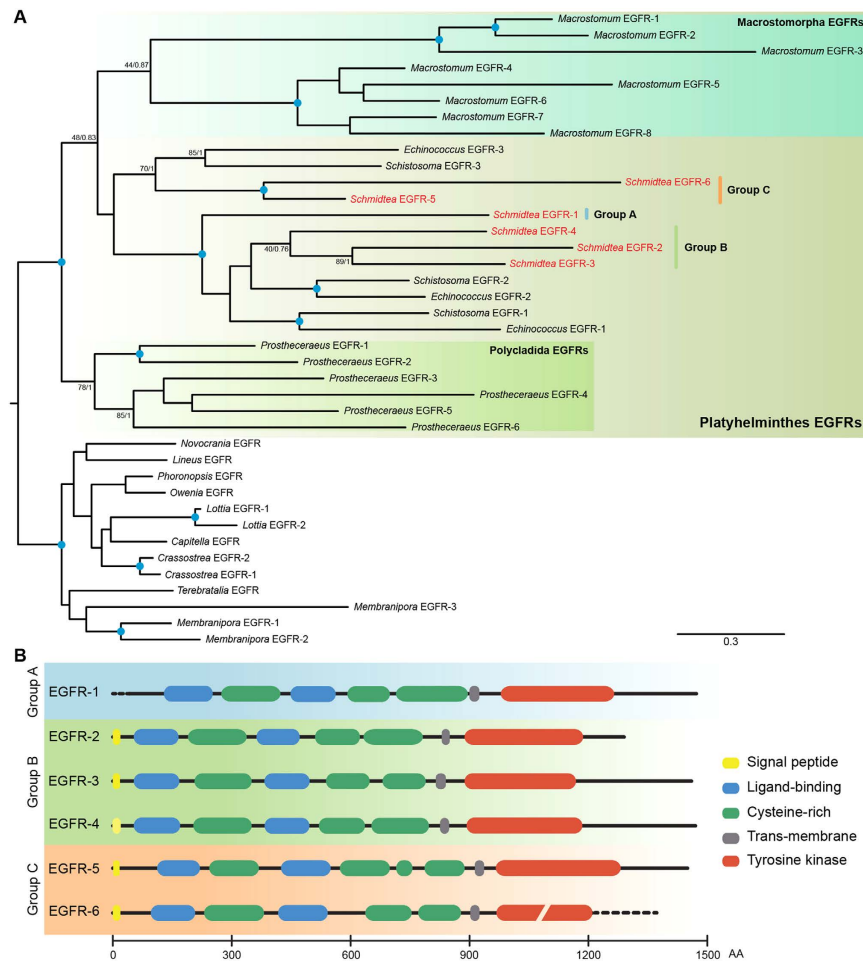
unknown. Accordingly, we also know little about the evolution and functional diversification of this important signaling pathway during animal history.

In this study, we use the numerous accessible genomic and transcriptomic datasets to present phylogenetic analyses of the two key EGFR signaling pathway components –EGFRs and EGF ligands– in representatives of most major animal clades, together with the identification and characterization of two new EGFRs and nine new EGF ligands in the planarian *S. mediterranea*. Our findings indicate that *S. mediterranea* has a total of six EGFRs and nine EGF ligands, which exhibit a wide variety of expression patterns. Silencing of the newly identified EGF ligand *Smed-egf-6* by RNA interference phenocopies the previously reported role of the EGFR *Smed-egfr-5*<sup>25</sup>, suggesting a mechanistic connection between them. Altogether, our data permit us to hypothesize that the last common metazoan ancestor had one EGFR and one probable EGF-like ligand, and that the neuregulin-like type of ligands appeared in the last common bilaterian ancestor. Our study indicates that the evolutionary history of the EGFR pathway is characterized by frequent independent expansions of an originally minimal set of one EGFR and one (Metazoa) or two (Bilateria) EGF ligands. These data may help us to understand how this signaling pathway has acquired its broad diversity of physiological roles, as observed in extant animals.

## Results and Discussion

**Distribution and diversification of EGFRs in Metazoa.** A thorough investigation of the genomic and transcriptomic data available for a wide diversity of animal lineages revealed that the EGFR subfamily of RTKs is present in almost all animal groups (Table 1), including ctenophores and sponges<sup>32</sup>, the two earliest branching animal clades<sup>26</sup>. Significant exceptions are cnidarians and placozoans, for which the absence of an EGFR ortholog has been previously reported<sup>33,34</sup>, and the gastrotrich *Lepidodermella squamata*. In this latter case, further sequencing and sampling efforts will be required to confirm this absence. The orthology assignment of all identified EGFR sequences was confirmed by a phylogenetic analysis of the tyrosine kinase (TK) domain, including representatives of all metazoan RTKs subfamilies and the RTKs of the amoebozoan *Acanthamoeba castellanii* (Supplementary Figure 1). This analysis was corroborated by a further phylogenetic analysis of the whole EGFR sequence (Fig. 1).

Despite particular relationships are not strongly supported in our orthology analysis, the number, distribution (Table 1) and general phylogenetic relationships (Fig. 1) of the EGFRs suggest that a single copy of this RTK was present in the last common metazoan ancestor, as well as at the base of the three major bilaterian groups: the Deuterostomia, the Ecdysozoa and the Spiralia. Subsequent independent expansions occurred in certain bilaterian lineages, such as xenacoelomorphs, tunicates, vertebrates, platyhelminthes, clitellate annelids, molluscs, and bryozoans. While the overall domain architecture of all EGFRs is highly similar (Fig. 1), the number of extracellular cysteine rich domains can vary considerably (Table 1). The previous assumption that vertebrate and invertebrate EGFRs differ in the number of cysteine rich domains (2 in vertebrate EGFRs; 3 in the EGFRs of *D. melanogaster* and *C. elegans*)<sup>35</sup> is not supported by our data. In fact, the copy number of this extracellular

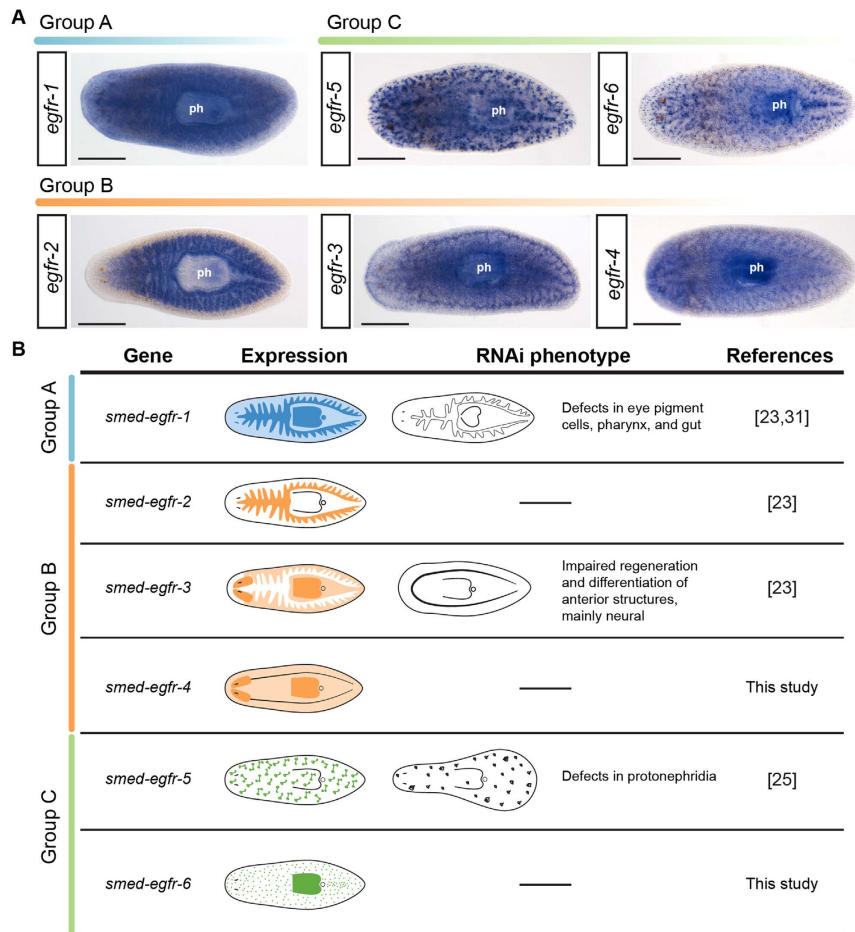


**Figure 2. Maximum likelihood (ML) phylogenetic tree of flatworm EGFRs as obtained by RAxML.**

(A) The tree was rooted with the EGFRs of representative spiralian taxa. The model of protein evolution used was LG + G + I. Nodal support was obtained by RAxML 550 replicates (bootstrap value [BV]) and Bayesian posterior probabilities (PP). Both values are shown in key branches. A blue dot in the node indicates BV > 95% and PP > 0.95. Planarian sequences are highlighted in red. (B) Schematic representation of the domain architecture of each *S. mediterranea* EGFR, drawn to scale.

domain seems to oscillate between 2 and 5, and rather than being species-specific, it appears to vary even between paralog EGFRs (Table 1). Interestingly, we found EGFR sequences with potentially inactive (or at least highly divergent) tyrosine kinase domains in the planarian *S. mediterranea* and the limpet *Lottia gigantea* (Table 1) (Supplementary Figure 2). This condition is similar to the situation observed in the human *ErbB3* receptor<sup>36–39</sup>, and suggests that EGFR duplication in some spiralian lineages has been accompanied by the evolution of alternative regulation/modulation of EGFR signaling.

**The EGFR complement of the planarian *S. mediterranea*.** The re-investigation of the most recent transcriptome and genome resources of *S. mediterranea*<sup>40,41</sup> using the newly identified EGFR sequences of the early branching polyclad flatworm *Prostheceraeus vittatus* (this study) as queries<sup>42</sup>, allowed us to recover two new EGFRs, which we named *Smed-egfr-4* and *Smed-egfr-6*. As such, *S. mediterranea* includes six EGFRs in its genome, which is in the range of the other flatworm species that we analyzed: 8 EGFRs in the macrostomid *Macrostomum lignano*, 6 EGFRs in the polyclad *P. vittatus*, and 3 EGFRs in the parasites *Schistosoma mansoni* and *Echinococcus multilocularis* (Table 1). An orthology analysis of the recovered flatworm EGFRs using other spiralian sequences as an outgroup (Fig. 2A) revealed that the expansions observed in Macrostomorpha, Polycladida and Neoophora (planarians and parasites) occurred independently, and thus there was likely one single EGFR copy in the last common platyhelminth ancestor, as in most other spiralian lineages. Importantly, planarian EGFRs group in three distinct clades, which we named group A, B, and C (Fig. 2A). The receptor *Smed-egfr-1* is the single representative of group A (Fig. 2B). Group B includes the receptors *Smed-egfr-2*, *Smed-egfr-3* and *Smed-egfr-4*, which exhibit canonical domain architectures, but a varying number of cysteine rich domains (Fig. 2B). Finally, group C consists of the receptors *Smed-egfr-5* and *Smed-egfr-6*, the latter with key amino acid substitutions that may inactivate the ATP binding and active sites of the tyrosine kinase domain (Supplementary Figure 2).



**Figure 3. Expression patterns of planarian EGFRs. (A)** Whole-mount *in situ* hybridizations of all planarian EGFRs on intact adult specimens of *S. mediterranea*, grouped according to their phylogenetic relationship. See main text for details of the expression patterns. **(B)** Schematic summary of the expression patterns of planarian EGFRs and the phenotype observed after their silencing by dsRNA injection, a dash indicates no phenotype was observed (see references for further details). In **(A)**, anterior is to the left. ph: pharynx. Scale bars: 500  $\mu$ m in all panels.

Previous studies characterized the expression and function of the *S. mediterranea* EGFRs *Smed-egfr-1*, *Smed-egfr-2*, *Smed-egfr-3* and *Smed-egfr-5* in adult worms<sup>23,25</sup>. The receptor *Smed-egfr-1* is expressed in the gut, pharynx, and eye pigment cells (Fig. 3A), and is required for the proper regeneration and homeostasis of the eyes, pharynx and gut (Fig. 3B)<sup>23,31</sup>. The paralog *Smed-egfr-2* is expressed in the gut (Fig. 3A), and no apparent phenotype is observed after silencing<sup>23</sup>. The receptor *Smed-egfr-3* is detected in the neoblast, pharynx and the cephalic ganglia (Fig. 3A) and is required for the proper growth of the blastema during regeneration, probably by regulating cell differentiation (Fig. 3B)<sup>23,24</sup>. Finally, *Smed-egfr-5* is expressed in the excretory system (Fig. 3A) and is required for its proper regeneration (Fig. 3B)<sup>25</sup>. We performed whole mount *in situ* hybridization experiments to characterize the expression domains of the newly identified *Smed-egfr-4* and *Smed-egfr-6* genes. The receptor *Smed-egfr-4* was mainly expressed in the central nervous system and pharynx, and weakly in the mesenchyme (Fig. 3A). The paralog *Smed-egfr-6* was expressed in the pharynx and in a discrete pattern throughout the body (Fig. 3A). The silencing by RNAi of either *Smed-egfr-4* or *Smed-egfr-6* gave no perceivable external phenotypes (data not shown). Altogether, our findings support that the expansion of EGFRs that occurred in the lineages leading to *S. mediterranea* has been accompanied by a molecular, transcriptional and functional diversification of the different paralogs. However, further expression and functional analysis in other platyhelminthes and spiralian lineages are essential to better understand the ancestral role of the EGFR in these organisms.

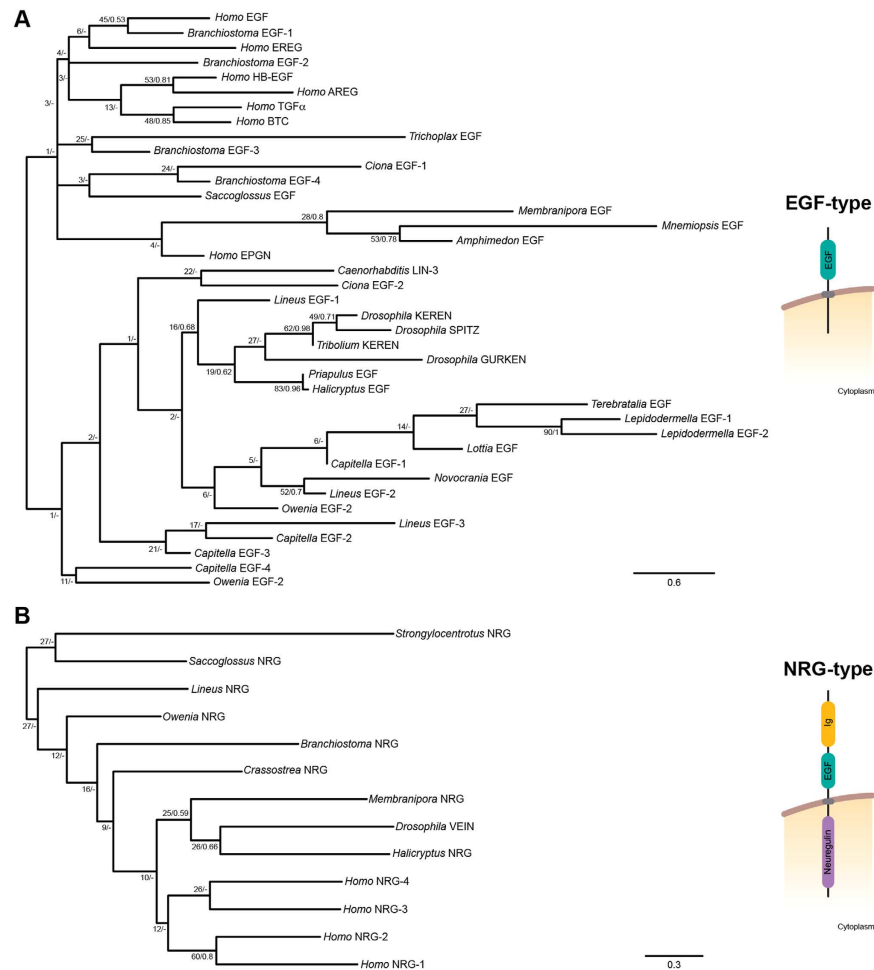
**Distribution of EGF ligands in Metazoa.** The evolution of EGF ligands has been mostly studied in vertebrates, *C. elegans*, and *D. melanogaster*<sup>17,43–45</sup>. Because the only common feature to all ligands is the presence of an EGF domain, which is a motif present in a tremendous diversity of proteins<sup>17</sup>, we designed a conservative approach to identify new putative EGF ligands. Candidate genes had to be full length (see Material and Methods for the very few exceptions to this criteria), with only one EGF motif, and with domain architectures consistent with the structures of already known EGF ligands<sup>17</sup>. Our *in silico* searches reported 75 new potential

Clade	Species	No. EGF ligands	EGF-like	Neuregulin-like
Ctenophora	<i>M. leidy</i>	1	1	0
Porifera	<i>A. queenslandica</i>	1	1	0
Placozoa	<i>T. adhaerens</i>	1	1	0
Cnidaria	<i>N. vectensis</i>	0	–	–
	<i>H. magnipapillata</i>	0	–	–
Xenacoelomorpha	<i>X. bocki</i>	2	1	1
	<i>M. stichopi</i>	5	4	1
	<i>I. pulchra</i>	5	2	3
	<i>C. macropyga</i>	2	0	2
Priapulida	<i>P. caudatus</i>	1	1	0
	<i>H. spinulosus</i>	2	1	1
Nematoda	<i>C. elegans</i>	1	1	0
Arthropoda	<i>T. castaneum</i>	1	1	0
	<i>D. melanogaster</i>	4	3	1
Gastrotricha	<i>L. squamata</i>	2	2	0
Platyhelminthes	<i>M. lignano</i>	2	1	1
	<i>P. vittatus</i>	9	7	2
	<i>S. mediterranea</i>	9	8	1
	<i>S. mansoni</i>	1	0	1
	<i>E. multilocularis</i>	2	1	1
Mollusca	<i>L. gigantea</i>	1	1	0
	<i>C. gigas</i>	1	0	1
Annelida	<i>O. fusiformis</i>	3	2	1
	<i>C. teleta</i>	4	4	0
	<i>H. robusta</i>	4	3	1
Nemertea	<i>L. ruber</i>	4	3	1
Bryozoa	<i>M. membranacea</i>	2	1	1
Brachiopoda	<i>T. transversa</i>	1	1	0
	<i>N. anomala</i>	1	1	0
Hemichordata	<i>S. kowalevskii</i>	2	1	1
Echinodermata	<i>S. purpuratus</i>	1	0	1
Chordata	<i>B. floridae</i>	5	4	1
	<i>C. intestinalis</i>	2	2	0
	<i>H. sapiens</i>	11	7	4

**Table 2. Distribution and diversity of EGF ligands in Metazoa.**

EGF ligands (Table 2), in addition to those already described for *Homo sapiens*, *C. elegans*, *D. melanogaster* and *Tribolium castaneum*. Consistent with the absence of an EGFR, we did not recover any putative EGF ligands in the cnidarian databases analyzed. An analysis of domain compositions revealed two basic architectures: the EGF-type of ligands, with only an EGF domain and a varying number of transmembrane regions (from zero, in soluble candidates, to two); and the neuregulin-type (NRG-type) of ligands, which showed an EGF domain combined with an immunoglobulin I-set domain and a transmembrane region (Supplementary Figure 3). Representatives of the first group are the *H. sapiens* EGF ligands *egf*, *hb-egf*, *TGF $\alpha$* , *areg*, *epr*, *btc*, and *epigen*, *lin-3* from *C. elegans*, and the *D. melanogaster* ligands *gurken*, *keren* and *spitz*. Examples of the NRG-type of EGF ligands are the human *nrg-1* and *nrg-2*, and the *D. melanogaster* gene *vein*. While we were able to recover putative ligands of the EGF-type in almost all animal lineages analyzed –with the exception of the sea urchin *Strongylocentrotus purpuratus* and the phoronid *Phoronopsis harmerii*–, NRG-type ligands were only identified in bilaterian groups (Table 2).

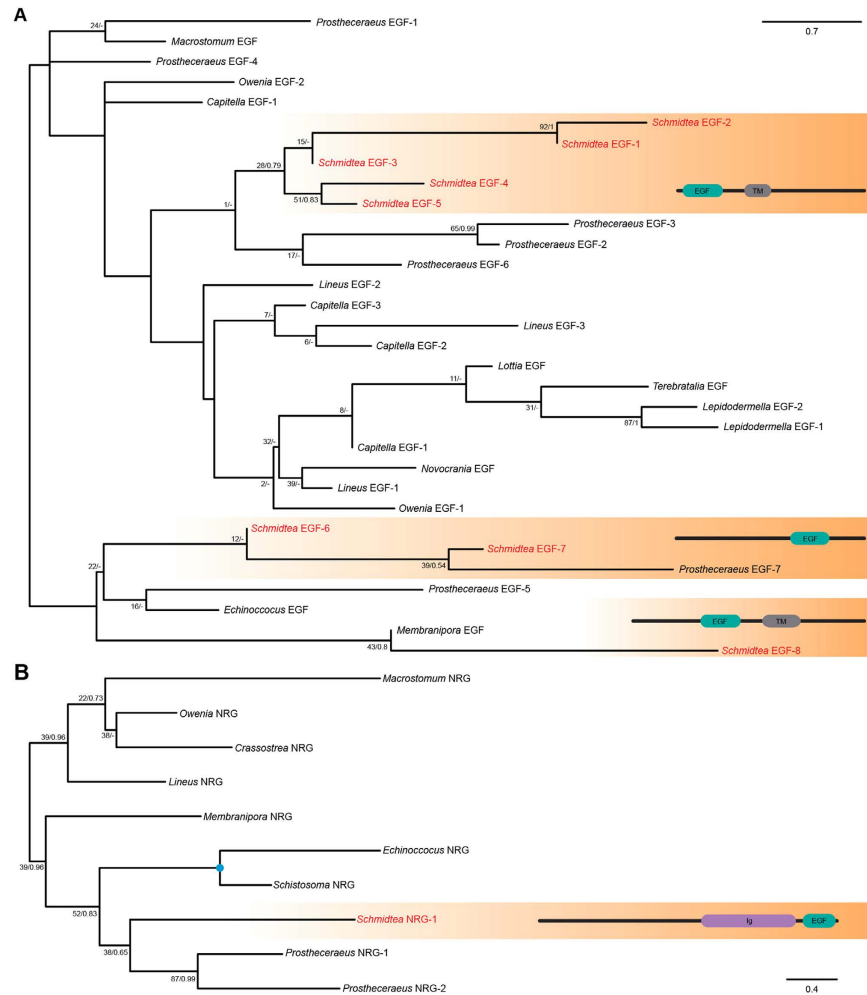
Phylogenetic studies of the EGF ligands tend to analyze vertebrate and invertebrate sequences separately<sup>43</sup>, to avoid artifacts associated with the low phylogenetic signal obtained from the EGF domain. Therefore, we first performed a sequence-similarity-based clustering analysis of all retrieved putative EGF ligands, with the aim of supporting our domain-based classification (Supplementary Figure 4). NRG-type ligands clustered together, which supported considering them a *bona fide* subtype, but most EGF-type ligands were highly dispersed and weakly interconnected. The failure to recover a well-defined cluster of EGF-type ligands is another sign of the low phylogenetic signal of EGF ligands. This is most likely caused by the extreme variability in the protein structure of these genes, and the fact that the only common region –the EGF motif– is also very variable, beyond the few conserved amino acids that are invariable among all species (e.g. the six conserved cysteines<sup>17</sup>). Following the results of the clustering analysis, we studied the orthology assignments of the two major subtypes of EGF ligands



**Figure 4. Maximum likelihood (ML) phylogenetic tree of the putative EGF ligands as obtained by RAxML.** (A) Unrooted tree of the EGF-type of EGF ligands, using the Whelan and Goldman (WAG) + G model of protein evolution. Nodal support was obtained by RAxML 500 replicates (bootstrap value [BV]) and Bayesian posterior probabilities (PP). (B) Unrooted tree of the NRG-type of EGF ligands, using the LG + G model of protein evolution. Nodal support was obtained by RAxML 1000 replicates (BV) and Bayesian posterior probabilities (PP). Fast evolving sequences, such as those of the planarian *S. mediterranea*, were not included in these analyses (for the entire dataset, see Additional files 4, 5: Figure S4, S5). The archetypal domain architectures of the EGF-type and NRG-type of ligands are shown on the left.

separately (Fig. 4A,B) (Supplementary Figures 5 and 6). In line with previous reports<sup>43</sup> and our findings after the sequence-similarity analysis, we could not resolve the deepest relationships between all metazoan EGF-type putative ligands. Regarding the NRG-type, we could tentatively assign orthology relationships between the two ambulacrarian NRGs, the two ecdysozoan NRGs, and the four human NRGs, which confirmed a previous analysis of vertebrate EGF ligands<sup>43</sup>. All things considered, the distribution of EGF ligands, the domain architecture of the diversity of candidate genes, and the clustering and phylogenetic analyses, support dividing the EGF ligands into two separate classes: the EGF-type and the NRG-type. The presence of the EGF-type in the ctenophore *Mnemiopsis leidyi*, the sponge *Amphimedon queenslandica* and the placozoan *Trichoplax adhaerens* indicates that this was likely the ancestral EGFR ligand. However, further biochemical analyses are needed to confirm that this putative EGF ligand actually binds the EGFRs present in the ctenophore and the sponge, and thus the existence of a truly functional EGFR signaling pathway in these early branching animal lineages. Consistent with its presence in xenacoelomorphs and most other bilaterian groups, the NRG-type appears to be an innovation of Bilateria.

**The EGF ligand complement of *S. mediterranea*.** Our search for EGF ligands in the genome<sup>40</sup> and publicly available transcriptomes of *S. mediterranea*<sup>41</sup> reported nine candidate genes (Table 2), including the previously described *Smed-epiregulin-1*<sup>30</sup> and the recently identified *Smed-nrg-1*<sup>31</sup>. Eight of these were of the EGF-type, and only one was of the NRG-type. *Smed-epiregulin-1* belongs to the EGF-type, and given the fact that the gene *epiregulin* is a vertebrate-specific paralog of an ancestral EGF-like gene<sup>44</sup>, we renamed this planarian candidate as *Smed-egf-1*. We named the rest of the planarian putative EGF-type ligands consecutively, from *Smed-egf-2* to *Smed-egf-8*, and the only NRG-type ligand as *Smed-nrg-1*<sup>31</sup>. The phylogenetic analysis of spiralian putative

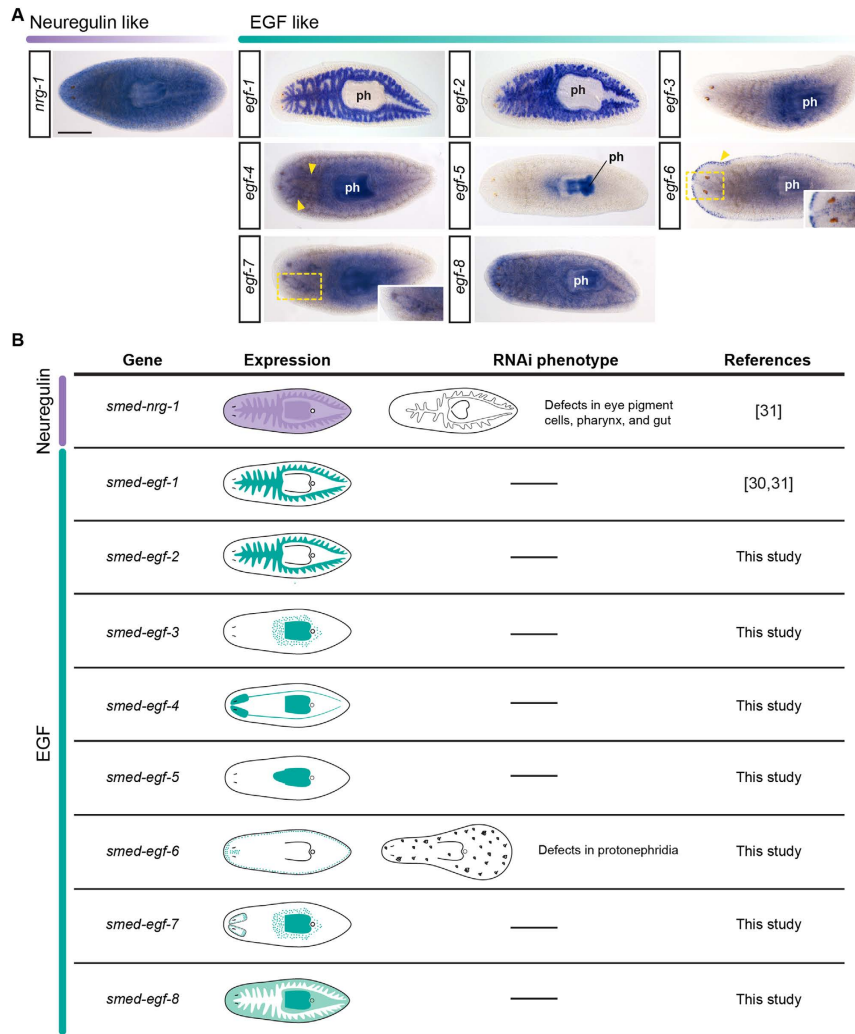


**Figure 5. Maximum likelihood (ML) phylogenetic tree of the putative flatworm EGF ligands as obtained by RAXML. (A)** Unrooted tree of the EGF-type of EGF ligands, using the Whelan and Goldman (WAG) + G model of protein evolution. Nodal support was obtained by RAXML 900 replicates (bootstrap value [BV]) and Bayesian posterior probabilities (PP). **(B)** Unrooted tree of the NRG-type of EGF ligands, using the LG + G model of protein evolution. Nodal support was obtained by RAXML 450 replicates (BV) and Bayesian posterior probabilities (PP). The domain architectures of each subgroup of planarian EGF ligands are shown on the right. Planarian sequences are highlighted in red.

EGF-type ligands did not recover monophyly of the platyhelminth sequences (Fig. 5A), most probably due to the low phylogenetic signal of the EGF domain. However, this analysis suggested three planarian-specific clusters (Fig. 5A). The first group includes from *Smed-egf-1* to *Smed-egf-5*, which are all characterized by presenting the EGF domain upstream of the transmembrane region. The second cluster consists of *Smed-egf-6* and *Smed-egf-7*, and both are secreted ligands without transmembrane domain. Finally, *Smed-egf-8* branches independently from other planarian EGF ligands in our orthology analyses, and has a domain architecture similar to the group of *Smed-egf-1* to *Smed-egf-5*. As observed in the current genomic version<sup>40</sup>, the presence of introns in all the identified ligands and the six EGFRs suggests that the expansion of these two gene families did not occur by retroposition, but rather by gene duplications.

We next analyzed the expression of the identified putative EGF ligands in adult intact specimens of *S. mediterranea* (Fig. 6A). The ligands *Smed-egf-1* and *Smed-egf-2* were expressed in the gut, and *Smed-egf-3* transcripts localized to the pharynx and in scattered mesenchymal cells around it, a pattern reminiscent of other pharynx-related genes, such as *foxA*<sup>46,47</sup>. Expression of *Smed-egf-4* was detected in the pharynx, the mesenchyme around it, and the central nervous system (Fig. 6A), while *Smed-egf-5* was expressed in the pharynx and esophagus (Fig. 6A). The putative ligand *Smed-egf-6* was mainly detected in isolated cells of the lateral margin of the planarian, at the boundary between the dorsal and the ventral epidermis. Notably, the expression was stronger and the number of cells greater in the margin of the head, where *Smed-egf-6* also localized to isolated cells of the anterior dorsal midline (Fig. 6A; inset). The EGF-type putative ligand *Smed-egf-7* was expressed in the brain – strongly in the posterior region of the cephalic ganglia –, pharynx and the mesenchyme around it (Fig. 6A), and *Smed-egf-8* was detected in the mesenchyme and pharynx. Finally, *Smed-nrg-1* was expressed in the pharynx and mesenchyme (Fig. 6A). Altogether, these findings show that the nine putative EGF ligands of *S. mediterranea*

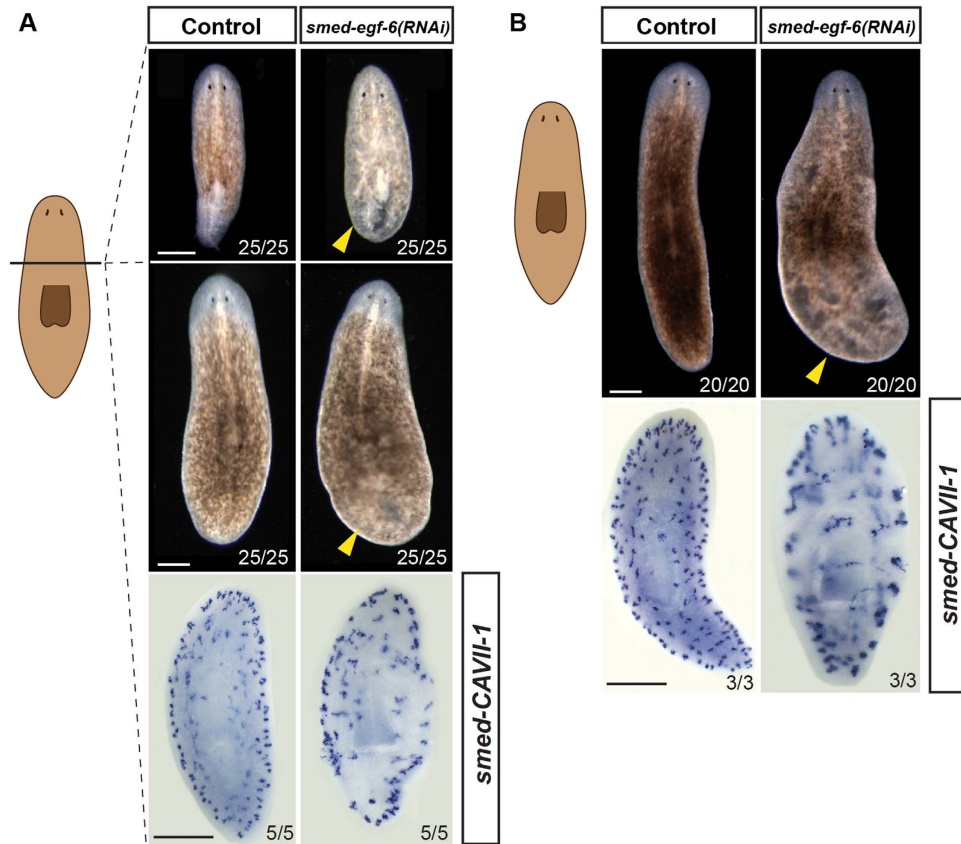




**Figure 6. Planarian epidermal growth factors ligands.** (A) Whole-mount *in situ* hybridizations of all putative planarian EGF ligands on intact adult specimens of *S. mediterranea*, grouped according to their subtype and phylogenetic relationships. See main text for details of the expression patterns. In *egf-4*, the yellow arrowheads indicate the central nervous system. In *egf-6*, the yellow arrowheads indicate expression in the body margin, and the inset (region delimited by the yellow rectangle) is a magnification of the expression in the dorsal anterior midline. Faint signal in the pharynx and central body region in *egf-6* is background staining. In *egf-7*, the inset is a magnification showing the expression in the brain. (B) Schematic summary of the expression patterns of the planarian putative EGF ligands and the phenotype observed after their silencing by dsRNA injection (see references and main text for further details). In (A), anterior is to the left. ph: pharynx. Scale bars: 500  $\mu$ m in all panels.

exhibit a broad diversity of expression patterns, relating mostly to pharyngeal and gut tissues as well as the nervous system (Fig. 6B).

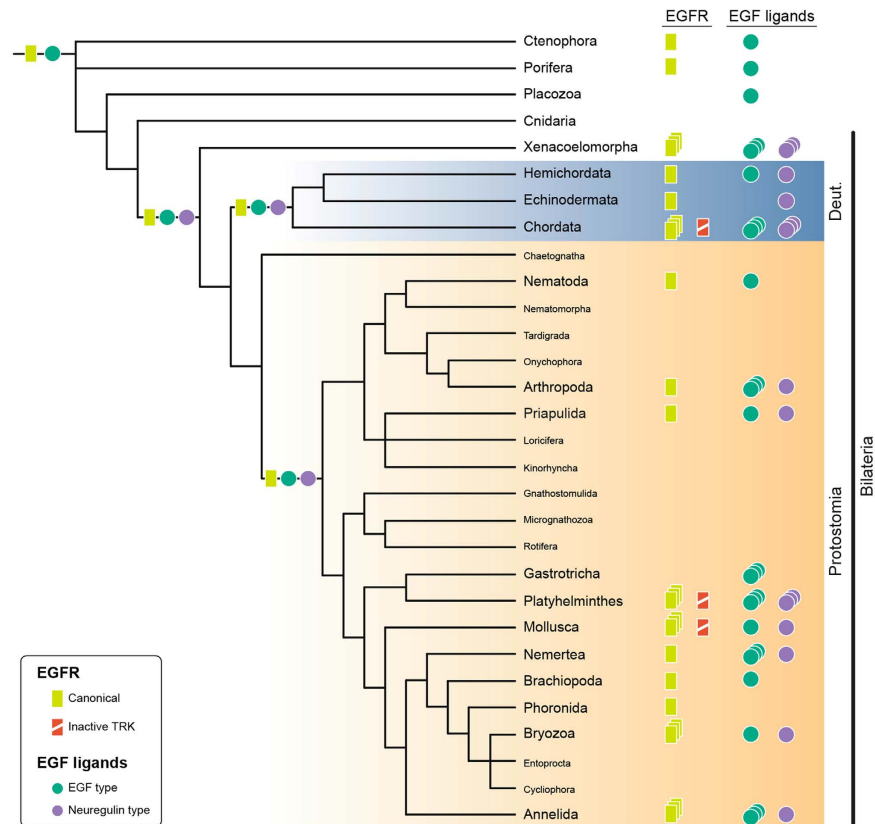
To gain a better understanding of the function of these candidate ligands, and their potential relationships with the described EGFRs, we performed single RNAi gene silencing by double stranded RNA (dsRNA) injections and assessed the consequences during anterior and posterior regeneration. Out of the eight putative EGF-type ligands, we only observed an apparent phenotype after *Smed-egf-6* RNAi, which could indicate some level of functional redundancy between these ligands. However, we cannot discard the possibility that those for which we did not observe any phenotype could be involved in more cell-specific roles that we could not discriminate at the gross morphological level. Silencing of *Smed-egf-6* did not affect the regeneration of missing structures, but rather, produced edemas (Fig. 7A), a phenotype remarkably similar to the one observed after *Smed-egfr-5* RNAi<sup>25</sup>. To determine if the swollen phenotype was caused by defects in the excretory system, as described for *Smed-egfr-5*(RNAi) animals, we analyzed the expression of the specific protonephridial marker carbonic anhydrase (*CAVII-1*)<sup>25</sup>. As expected, *Smed-egf-6*(RNAi) regenerating animals showed a reduced number of, and aberrant, protonephridial tubules (Fig. 7A), suggesting that the edemas were caused by an abnormal regeneration of this organ system. We further confirmed these results during adult homeostasis in intact animals (Fig. 7B). These findings suggest that *Smed-egf-6* could act through *Smed-egfr-5* to regulate planarian excretory



**Figure 7. Role of *Smed-egf-6* during adult planarian regeneration and homeostasis.** Regenerating (A) and intact (B) animals were injected on three consecutive days, and fixed two weeks after the last injection. Treated animals form edemas (yellow arrowheads) during both regeneration (A) and homeostasis (B). Whole-mount *in situ* hybridization of the protonephridial marker *Smed-CAVII-1* in *Smed-egf-6(RNAi)* animals demonstrates a decrease in the number of protonephridial tubules compared to control animals. Numbers in each panel refer to the frequency of the phenotype. In all panels, anterior is to the top. Scale bars: 500  $\mu$ m in all panels.

system regeneration and homeostasis. We have also recently reported that silencing *Smed-nrg-1* causes an abnormal regeneration of the eyes, gut and pharynx<sup>31</sup>, suggesting that this putative ligand signals via the *Smed-egfr-1* receptor. Although we have only observed an apparent phenotype for two of the nine putative ligands, their roles during regeneration and homeostasis are consistent with those previously assigned to planarian EGFRs. We propose that this concurrence validates our bioinformatic approach for ligand identification.

In summary, the demonstrable relevance of the EGFR pathway in the fields of biomedicine, cancer and developmental biology<sup>3</sup> contrasts with our limited understanding of the evolution and functional diversification of this pathway in the animal tree of life. Our broad genomic survey of representatives of 19 different major animal clades suggests that the EGFR pathway probably evolved within the metazoan stem lineage (Fig. 8)<sup>32</sup> and comprised an EGF receptor and a single putative EGF-like ligand, as observed in extant ctenophores and sponges. The EGFR pathway appears to be missing from placozoan and cnidarian lineages, while the presence of a neuregulin-like putative ligand in the vast majority of bilaterian species analyzed indicate that this new type of EGFR ligand appeared together with the Bilateria (Fig. 8). Subsequently, the EGFR and/or the EGF ligands have frequently become expanded in many bilaterian lineages, and new modes of signaling modulation (e.g. via inactive Tyr kinase domains) have appeared (Fig. 8). The planarian *S. mediterranea* (Platyhelminthes) is a prototype of these evolutionary events, because it exhibits an EGFR pathway with 6 EGF receptors –including one putative inactive EGFR–, and 9 EGF ligands –1 neuregulin-like and 8 EGF-like. Gene expression and functional analyses demonstrate that these components are detected in most of the differentiated tissues of the planarian (Figs 3 and 6)<sup>23,25</sup>, and control the regeneration and homeostasis of different organs, tissues, and body regions of the animal (Fig. 7), probably by regulating cell differentiation<sup>23–25,31</sup>. Altogether, our findings deliver an expanded and detailed evolutionary framework that makes a significant contribution towards our understanding of the molecular and functional diversification of the EGFR pathway in animals, and which will improve future comparisons between emerging biomedical systems, such as planarians, and better-established classical model organisms.



**Figure 8. Summary scenario for the evolution of the EGFR signaling pathway in Metazoa.** On the right, distribution of the identified EGFRs (yellow rectangles) and putative EGF ligands (EGF-type, green dots; NRG-type, purple dots) in the analyzed animal lineages is shown. A triplicate symbol indicates that this particular type of protein (EGFR, EGF-type and NRG-type of ligands) appears expanded in this lineage. Our findings indicate that the last common metazoan ancestor had one EGFR and one putative EGF-type ligand. The NRG-type of EGF ligand is Bilateria-specific. The ancestral bilaterian set of one EGFR, one EGF-type and one NRG-type of ligands seems to be retained in the last common ancestors of both Deuterostomia and Protostomia. Subsequently, particular bilaterian lineages have experienced expansions of one or more of these basic EGFR signaling components. Additionally, vertebrates, the limpet *L. gigantea* and the planarian *S. mediterranea* have an EGFR with an inactive tyrosine kinase domain, which suggests the independent evolution of alternative regulatory mechanisms of this signaling pathway. Tree topology based on<sup>26,59,60</sup>.

## Methods

**Database searching and phylogenetic analyses.** All potential EGFRs and EGF ligands were identified by BLAST and HMMER searches against completed genome/transcriptome databases publicly available or that are being generated in our laboratories (Supplementary Material). Searches were conducted with default parameters and an inclusive E-value of 0.05. For the EGFRs, only sequences including the tyrosine kinase domain (complete or partial) were considered. Regarding the EGF ligands, only full-length sequences with coherent domain architectures were used in subsequent analyses. The only exceptions for this rule were the previously described planarian ligand (*Smed-epiregulin-1*)<sup>30</sup>, which lacks the N-terminal region, as well as the newly identified NRG-type ligand of *X. bockii* and EGF-type ligands of *P. caudatus*, *H. spinulosus* and *S. kowalevskii*, which all lack the very C-terminal end of the protein but give high BLAST and HMMER similarities to other described EGF ligands. The retrieved sequences were aligned using MAFFT v7<sup>48</sup>. Poorly aligned regions of the multiple protein alignment of EGFRs were removed with Gblocks<sup>49</sup> using the least stringent parameters. Only the EGF domain of the candidate EGF ligands was used for phylogenetic analyses. In the case of the NRG-type ligands, also the immunoglobulin domain was included in the analyses. Maximum likelihood (ML) phylogenetic analyses were conducted with RAXML v8<sup>50</sup> with the autoMRE option on to calculate the bootstrap support values. Bayesian inference analyses were performed with MrBayes v3<sup>51</sup> using two parallel runs, sampling every 100 generations. Bayesian posterior probabilities were used for assessing the statistical support of each bipartition. The domain architecture of each identified sequence was analyzed using InterProScan 5<sup>52</sup> and SignalP 4.1<sup>53</sup>, and manually verified when automatic predictions were dubious. The domain information was used to assess the reliability of each sequence of the initial dataset and to help define protein families according to their architectural coherence.

**Clustering analysis of the potential EGF ligands.** A sequence-similarity-based (PSI-BLAST *P*-values) clustering approach to analyze the global phylogenetic relationships of the potential EGF ligands was performed with CLANS<sup>54</sup>. A cutting *P*-value of 1e-15 was applied.

**Planarian culture.** Asexual *S. mediterranea* from the BCN-10 clonal line were maintained in artificial water<sup>55</sup>. Animals were fed with veal liver and starved for at least 1 week before conducting any experiment.

**Gene cloning and whole-mount *in situ* hybridization.** The newly identified planarian EGFRs and putative EGF ligands were isolated by gene-specific PCR and cloned into the pGEM-T vector (Promega). Whole-mount *in situ* hybridization experiments were performed as previously described<sup>156,57</sup>. All samples were processed through a Leica MZ16F stereomicroscope and images were captured with a ProgResC3 camera (Jenoptik). Images were processed with Photoshop CS6 (Adobe) and figures were mounted in Illustrator CS6 (Adobe). Brightness/contrast and color balance adjustments were applied to the whole image, not parts.

**RNA interference.** Silencing by RNAi was performed as described elsewhere<sup>58</sup>. Control animals were injected with double-stranded RNA for green fluorescent protein (GFP). In *Smed-egf-6*(RNAi) experiments, all animals received one round of injection (each consisting of three injections on consecutive days). For experiments on regenerating animals, planarians were amputated 1 day after the last injection and allowed to regenerate 14 days before fixation. For homeostatic experiments dsRNA-injected animals were kept in starvation for 2 weeks before fixation. In the RNAi experiments of the other components of the EGFR pathway in *S. mediterranea* identified in this study, all animals received two rounds of injections (each consisting of three injections on consecutive days) separated by 3–4 days and were amputated 1 day after both rounds of injection. Since no phenotype was observed during regeneration with these genes, we did not proceed with homeostatic RNAi experiments.

## References

- Schlessinger, J. Receptor tyrosine kinases: legacy of the first two decades. *Cold Spring Harb Perspect Biol* **6**, doi: 10.1101/cshperspect.a008912 (2014).
- Pires-daSilva, A. & Sommer, R. J. The evolution of signalling pathways in animal development. *Nat Rev Genet* **4**, 39–49, doi: 10.1038/nrg977 (2003).
- Sibilia, M. *et al.* The epidermal growth factor receptor: from development to tumorigenesis. *Differentiation* **75**, 770–787, doi: 10.1111/j.1432-0436.2007.00238.x (2007).
- Schweitzer, R. & Shilo, B. Z. A thousand and one roles for the Drosophila EGF receptor. *Trends Genet* **13**, 191–196 (1997).
- Chang, C. & Sternberg, P. W. C. *elegans* vulval development as a model system to study the cancer biology of EGFR signaling. *Cancer Metastasis Rev* **18**, 203–213 (1999).
- Miettinen, P. J. *et al.* Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor. *Nature* **376**, 337–341, doi: 10.1038/376337a0 (1995).
- Birchmeier, C. & Birchmeier, W. Molecular aspects of mesenchymal-epithelial interactions. *Annu Rev Cell Biol* **9**, 511–540, doi: 10.1146/annurev.cb.09.110193.002455 (1993).
- Galvez-Contreras, Gonzalez-Castaneda, T, C.-O., S, L. & O, G.-P. Phenytoin enhances the phosphorylation of EGFR and FGFR in the subventricular zone and promotes the proliferation of neural precursor cells and oligodendrocyte differentiation. *The European Journal of Neuroscience* (2015).
- Wiesen, J. F., Young, P., Werb, Z. & Cunha, G. R. Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. *Development* **126**, 335–344 (1999).
- Guo, G. *et al.* Ligand-Independent EGFR Signaling. *Cancer Res* **75**, 3436–3441, doi: 10.1158/0008-5472.CAN-15-0989 (2015).
- Jorissen, R. N. *et al.* Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res* **284**, 31–53 (2003).
- Lax, I. *et al.* Epidermal growth factor (EGF) induces oligomerization of soluble, extracellular, ligand-binding domain of EGF receptor. A low resolution projection structure of the ligand-binding domain. *J Biol Chem* **266**, 13828–13833 (1991).
- Aroian, R., Koga, M., Mendel, J., Ohshima, Y. & Sternberg, P. The *let-23* gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. *Nature* **348**, 693–699 (1990).
- Shilo, B. Z. Signaling by the Drosophila epidermal growth factor receptor pathway during development. *Exp Cell Res* **284**, 140–149 (2002).
- Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2**, 127–137 (2001).
- Harris, R. C., Chung, E. & Coffey, R. J. EGF receptor ligands. *Exp Cell Res* **284**, 2–13 (2003).
- Wouters, M. A. *et al.* Evolution of distinct EGF domains with specific functions. *Protein Sci* **14**, 1091–1103, doi: 10.1110/ps.041207005 (2005).
- Hill, R. & Sternberg, P. The gene *lin-3* encodes an inductive signal for vulval development in *C. elegans*. *Nature* **358**, 470–476 (1992).
- Freeman, M. Complexity of EGF receptor signalling revealed in Drosophila. *Curr Opin Genet Dev* **8**, 407–411 (1998).
- Massague, J. & Pandiella, A. Membrane-anchored growth factors. *Annu Rev Biochem* **62**, 515–541, doi: 10.1146/annurev.bi.62.070193.002503 (1993).
- Falls, D. L. Neuregulins: functions, forms, and signaling strategies. *Exp Cell Res* **284**, 14–30 (2003).
- Austin, C. L., Manivannan, S. N. & Simcox, A. TGF- $\alpha$  ligands can substitute for the neuregulin *Vein* in Drosophila development. *Development* **141**, 4110–4114, doi: 10.1242/dev.110171 (2014).
- Fraguas, S., Barberán, S. & Cebrià, F. EGFR signaling regulates cell proliferation, differentiation and morphogenesis during planarian regeneration and homeostasis. *Dev Biol* **354**, 87–101, doi: 10.1016/j.ydbio.2011.03.023 (2011).
- Fraguas, S., Barberán, S., Iglesias, M., Rodríguez-Esteban, G. & Cebrià, F. *egr-4*, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians. *Development* **141**, 1835–1847, doi: 10.1242/dev.101345 (2014).
- Rink, J. C., Vu, H. T. & Sánchez Alvarado, A. The maintenance and regeneration of the planarian excretory system are regulated by EGFR signaling. *Development* **138**, 3769–3780, doi: 10.1242/dev.066852 (2011).
- Dunn, C., Giribet, G., Edgecombe, G. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu Rev Ecol Evol Syst* **45**, 371–395 (2014).
- Handberg-Thorsager, M., Fernandez, E. & Salo, E. Stem cells and regeneration in planarians. *Front Biosci* **13**, 6374–6394 (2008).
- Spiliotis, M., Kroner, A. & Brehm, K. Identification, molecular characterization and expression of the gene encoding the epidermal growth factor receptor orthologue from the fox-tapeworm *Echinococcus multilocularis*. *Gene* **323**, 57–65 (2003).
- Ramachandran, H., Skelly, P. J. & Shoemaker, C. B. The *Schistosoma mansoni* epidermal growth factor receptor homologue, SER, has tyrosine kinase activity and is localized in adult muscle. *Mol Biochem Parasitol* **83**, 1–10 (1996).

30. Wenemoser, D., Lapan, S. W., Wilkinson, A. W., Bell, G. W. & Reddien, P. W. A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev* **26**, 988–1002, doi: 10.1101/gad.187377.112 (2012).
31. Barberán, S., Fraguas, S. & Cebrià, F. The EGFR signaling pathway controls gut progenitor differentiation during planarian regeneration and homeostasis. *Development*, doi: 10.1242/dev.131995 (2016).
32. Suga, H. *et al.* Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci Signal* **5**, ra35, doi: 10.1126/scisignal.2002733 (2012).
33. Srivastava, M. *et al.* The Trichoplax genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
34. D'Aniello, S. *et al.* Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Mol Biol Evol* **25**, 1841–1854 (2008).
35. Bogdan, S. & Klambt, C. Epidermal growth factor receptor signaling. *Curr Biol* **11**, R292–295 (2001).
36. Citri, A., Skaria, K. B. & Yarden, Y. The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp Cell Res* **284**, 54–65 (2003).
37. Guy, P. M., Platko, J. V., Cantley, L. C., Cerione, R. A. & Carraway, K. L., 3rd. Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proc Natl Acad Sci USA* **91**, 8132–8136 (1994).
38. Sierke, S. L., Cheng, K., Kim, H. H. & Koland, J. G. Biochemical characterization of the protein tyrosine kinase homology domain of the ErbB3 (HER3) receptor protein. *Biochem J* **322** (Pt 3), 757–763 (1997).
39. Shi, F., Telesco, S. E., Liu, Y., Radhakrishnan, R. & Lemmon, M. A. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc Natl Acad Sci USA* **107**, 7692–7697, doi: 10.1073/pnas.1002753107 (2010).
40. Robb, S. M., Gotting, K., Ross, E. & Sanchez Alvarado, A. SmedGD 2.0: The *Schmidtea mediterranea* genome database. *Genesis* **53**, 535–546, doi: 10.1002/dvg.22872 (2015).
41. Brandl, H. *et al.* PlanMine - a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res* **44**, D764–773, doi: 10.1093/nar/gkv1148 (2016).
42. Martin-Duran, J. M., Ryan, J. F., Cossermelli Vellutini, B. C., Pang, K. & Hejnal, A. A novel approach using increased taxon sampling reveals thousands of hidden orthologs in flatworms. *bioRxiv*, doi: 10.1101/050724 (2016).
43. Stein, R. A. & Staros, J. V. Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol Biol* **6**, 79, doi: 10.1186/1471-2148-6-79 (2006).
44. Lainsney, J. A., Braasch, I., Walter, R. B., Meierjohann, S. & Schartl, M. Lineage-specific co-evolution of the Egf receptor/ligand signaling system. *BMC Evol Biol* **10**, 27, doi: 10.1186/1471-2148-10-27 (2010).
45. Stein, R. A. & Staros, J. V. Evolutionary analysis of the ErbB receptor and ligand families. *J Mol Evol* **50**, 397–412 (2000).
46. Adler, C. E., Seidel, C. W., McKinney, S. A. & Sanchez Alvarado, A. Selective amputation of the pharynx identifies a FoxA-dependent regeneration program in planaria. *Elife* **3**, e02238, doi: 10.7554/eLife.02238 (2014).
47. Martín-Durán, J. M., Amaya, E. & Romero, R. Germ layer specification and axial patterning in the embryonic development of the freshwater planarian *Schmidtea polychroa*. *Dev Biol* **340**, 145–158, doi: 10.1016/j.ydbio.2010.01.018 (2010).
48. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780, doi: 10.1093/molbev/mst010 (2013).
49. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577, doi: 10.1080/10635150701472164 (2007).
50. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi: 10.1093/bioinformatics/btu033 (2014).
51. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
52. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240, doi: 10.1093/bioinformatics/btu031 (2014).
53. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786, doi: 10.1038/nmeth.1701 (2011).
54. Frickey, T. & Lupas, A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704, doi: 10.1093/bioinformatics/bth444 (2004).
55. Cebrià, F. & Newmark, P. A. Planarian homologs of *netrin* and *netrin receptor* are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development* **132**, 3691–3703, doi: 10.1242/dev.01941 (2005).
56. Umesono, Y., Watanabe, K. & Agata, K. A planarian orthopedia homolog is specifically expressed in the branch region of both the mature and regenerating brain. *Dev Growth Differ* **39**, 723–727 (1997).
57. Molina, M. D., Salo, E. & Cebrià, F. The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Dev Biol* **311**, 79–94, doi: 10.1016/j.ydbio.2007.08.019 (2007).
58. Sanchez Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc Natl Acad Sci USA* **96**, 5049–5054 (1999).
59. Laumer, C. E. *et al.* Spiralian phylogeny informs the evolution of microscopic lineages. *Curr Biol* **25**, 2000–2006, doi: 10.1016/j.cub.2015.06.068 (2015).
60. Borner, J., Rehm, P., Schill, R. O., Ebersberger, I. & Burmester, T. A transcriptome approach to ecdysozoan phylogeny. *Mol Phylogenet Evol* **80**, 79–87, doi: 10.1016/j.ympev.2014.08.001 (2014).

## Acknowledgements

We thank Andreas Hejnal and all the members of Emili Saló's lab for support and discussion, as well as D. Thiel for help with the CLAN analysis. We thank Gemma S. Richards for a critical read of the manuscript. All identified sequences are available in Additional File 7. Additionally, *S. mediterranea* sequences were deposited in GenBank (accession numbers *egf-1* to *egf-7*: KU850488 to KU850494; *egf-8*: KT947468; *nrg-1*: KU850495; *egfr-1* to *egfr-6*: KU850496 to KU850501). This study was funded by grants, BFU2012-31701 and BFU2015-65704-P (Ministerio de Economía y Competitividad/Feder (Spain)) and 2009-SGR-1018 from the Generalitat de Catalunya to F.C. SB was supported by a FI fellowship from the Generalitat de Catalunya and a collaboration fellowship from IBUB. JMMD was supported by Marie Curie IEF 329024 fellowship and Sars core budget. The authors declare no competing financial interests.

## Author Contributions

All authors conceived and designed the study. S.B. and J.M.M.-D. conducted the experiments and drafted the manuscript. All authors discussed and commented on the data. F.C. edited the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Barberán, S. *et al.* Evolution of the EGFR pathway in Metazoa and its diversification in the planarian *Schmidtea mediterranea*. *Sci. Rep.* **6**, 28071; doi: 10.1038/srep28071 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>