Convergent Acquisition of Nonembryonic Development in Styelid Ascidians

Alexandre Alié, Laurel Sky Hiebert, Paul Simion, Marta Scelzo, Maria Mandela Prünster, Sonia Lotito, Frédéric Delsuc, Emmanuel J.P. Douzery, Christelle Dantec, Patrick Lemaire, et al.

To cite this version:
Alexandre Alié, Laurel Sky Hiebert, Paul Simion, Marta Scelzo, Maria Mandela Prünster, et al.. Convergent Acquisition of Nonembryonic Development in Styelid Ascidians. Molecular Biology and Evolution, Oxford University Press (OUP), 2018, 35 (7), pp.1728 - 1743. 10.1093/molbev/msy068. hal-01879150

HAL Id: hal-01879150
https://hal-sde.archives-ouvertes.fr/hal-01879150
Submitted on 22 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Convergent Acquisition of Nonembryonic Development in Styelid Ascidians

Alexandre Alié,1, f, 1 Laurel Sky Hiebert, f, 2, 3 Paul Simion, f, 1 Marta Scelzo, 1 Maria Mandela Prünster, 1 Sonia Lotito, 1 Frédéric Delsuc, 4 Emmanuel J.P. Douzery, 4 Christelle Dantec, 5 Patrick Lemaire, 5 Sébastien Darras, 6 Kazuo Kawamura, 7 Federico D. Brown, 2, 3, 8 and Stefano Tiozzo*, 1

1 Sorbonne Université, CNRS, Laboratoire de Biologie du Développement de Villefranche-sur-mer (LBDV), 06230 Paris, France
2 Departamento de Zoologia – Instituto Biociências, Universidade de São Paulo, São Paulo, Brazil
3 Centro de Biologia Marina (CEBIMar), Universidade de São Paulo, São Paulo, Brazil
4 ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France
5 CRBM, Université de Montpellier, CNRS, Montpellier, France
6 Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, Banyuls/Mer, 06230 Paris, France
7 Laboratory of Cellular and Molecular Biotechnology, Faculty of Science, Kochi University, Kochi, Japan
8 Instituto Nacional de Ciência e Tecnologia em Estudos Interdisciplinares e Transdisciplinares em Ecologia e Evolução (IN-TREE), Salvador, Brazil

† These authors contributed equally to this work.
* Corresponding author: E-mail: tiozzo@obs-vlfr.fr.

Associate editor: Nicole King

Abstract

Asexual propagation and whole body regeneration are forms of nonembryonic development (NED) widespread across animal phyla and central in life history and evolutionary diversification of metazoans. Whereas it is challenging to reconstruct the gains or losses of NED at large phylogenetic scale, comparative studies could benefit from being conducted at more restricted taxonomic scale, in groups for which phylogenetic relationships are well established. The ascidian family of Styelidae encompasses strictly sexually reproducing solitary forms as well as colonial species that combine sexual reproduction with different forms of NED. To date, the phylogenetic relationships between colonial and solitary styelids remain controversial and so is the pattern of NED evolution. In this study, we built an original pipeline to combine eight genomes with 18 de novo assembled transcriptomes and constructed data sets of unambiguously orthologous genes. Using a phylogenomic super-matrix of 4,908 genes from these 26 tunicates we provided a robust phylogeny of this family of chordates, which supports two convergent acquisitions of NED. This result prompted us to further describe the budding process in the species Polyandrocarpa zorritensis, leading to the discovery of a novel mechanism of asexual development. Whereas the pipeline and the data sets produced can be used for further phylogenetic reconstructions in tunicates, the phylogeny provided here sets an evolutionary framework for future experimental studies on the emergence and disappearance of complex characters such as asexual propagation and whole body regeneration.

Key words: asexual reproduction, Botryllinae, cross-contamination, Polyandrocarpa, regeneration, tunicate.

Introduction

Most animals reproduce sexually (White 1978). Male and female gametes combine their genetic material in a zygote, develop by embryogenesis, and produce individuals that are genetically different from their parents. This reproductive strategy is regarded as evolutionarily advantageous by allowing rapid adaptation to constantly changing environments; for example, it can efficiently eliminate deleterious mutations or rapidly incorporate beneficial mutations (reviewed by West et al. 1999). However, other reproductive modes coexist within the metazoans. Nearly half of all animal phyla contain species that are able to propagate asexually via agamic reproduction (i.e., cloning), and the number is higher if one includes parthenogenetic species (Bell 1982; Hughes and Cancino 1985; Nilsson Sköld et al. 2009; Brusca et al. 2016). Organisms that reproduce clonally mobilize somatic cells, either as undifferentiated progenitors or terminally differentiated cells, to produce new individuals genetically identical to their parent. Agametic asexual reproduction is often linked with extensive regenerative capacity, as many clonally propagating organisms can regenerate their body entirely, including the germline (Hughes 1989; Sanchez-Alvarado and Yamanaka 2014). This relationship strongly suggests that clonal propagation and regeneration are two
processes of “nonembryonic development” (NED) that involve common cellular and molecular mechanisms (Martinez et al. 2005; Agata et al. 2007; Bosch 2007; Galliott and Chera 2010; Gurlay et al. 2010).

Some animals undergoing NED can form colonies, where individuals remain either physically connected to each other or develop clusters of genetically identical modules termed zooids (e.g., in tunicates), ramets (e.g., in bryozoans), or polyps (e.g., in cnidarians; Boardman et al. 1973). Depending on the species, the modules can be either morphologically identical or different, so that physiological functions, like reproduction or feeding can be subdivided among all units of the colony. This type of growth and modular organization has important physiological and ecological consequences, such as allowing competitive dominance over solitary forms on space-limited hard substrata (Jackson 1977) or reducing gene flow compared with solitary species (Jackson and Coates 1986). Solitary organisms show an inverse relationship between body size and reproductive rate due to reduced metabolic rate with high body mass. However, colonial species escape this limitation and show mass-specific metabolic rates related to module size rather than colony size. Thus, colonial species can grow faster than solitary species and propagate indefinitely without loss in reproductive potential (Hughes 1989; Jackson and Coates 1986; Burgess et al. 2017). Therefore, widespread sexual reproduction by gamete production only represents one facet of species perpetuation and evolutionary diversification of metazoans.

Species employing asexual clonal reproduction are distributed across animal phyla, including nonbilaterian animals such as placozoans, cnidarians or sponges, many spiralian groups such as bryozoans, annelids or planarians, and deuterostome phyla such as hemichordates, echinoderms or tunicates, with the notable exception of vertebrates (Hughes 1989; Bely and Nyberg 2010; Brusca et al. 2016; Alié et al. 2015; Funayama et al. 2010). Regenerative abilities are also widespread and unevenly distributed across metazoans (reviewed in Tiozzo and Copley 2015). While many authors consider regeneration an ancestral trait for animals (Sánchez Alvarado 2000; Bely and Nyberg 2010), the more scattered phylogenetic distribution and diversity of agametic asexual reproductive mechanisms is suggestive of multiple independent acquisitions (Nilsson Sköld et al. 2009). For example, in annelids asexual reproduction by fission was gained at least 19 times (Zattara and Bely 2016). In anthozoan cnidarians fission has likely arisen from an aclonal ancestor at least four times (Barbeitos et al. 2010). It is difficult to reconstruct the cellular and molecular mechanisms underlying the evolution of NED by comparing distant phyla, first because in many cases phylogenetic relationships are unresolved, and because, the cell types involved and their behaviors are hardly comparable at such phylogenetic scales (Carlson 2007). Therefore, in order to understand how novel reproductive and regenerative strategies evolved and the directionality of these transitions (i.e., gains or losses) it is essential to compare closely related species undergoing NED or lacking such traits with well-defined phylogenetic relationships (Tiozzo and Copley 2015; Nogueira et al. 2016).

Ascidians belong to the subphylum Tunicata, the closest living relatives to vertebrates (Delsuc et al. 2006) in which it is possible to find strictly sexually reproducing species and others that combine sexual reproduction with different forms of budding and whole body regeneration (Kürm et al. 2011; Brown and Swalla 2012; Piette and Lemaire 2015). Whereas solitary ascidians reproduce only sexually, colonial species can develop both sexually and via NED. Interestingly, colonial and solitary species are scattered across the phylogenetic tree of ascidians, suggesting that NED, and therefore coloniality, has been acquired and/or lost several times independently (Brown and Swalla 2012). Ascidians are polyphyletic and classically divided in three orders—Aplousobranchia, Phlebobranchia, and Stolidobranchia—in each of which budding mechanisms are very different, further supporting convergent acquisitions of coloniality (Brown and Swalla 2012). The order Stolidobranchia comprises three traditionally recognized families (Molgulidae, Pyuridae, and Styelidae). Molgulidae and Pyuridae are solitary, whereas approximately one third of Styelidae are colonial species (WoRMS Editorial Board 2017; fig. 1A–F). Thus, Stolidobranchia has long been considered a suitable group to study the plastic evolution of NED and coloniality (Zeng et al. 2006; Pérez-Portela et al. 2009; Brown and Swalla 2012).

In colonial styelids, two budding modes have been described and sometimes coexist in the same species: peribranchial budding and vascular budding (fig. 1G–H). In peribranchial budding, a portion of the zooid epithelium lining the peribranchial chamber starts to bud in a stereotyped process (fig. 1G) and gives rise to new adult zooids (Selys-Longchamps 1917; Berrill 1935; Kawamura and Watanabe 1981; Fujiwara et al. 2011; Manni et al. 2014). In addition to peribranchial budding, several species also use vascular budding either for asexual propagation or in response to injury (fig. 1H). During this process circulatory cells cluster, proliferate, and develop into a new individual within the vasculature of the colony (Kawamura and Sunanaga 2010; Rinkevich et al. 2010; Ricci et al. 2016). In both processes, the surrounding epidermis envelopes the bud, forming a double vesicle of monolayered tissue (fig. 1G and H) that will later undergo organogenesis. Therefore, in Styelidae the two described budding processes pass through a similar stage, the double vesicle, which has different cellular origins and is triggered by different mechanisms. Whereas peribranchial budding mainly involves epithelial morphogenesis, vascular budding is ensured by circulating putative stem cells (Laird et al. 2005).

Previous attempts to reconstruct the phylogeny of the Styelidae have not reached a consensus on the relationships between colonial versus solitary species. Kott (1985) splits the styelids into three subfamilies, solitary Styelinae, colonial Polyzoinae, and colonial Botryllinae based on morphological characters. However, molecular phylogenies based on ribosomal and mitochondrial markers have led to conflicting results, either retrieving all colonial species together in a single clade and suggesting a unique acquisition of asexual...
reproduction (Zeng et al. 2006; Tsagkogeorga et al. 2009; Shenkar et al. 2016), or retrieving several lineages of colonial species and thus proposing multiple convergent acquisitions of asexual reproduction (Pérez-Portela et al. 2009). Whereas the former studies included only a limited number of colonial styelids, Pérez-Portela et al. (2009) included more of them but the key branches were not statistically supported.

In the present work, we used Illumina-based RNA-seq comparisons to address phylogenetic relationships between colonial and solitary styelids. We used an original pipeline to combine de novo assembled transcriptomes of 18 stolidobranchs with 8 ascidian genomes to build a supermatrix of orthologous genes, which lead to a maximally supported tree topology. Our results showed that NED was acquired at least twice independently in the family Styelidae and led us to uncover a novel budding mechanism in Polyandrocarpa zurritensis. In light of these results, we discuss the evolution of asexual reproduction and coloniality emphasizing the evolutionary plasticity of asexual development in tunicates.

**Results**

**Species Selection from 18S rRNA-COI Phylogeny**

In order to a priori select species that span the diversity of colonial and solitary styelids, we first conducted a preliminary phylogenetic analysis based on a concatenated alignment of 18S rRNA and mitochondrial COI markers, combining data from Pérez-Portela et al. (2009), publicly available sequences, and sequences obtained from 12 specimens that we collected (see Materials and Methods and supplementary fig. 1, Supplementary Material online). Whereas most of the deep nodes were not statistically supported, the topology of the resulting tree was helpful to guide a posteriori selection of the colonial and solitary species to be used in the study. We used colonial ascidians from all different clades recovered in the phylogeny, as well as solitary species of Styela, Polycarpa, and Asterocarpa that were located in different phylogenetic lineages (supplementary fig. 1, Supplementary Material online).

In order to test the monophyly of styelids, we chose an outgroup that is a priori paraphyletic, including four Pyuridae, two Molgulidae, and four other distantly related species belonging to the genera Ciona and Phallusia (order Phlebobranchia). The final list of selected species is given in supplementary table 1, Supplementary Material online. Outgroup species of choice were mainly selected by the availability of the full genome sequence, which later simplified the steps of data set construction (supplementary table 1, Supplementary Material online).

**A New Pipeline for the De Novo Assembly of Ascidian Transcriptomes**

We generated cDNA libraries from the 16 species we collected, from most of which 30–50 million reads were sequenced using the Illumina HiSeq 4000 platform (table 1). The number of sequence reads, accession numbers, and other basic statistics to assess the quality of the transcriptomes are shown in table 1. The initial assembly of these reads using Trinity software (Grabherr et al. 2011) led to transcriptomes composed of between 76,825 and 410,479 contigs depending...
on the species (supplementary table 1, Supplementary Material online). These values are far higher than expected from current knowledge of ascidian genome complexity (Dehal et al. 2002; Voskoboynik et al. 2013; Velandia-Huerto et al. 2016). We therefore sought to eliminate redundant and low-quality contigs by developing a custom pipeline (detailed in fig. 2A and Materials and Methods). Briefly, we used stringent parameters to filter reads, contigs and ORFs based on sequence quality, redundancy, length, and expression levels. We additionally detected and removed cross-contaminations between samples that were sequenced in the same batch using CroCo (Simion et al. 2018). An average of 5% of all filtered contigs were detected as cross-contaminants (supplementary table 1, Supplementary Material online), and most cross-contaminations arose between samples sequenced in the same Illumina run (supplementary fig. 2 and supplementary table 2, Supplementary Material online).

On average, our pipeline retained ~20% of initial contigs (supplementary table 1, Supplementary Material online). BUSCO indices ranged from 87% and 96% (except for Eusynstyela tincta for which it was 74%), showing that our pipeline did not massively reduce gene numbers but rather eliminated undesired redundancy and/or contaminations. As an example, whereas Trinity assembly of reads from Polycarpa aurata gave 118,198 contigs with BUSCO index of 95% we ended up with 29,348 contigs. Final transcriptomes are available at https://github.com/AlexAlie/styelidae, last accessed April 2018, and raw Illumina reads can be retrieved from NCBI BioProject PRJNA422120.

A New Data Set of Orthologous Genes for Ascidian Phylogeny

Combining our refined de novo transcriptome assemblies with sets of genes derived from genomes allowed us to build a high quality data set of orthologous tunicate genes (fig. 2). The contigs of our 18 transcriptomes (fig. 2A) were first attributed to 14,552 clusters that were built using 18 chordate genomes (fig. 2B, supplementary table 3, Supplementary Material online). Among them we selected 8,187 clusters with no apparent paralogy in tunicate genomes (fig. 2C). After removing all nontunicate species, the contigs within each cluster were aligned (fig. 2D) and we performed alignment curation by joining end-to-end fragments of single transcripts that were split due to imperfect transcriptome assembly (fig. 2D, step “fusion of nonoverlapping sequences”). The combination of high quality transcriptomes, fusion of fragmented sequences, and subsequent filtering (see Materials and Methods) led to the retention of 4,908 alignments devoid of apparent paralogy in tunicates (fig. 2D, TOTAL data set of 1,740,663 amino acid sites and 36.5% missing data). A reduced version of this data set comprising 1,306 genes was selected, in which virtually all species were present (fig. 2D, REDUCED data set of 605,265 amino acid sites and 14.4% missing data). Functional annotations showed that the genes belong to different families, functional groups, and have different kinds of protein domains (supplementary table 4, Supplementary Material online). Both TOTAL and REDUCED matrices are provided at https://github.com/AlexAlie/styelidae, last accessed April 2018.

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Number</th>
<th>Sampling Location</th>
<th>Tissues Used for De Novo Transcriptome</th>
<th>n Raw Reads</th>
<th>n Contigs</th>
<th>N50</th>
<th>Average Length</th>
<th>Busco Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asterocarpa humilis</td>
<td>PRJNA422120</td>
<td>Brest (France)</td>
<td>1 adult (no tunic, no stomach)</td>
<td>39,963,596</td>
<td>24,341</td>
<td>1,664</td>
<td>1,180.46</td>
<td>88%</td>
</tr>
<tr>
<td>Botryllodes leachii</td>
<td>SAMN04161391-98</td>
<td>Otago (New Zealand)</td>
<td>Embryos and regenerative tissues</td>
<td>315,587,750</td>
<td>42,036</td>
<td>1,365</td>
<td>903</td>
<td>96%</td>
</tr>
<tr>
<td>Botryllus schlosseri</td>
<td>SRX726446-59,</td>
<td>Santa Barbara (USA)</td>
<td>Whole colonies</td>
<td>621,300,622</td>
<td>65,586</td>
<td>1,224</td>
<td>905</td>
<td>95%</td>
</tr>
<tr>
<td>Dendrodo grossularia</td>
<td>PRJNA422120</td>
<td>Roscoff (France)</td>
<td>1 adult + larvae (no tunic, no stomach)</td>
<td>42,386,509</td>
<td>57,564</td>
<td>1,194</td>
<td>852.74</td>
<td>95%</td>
</tr>
<tr>
<td>Distomus variolosus</td>
<td>PRJNA422120</td>
<td>Roscoff (France)</td>
<td>1 adult (no tunic)</td>
<td>39,810,612</td>
<td>44,516</td>
<td>1,458</td>
<td>1,004.97</td>
<td>92%</td>
</tr>
<tr>
<td>Eusynstyela tincta</td>
<td>PRJNA422120</td>
<td>Sao Sebastiao (Brazil)</td>
<td>Siphon</td>
<td>33,225,966</td>
<td>39,494</td>
<td>837</td>
<td>694</td>
<td>74%</td>
</tr>
<tr>
<td>Microcosmus sabatieri</td>
<td>PRJNA422120</td>
<td>Banyuls (France)</td>
<td>1 adult (no tunic)</td>
<td>29,425,964</td>
<td>22,951</td>
<td>1,471</td>
<td>1,022.8</td>
<td>93%</td>
</tr>
<tr>
<td>Polyandrocarpa misakiensis</td>
<td>PRJNA422120</td>
<td>Misaki (Japan)</td>
<td>Asexual zooids and buds</td>
<td>39,178,452</td>
<td>52,545</td>
<td>1,194</td>
<td>851.48</td>
<td>96%</td>
</tr>
<tr>
<td>Polyandrocarpa zorrantis</td>
<td>PRJNA422120</td>
<td>La Spezia (Italy)</td>
<td>2 adults and buds</td>
<td>158,133,887</td>
<td>57,781</td>
<td>1,197</td>
<td>816.91</td>
<td>96%</td>
</tr>
<tr>
<td>Polycarpa aurata</td>
<td>PRJNA422120</td>
<td>Cebu (Philippines)</td>
<td>1 adult (no tunic, no stomach)</td>
<td>34,876,779</td>
<td>29,348</td>
<td>1,443</td>
<td>982.17</td>
<td>95%</td>
</tr>
<tr>
<td>Polycarpa mamillaris</td>
<td>PRJNA422120</td>
<td>Banyuls (France)</td>
<td>Oocytes</td>
<td>12,985,467</td>
<td>25,221</td>
<td>1,599</td>
<td>1,056.42</td>
<td>96%</td>
</tr>
<tr>
<td>Polycarpa pomaria</td>
<td>PRJNA422120</td>
<td>Brest (France)</td>
<td>1 adult (no tunic)</td>
<td>33,554,939</td>
<td>41,730</td>
<td>1,386</td>
<td>949.48</td>
<td>95%</td>
</tr>
<tr>
<td>Pyura dura</td>
<td>PRJNA422120</td>
<td>Banyuls (France)</td>
<td>Gonads</td>
<td>33,619,672</td>
<td>17,739</td>
<td>1,854</td>
<td>1,216.9</td>
<td>96%</td>
</tr>
<tr>
<td>Stolonica socialis</td>
<td>PRJNA422120</td>
<td>Roscoff (France)</td>
<td>1 adult</td>
<td>35,870,064</td>
<td>43,474</td>
<td>1,431</td>
<td>934.34</td>
<td>94%</td>
</tr>
<tr>
<td>Styela canopus</td>
<td>PRJNA422120</td>
<td>Sao Sebastiao (Brazil)</td>
<td>1 adult (mantle)</td>
<td>44,826,592</td>
<td>42,420</td>
<td>1,047</td>
<td>815.45</td>
<td>87%</td>
</tr>
<tr>
<td>Styela clava</td>
<td>PRJNA422120</td>
<td>Brest (France)</td>
<td>1 adult (no tunic, no stomach)</td>
<td>47,686,633</td>
<td>26,553</td>
<td>1,647</td>
<td>1,164.48</td>
<td>89%</td>
</tr>
<tr>
<td>Styela limata</td>
<td>PRJNA422120</td>
<td>Banyuls (France)</td>
<td>Embryos</td>
<td>38,186,700</td>
<td>19,288</td>
<td>1,870</td>
<td>1,253.51</td>
<td>97%</td>
</tr>
<tr>
<td>Polycarpa sp.</td>
<td>PRJNA422120</td>
<td>Roscoff (France)</td>
<td>1 adult (no tunic)</td>
<td>43,522,115</td>
<td>46,672</td>
<td>1,311</td>
<td>918.30</td>
<td>90%</td>
</tr>
</tbody>
</table>
A Robust Phylogeny of Styelid Ascidians

We conducted maximum likelihood analyses using models with (LG4X) and without (LG) site heterogeneity, as well as Bayesian inference under the site-heterogeneous CAT mixture model. All three strategies generated the same topology, with maximum bootstrap and maximum posterior probabilities obtained for all branches (fig. 3 and supplementary figs. 3–5, Supplementary Material online). We retrieved a...
monophyletic Stolidobranchia, in which a monophyletic Styelidae was embedded within a paraphyletic Pyuridae, and in which a monophyletic Molgulidae was the sister-group to all other Stolidobranchia. Within Styelidae, a clade of solitary *Styela* and *Asterocarpa* was the sister-group to all the rest (fig. 3). Colonial species, that is, species adopting NED, fell in two distinct locations in our inferred phylogeny. First, we recovered a clade containing *Botryllus schlosseri*, *Botrylloides leachii*, *Polyandrocarpa misakensis*, *Eusynstyela tincta*, *Distomus variolosus*, and *Stolonica socialis* (fig. 3). In all species belonging to this clade, colonial growth occurs by peribranchial budding, and therefore we named this clade the peribranchial budding clade (PB clade). In this clade, *B. schlosseri* and *B. leachii* (Botryllinae) are also able to regenerate colonies through vascular budding (fig. 3). The Botryllinae was the sister group to a clade containing four Polyzoinae: *E. tincta* together with *P. misakensis* and *S. socialis* together with *D. variolosus*. The fifth Polyzoinae and last colonial species of our sampling, *P. zorritensis* grouped together with solitary *Polycarpa* and *Dendrodoa grossularia*, as sister to *P. aurata* (fig. 3). Therefore, the subfamily Polyzoinae appeared as polyphyletic, with *P. zorritensis* being more closely related to solitary species than to other colonial styelids. The *Polycarpa* genus was also polyphyletic, with *P. aurata* being separated from other *Polycarpa* species. The distant phylogenetic position of *P. zorritensis* apart from the PB clade led us to further describe budding in this species.

**Novel Budding in *P. zorritensis***

In a previous report, budding of *P. zorritensis* was described as vascular, involving "blood vessel wall and lymphocytes" (Brunetti and Mastrototaro 2004), but this was based on in vivo observations only. To gain more insights into the cellular processes leading to budding, we conducted histological observations of *P. zorritensis* budding. We observed that the budding process in this species differed from those known in other ascidians. After a few days in culture, the adult zooid emits stolons (fig. 4A, B) that consist of a single blood vessel bearing short blind ramifications, named ampullae, and surrounded by a thin layer of tunic (fig. 4B). Budding starts at several spots along each stolon, where ampullae become very numerous, short, and oriented in every direction, taking a flower-shape (fig. 4C). The growing bud appears in the center.
of these ramifications in form of a round vesicle (fig. 4D) and later develops into a new functional zooid (fig. 4E). Histological sections at successive stages showed that the vascular epidermis of the central stolon invaginates (fig. 4F), until invagination borders come into contact and fuse with each other (fig. 4G). This process results in the formation of an inner vesicle, separated from the outer vesicle by mesenchymal space. A few hours later, the inner vesicle shows distinct foldings (fig. 4H) and later organ primordia become visible. Figure 4I summarizes the budding process until the double vesicle stage.

Expression of NK4 Delineates the Prospective Inner Vesicle in B. schlosseri and P. zorritensis

In order to find molecular markers of budding tissues across stylids, we studied the expression of seven candidate genes based on known expression in Botryllus (supplementary table 6, Supplementary Material online) and identified the homeobox containing NK4, having a similar expression pattern in the prospective territory of the future inner vesicle in both P. zorritensis and B. schlosseri (fig. 5). NK4 is the name given to the ascidian orthologue of NkX2.5/2.6/Tinman family of homeobox cardiac determinants (fig. 5A, Wang et al. 2013).
During peribranchial budding in *B. schlosseri* this gene was first expressed in the thickened, evaginating peribranchial epithelium, neither beyond this region of budding nor in the surrounding epidermis (fig. 5B, C″ and supplementary fig. 6, Supplementary Material online). When the inner vesicle was fully formed, NK4 became restricted to the anterior-right side of the budlet (fig. 5B). Similarly, during the earliest stages of vascular budding in *B. schlosseri*, NK4 was specifically
expressed in the growing inner vesicle (fig. 5D–F). NK4 was not detected either in circulating cells or in the vascular epidermis of the ampullae (fig. 5D), but because the tunic stains nonspecifically, we cannot rule out expression in the epidermal outer vesicle of the budlet. In *P. zorritensis* the earliest detected expression of NK4 began during epidermal invagination, and expanded throughout the whole invaginating region but not in the noninvaginating epithelium (fig. 5E). The expression persisted until the invagination borders fused to each other (fig. 5F). Thus, NK4 expression defines the onset of the budding epithelia in both *B. schlosseri* and *P. zorritensis* in three different budding modes.

**Discussion**

In this study, we generated 18 new transcriptomes of solitary and colonial ascidians belonging to the Styelidae family. We designed a bioinformatic pipeline that allowed us to conduct a phylogenomic analysis from heterogeneous sources of samples, including genomic and transcriptomic data. We provided a robust phylogeny of a family of chordates, which permitted us to trace the evolution of a complex character, that is, NED, and to highlight a previously undescribed mechanism of asexual development in colonial ascidians.

**A Reference Set of Orthologous Genes for Tunicate Phylogeny**

De novo transcriptome assembly often yields a high number of redundant contigs due to assembly artifacts, but also because of heterozygosity, isoforms, and polymorphism generated when several individuals are pooled together for RNA extraction (Cahais et al. 2012). Contaminations from other organisms, such as epibionts or micro and macroorganisms present in the ascidian pharynx, as well as poorly assembled rare transcripts, may also participate to increase the number of undesired contigs. In addition, some colonial ascidians form chimeric colonies (Kassmer et al. 2015), which adds another source of polymorphism to the transcriptome assembly. Reducing such redundancy and eliminating contaminations have proven essential for downstream analyses such as gene expression, functional enrichment, and orthology assessment (Davidson and Oshlack 2014; Philippe et al. 2017). In this paper, we designed an original procedure that was aimed at increasing transcriptome quality. The importance of good quality transcriptome assemblies only now is starting to be recognized (Cahais et al. 2012; Cabau et al. 2017) and therefore our new approach represents a straightforward solution to this problem with light computational demands. The merging step using CD-HIT-EST (fig. 2A) led to a drastic reduction of contig numbers (supplementary table 1, Supplementary Material online), showing that part of the redundancy comes from heterozygosity, assembly artifacts, or polymorphism of colonial species. Further drastic reduction of contig numbers was obtained from ORF selection, showing that many contigs assembled by Trinity had small ORF size (smaller than our threshold of 90 aa), possibly due to the inclusion of rare transcripts that split into multiple small ORFs.

Cross-contaminations in RNA-seq experiments are known to arise when RNA molecules from one sample accidentally pass to another sample, leading to a misattribution of sequencing reads to the contaminated sample (Ballenghien et al. 2017). Because of the Illumina sequencing depth, even extremely minor contamination events can be sequenced and result in unwanted assembled transcripts (Ballenghien et al. 2017; Simion et al. 2018). The level of cross-contamination was important in our samples, ranging between 0.07% and 24% of the contigs, depending on the species. Therefore removing cross-contaminants using recently developed CroCo software (Simion et al. 2018) turned out to be a crucial step. Interestingly, the level of cross-contamination was of the same order of magnitude regardless of whether RNA was extracted in the same or different laboratories, but it was much higher between samples sequenced on the same Illumina lane than between samples sequenced on different lanes performed at different dates (supplementary fig. 2, Supplementary Material online). These observations show that, even if all care is taken during tissue collection and RNA extractions, cross-contaminations mainly arise at the sequencing platform, when samples are handled for library preparation or Illumina sequencing. It is important to note that the amount of cross contaminants produced by sequencing facilities is drastically variable, as we observed an average of 0.2% of cross contaminants for samples handled by one company and an average of 7.3% of cross contaminants for samples handled by a different company (supplementary table 1, Supplementary Material online). Our study represents a typical case that highlights a recurrent problem that cannot be underestimated (Ballenghien et al. 2017; Simion et al. 2018), as it could produce artifacts not only for phylogenetic reconstruction, but also in any RNA-seq study.

Phylogenomics is believed to outperform single-gene phylogeny by adding a high number of genes that will supposedly convey sufficient phylogenetic signal (Philippe et al. 2011). However, if the genes selected for the data matrix are not conscientiously curated, a large amount of signal noise from sequences that deviate from the species tree can result in a fallacious phylogeny (Philippe et al. 2011). Careful selection of orthologous genes is therefore indispensable. The choice of the probabilistic model that best fits the data is also crucial in order to reduce systematic errors such as long-branch attraction between fast evolving taxa (Lartillot et al. 2007; Philippe et al. 2011). We built 4,908 alignments devoid of apparent paralogy by selecting single-copy genes thanks to a customized pipeline that was facilitated by genomic data available for eight ascidian species as well as by the elimination of spurious redundancy in the de novo assembled transcriptomes. All phylogenetic analyses conducted under site-homogeneous (LG) and site-heterogeneous (LG4X and CAT) models led to the same topology with maximum statistical support for all nodes. Therefore, the data set produced here turned out to be particularly appropriate to robustly recover phylogenetic relationships of Styelidae and relatives at the family level. It can thus be used as a reference set of orthologous genes in future tunicate phylogenetic studies. The REDUCED data set...
of 1,306 alignments shared by at least 24 species provided the same resolution and can be used as an alternative when the computational power is limited, for instance when using elaborated evolutionary models.

Evolutionary and Developmental Implications for Asexual Reproduction and Coloniality

A diverse range of budding modes has been reported across the whole subphylum of tunicates. In planktonic thaliaceans, buds arise from strobilation of stolons that contain tissues from all the three germ layers (Godeaux et al. 1998; Piette and Lemaire 2015). In phlebobranch ascidians such as Perophora viridis, buds arise from a stolonal vessel composed of an epithermis and containing a transverse mesenchymal septum (Freeman 1964; Kawamura et al. 2007). In aplousobranch ascidians, budding involves different kinds of tissues and mechanisms, from epithelial epicardium in Apodium during abdominal strobilation, to mesenchymal septum during stolonal budding in Clavelina (Berrill 1950; Nakauchi 1982). NED mechanisms in different orders are so diverse that unifying principles are difficult to identify and study. Moreover, a full understanding of the evolution of asexual reproduction and regeneration in tunicates is limited by the current state of the tunicate phylogeny, in which the relationships among orders have only recently started to be resolved (Turon and López-Legentil 2004; Tsagkogeorga et al. 2009; Tatián et al. 2011; Shenkar et al. 2016; Kocot et al. 2018, Delsuc et al. 2018). However, focusing on one particular order or family allows one to decipher phylogenetic relationships between closely related solitary and colonial species and thus to reconstruct the evolution of NED at the intrafamily level (Shenkar et al. 2016).

In the phylogeny presented here, the solitary Styela and Asterocarpa form an early-diverging clade together with the families Pyuridae and Molgulidae, the latter being exclusively composed of solitary species. Therefore, the most parsimonious interpretation is that the last common ancestor of Styelidae was a solitary animal, and as a consequence NED and coloniality probably arose at least twice independently in Styelidae in two branches: 1) the monophyletic PB clade, that is, Polozoainea + Botryllinae; and 2) P. zorritensis. An alternative interpretation of NED evolution, considering its acquisition in deeper nodes, can only be explained with multiple independent reversions to solitary lifestyle (see examples of different scenarios in supplementary fig. 7A, Supplementary Material online). In general, our topology of major Stolidobranch clades is in accordance with previously reported phylogenies that retrieved paraphyletic pyurids (Zeng et al. 2006; Tsagkogeorga et al. 2009; Shenkar et al. 2016). Within styelids, our result contrasts with previous phylogenies based on ribosomal and mitochondrial genes showing either a single acquisition (Zeng et al. 2006; Tsagkogeorga et al. 2009; Shenkar et al. 2016) or at least five independent acquisitions of coloniality (Pérez-Portela et al. 2009). However, these studies either included a limited number of colonial styelids, notably lacking P. zorritensis, or their topologies received poor statistical support.

In the phylogenic tree resulting from this study, only members of Botryllinae are known to perform vascular budding within the PB clade (for colony growth or regeneration), while in Polozoinea only peribranchial budding has been reported (Selys-Longchamps 1917; Berrill 1950; Abbott 1953; Watanabe and Tokioka 1972; Kawamura and Watanabe 1981). Thus the species tree implies that vascular budding is a Botryllinae innovation, and secondarily acquired in an ancestor that already possessed the peribranchial budding. Our understanding of budding evolution could be further improved in the future by including more colonial styelids to our phylogenomic data set. For instance, adding species of the genera Polyzoa (peribranchial budding), Metandrocarpa (peribranchial budding), and Symplegma (vascular/peribranchial budding).

The fact that budding tissues are of a different nature between P. zorritensis and the PB clade species reinforces the hypothesis that NED has been acquired convergently. Brunetti and Mastrototaro (2004) provided a general description of P. zorritensis and defined NED in this species as a form of vascular budding. However, a detailed morphological description was not provided. In the Stolidobranchia, only peribranchial budding and vascular budding have been extensively studied. Both these NEDs pass through a double vesicle stage (fig. 1G, H), where the outer vesicle in peribranchial budding derives from the parental epithermis and gives rise to the epithermis of the new zooid, and an inner vesicle, which will give rise to most of the internal organs, derives either from a portion of peribranchial epithelia (fig. 1G) or from population(s) of circulating cells (fig. 1H), in peribranchial and vascular budding respectively. The histological analyses of P. zorritensis budding show that the new zooids originate from the blood vessel epithelium but follow completely different ontogenetic trajectories, in which both the inner and the outer vesicles derive from the same tissue, namely the vascular epithermis (fig. 4). To our knowledge, this mode of budding has never been described in any other species.

Each of these three events of NED acquisition involved nonhomologous tissues for the formation of the inner vesicle (vascular epithermis, peribranchial epithelium and hemocytes, respectively) but it remains unknown whether budding is triggered by orthologous genes. The example of NK4 highlights the importance of having a robust phylogenetic framework to select relevant species for comparative studies but also to interpret gene expression patterns in an evolutionary perspective. NK4 is a highly conserved transcription factor involved in heart specification during ascidian embryogenesis (Wang et al. 2013, Prünster et al. unpublished). It also shows localized expression patterns in the anterior endoderm of larve and juveniles in Ciona intestinalis, and in the ventral ectoderm of larvae in both C. intestinalis and Molgula occidentalis (Christiaen et al. 2010; Stolfi et al. 2014). In B. schlosseri and P. zorritensis NK4 shows an additional domain of expression restricted to the tissues of the emerging bud, thus suggesting a putative new role in NED. By exploring the function of NK4 and by adding the description of other bud-specific molecular markers it will be possible to establish whether
homologous programs are directing budding in different context and species, and to explore the plastic and modular nature of gene regulatory networks that are responsible for nonembryonic developmental processes (Ricci et al. 2016).

New Insights on Styelidae Systematics and Future Directions

Relationships of major Styelidae clades have been debated for more than a century. Berrill (1948) considered that all colonial species should be unified as a natural group because he assumed that they all perform peribranchial budding. In contrast Kott (1985), in her revision of Australian ascidians, suspected that budding modes may be more diverse than expected and proposed that “many species [of Polyandrocarpa] have a closer relationship with species of Polycarpa than they do with one another [...]” and that “Accurate resolution of their taxonomy must await information on the process of vegetative reproduction, which may indicate their true phylogeny.” Our phylogenetic tree and the histological investigations on budding mode reconcile the views of Berrill and Kott. As Berrill suspected all peribranchial budding species form a monophyletic group, and as Kott suspected, the genus Polyandrocarpa is polyphyletic and P. zorritensis has an original budding mode.

The phylogenetic position of P. zorritensis is further supported by its gonad morphology, which is of the polycarpid type, that is, a large gonad containing numerous male and female lobules all together embedded in a common sac, also found in Polycarpa and Dendrodoa (Monniot and Monniot 1972; Monniot 2016). In contrast, other colonial styelids have either sex-separated gonads or small hermaphroditic gonads with no more than two testes (Monniot and Monniot 1972).

It would be informative to incorporate into the existing phylogenomic matrix other Polyandrocarpa species having polycarpid gonad, e.g. Polyandrocarpa arianae (Monniot 2016), and to compare their budding mechanisms. Interestingly, our phylogenetic tree is also in agreement with the view that the acquisition of brooding predates the evolution of coloniality (Brown and Swalla 2012), and that internal brooding was secondarily lost in the oviparous P. misakiensis (see supplementary fig. 7B, Supplementary Material online for comparison between alternative scenarios). However, broader sampling of PB clade species would be necessary to better trace the evolution of this character.

Conclusions

Experimental approaches to study macroevolutionary transitions require a priori definitions of homologous characters, but also the identification of the most suitable species in which to trace the evolution of such characters, as well as the knowledge of the phylogenetic relationships among these species. The phylogeny provided in this paper sets an initial framework to characterize the emergence and disappearance of other complex sets of characters, such as agamic asexual propagation and whole body regeneration. In this regard, the tools we generated already draw our attention to the plasticity of NED in an ascidian family and led to the discovery of an undescribed form of asexual development. Comparing the cellular and molecular mechanisms behind each form of NED can also lead to the discovery of conserved developmental modules that regulate regeneration and asexual development. The phylogenetic framework presented here needs further taxonomic sampling to fully understand the evolutionary transitions to NED, but we can already use these findings to trace and study other characters with known plastic evolution, such as larval photoreceptors (Sorrentino et al. 2001), adulthood (Jeffery 2007) or brooding strategies.

Materials and Methods

Animal Collection and Identification

Sampling locations are provided in table 1. Brazilian collections were authorized by license No. 44082-1 of the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). Before relaxation and fixation in menthol and 10% formalin, animals were kept for ~12 h in running, grossly filtered sea water, shortly after which they were fixed. Solitary species are not externally recognizable and need to be dissected, formaldehyde-fixed, and carefully examined. For this reason, a single individual cannot be used for both morphology-based identification and RNA extraction, since fresh tissues have to be frozen very rapidly to get good quality RNA. Therefore, we used the following strategy: when two individuals looked similar, one was dedicated to morphology-based identification and fixed in 4% formaldehyde, the other was used for RNA extraction and flash-frozen in liquid nitrogen. From both, a small piece of tissue was first removed and transferred in 80% ethanol in order to extract DNA and proceed to mitochondrial cytochrome oxidase subunit I (COI) and nuclear 18S rRNA barcoding. By this method, both individuals could be compared at the DNA level, allowing confirmation that the morphology-based identified specimen was of the same species as the frozen one, from which the transcriptome was to be sequenced. When COI and/or 18S rRNA sequences had been deposited in databases by others, morphology-based species determination was confirmed by sequence identity. Before flash-freezing specimens in liquid nitrogen, the tunic and digestive tract were removed in order to decrease the level of contaminating RNA. Concerning colonial species, three zooids or pieces of the same colony were fixed in ethanol, formaldehyde, and liquid nitrogen respectively. In some colonial species, the zooids were so small and so deeply embedded in the tunic that digestive tract and tunic could not be removed before fixation. In these cases, the animals were starved for 24 h and the tunic was cleaned by brush and forceps, as much as possible.

DNA Barcoding and Phylogenetic Analyses Based on COI and 18S rRNA

COI was amplified using primers modified from Folmer et al. (1994), Meyer (2003), and Schwendinger and Giribet (2005). Ribosomal 18S rRNA gene was amplified using 18S-TF and 18S-TR primers from Carreras-Carbonell et al. (2005). A list of primers used for each species is provided in supplementary table 5, Supplementary Material online. For Brazilian samples,
DNA was extracted with DNeasy Blood & Tissue kit (Qiagen) following manufacturer’s instructions, and then quantified on a NanoDrop 2000 Spectrophotometer (Thermo Scientific). The PCR amplifications were carried out in 20 μl reactions with 0.15 μl GoTaq polymerase in supplied buffer (Promega), 10 mM dNTPs, and 10 mM of each primer. Cycling parameters were as follows: a 94 °C denaturation for 5 min, followed by 35 cycles of: 94 °C denaturation for 30 s, 45 °C annealing for COI or 66 °C for 18S for 30 s, 72 °C extension for 1 min, and a final extension at 72 °C for 5 min. The PCR products were labeled using BigDye Terminator ver. 3.1 Cycle Sequencing kit chemistry (Applied Biosystems), and sequenced on an Applied Biosystems 3730 DNA sequencer. For French samples, DNA was extracted as follows: tissues was first digested by 200 μg/ml Proteinase K at 37°C for 30 min then mashed with a sterile tip; nucleic acids were purified by phenol/chloroform/isoamyl alcohol (25:24:1), precipitated in sodium chloride 0.2 M and finally quantified on a NanoDrop 2000 Spectrophotometer (Thermo Scientific). Cycling parameters were the same as in Brazil and France. Successful amplifications were sequenced on an ABI 3730 at Genewiz, Inc. (South Plainfield, NJ, USA). New COI and 18S rRNA sequences have were in GenBank (supplementary table 5, Supplementary Material online) for accession numbers).

All sequences were aligned and trimmed using Geneious R7 (Biomatters LTD). We aligned the sequences with those from Pérez-Portela et al. (2009) and some available on Genbank (NCBI; see supplementary table 5, Supplementary Material online for accession numbers) using CLUSTALW (Higgins et al. 1996) and trimmed and confirmed by eye. The 18S rRNA alignment was 801 bp and the COI alignment was 679 bp. As in Pérez-Portela et al. (2009), we concatenated the COI and 18S rRNA sequences. Two partitions were specified, one for each gene. We conducted a Maximum Likelihood analysis using the program RAxML v.8 (Stamatakis 2014) with nucleotide substitution model GTR + G + I, and run 100 bootstrap replicates on our alignment of 100 sequences.

RNA Isolation and cDNA Sequencing

Information on tissue used for RNA extraction is provided in table 1. For the species collected in Roscoff, total RNA was extracted in the laboratory of S.Tiozzo (France), using and NucleoSpin RNA II kit (Macherey-Nagel), including a DNAase treatment, and following manufacturer’s instructions. For P. misakiensis, RNA was extracted in the laboratory of K. Kawamura (Japan), using NucleoSpin RNA kit (Takara-Clontech). For species collected in Brazil, RNA was extracted in the laboratory of F. Brown (Brazil) using a NucleoSpin RNA Plus kit (Macherey-Nagel). Total RNA were sent to Beijing Genome Institute (BGI, China) where cDNA libraries were prepared before sequencing. For species collected in Banyuls, RNA was extracted in the laboratory of S. Darras (France) using Trizol (Life Technologies), followed by purification using the RNasy kit (Qiagen) with a DNase treatment. cDNA preparation and sequencing were performed by Sistemas Genomicos (Spain). RNA-seq data were obtained using Illumina HiSeq 4000 technology. The exact number of reads per species is given in table 1.

De Novo Transcriptome Assembly

For all the species for which de novo transcriptomes were assembled, we applied the following eight-step pipeline: 1) Illumina reads were clipped and trimmed to eliminate low quality regions using Trimmomatic v. 0.35 (Bolger et al. 2014) with the following parameters: PE reads_1.fq reads_2.fq ILLUMINACLIP: adapters/TruSeq3-PE.fa: 2: 30: 10 LEADING: 36 TRAILING: 36 SLIDINGWINDOW: 4: 34 MINLEN: 100 and then sequence quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed April 2018); 2) reads were assembled with Trinity v. 2.4.0 (Grabherr et al. 2011) using default parameters; 3) open reading frames (ORF’s) of a minimum of 90 amino acids were kept using Transdecoder v. 3.0.1 (https://github.com/TransDecoder/TransDecoder/wiki; last accessed April 2018); 4) very similar ORF’s (e.g., putative isoforms or alleles) were collapsed using CD-HIT-EST v4.6 (Fu et al. 2012) with default parameters; 5) reads were then mapped back onto the remaining contigs using Kallisto v. 0.42.4 (Bray et al. 2016) with default parameters; 6) density plots were drawn with R, using hist() and density() functions. On the basis of these plots, contigs with low expression levels were eliminated. The threshold above which contigs were kept corresponds to the lower peak of the density plot; 7) exhaustiveness was assessed by running BUSCO (Simão et al. 2015) against metazoan database; 8) cross-contaminations were removed using CoCo (Simion et al. 2018). Number of contigs at each step for each species is given in supplementary table 1, Supplementary Material online.

Data Availability

The 18 de novo assembled transcriptomes can be downloaded from https://github.com/AlexAlie/styelidae; last accessed April 2018. Raw Illumina reads obtained in this study can be retrieved from NCBI BioProject PRJNA422120. Accession numbers or Illumina reads from B. schlosseri (Rodriguez et al. 2014) and B. leachii (Zondag et al. 2016) are provided in table 1.


Construction of a New Set of Orthologous Alignments

We selected 15 chordate species for which genomic data were available (i.e., six tunicates, six vertebrates, one cephalochordate, one echinoderm, and one hemichordate; see supplementary table 3, Supplementary Material online), and we retained only the longest transcript for each gene of their corresponding proteomes. These proteomes were further dereplicated using CD-HIT (~c 1.0; Fu et al. 2012) and sequences <30 amino acids were removed. These 15 proteomes were clustered using OrthoMCL (~i 1.3, Li et al. 2003). We later added three new proteomes from tunicate species for which genomic data

1739
became available (supplementary table 3, Supplementary Material online) into the clusters using Forty 2 (https://bitbucket.org/dbaurain/42/). The few sequences that were simultaneously added to different clusters were discarded, except in the cluster in which taxonomic diversity was the highest. We discarded all resulting clusters that did not possess at least: 1) two tunicate sequences or 2) one tunicate and one vertebrate sequence, retaining 14,552 clusters.

**Gene Augmentation, Selection, and Concatenation**

Transcriptomic data from 18 stolidobranchs including 16 styelds (table 1) were added to the 14,552 clusters of orthologous sequences using Forty 2. Clusters that contained apparent paralogy for at least one tunicate species among those that correspond to the genomic data set (i.e., species belonging to the Ciona, Phallusia, Molgula, and Halocynthia genera) or which contained <10 tunicate species or which contained >80 sequences were discarded. All nontunicate sequences were removed from the remaining 8,187 clusters in order to improve the accuracy of the subsequent aligning step, a. All clusters were aligned using Mafft (–maxiterate 1000 –localpair, Katoh and Standley 2013). For every alignment, nonoverlapping sequences were fused as they most likely correspond to sequence fragments from the same transcript. We used an extended definition of a “nonoverlap” between two sequences: if actually overlapping, the total overlap is smaller than 200 amino acids and the two sequences differ for 44 or 68 that 58.

We used an extended definition of a “nonoverlap” between two sequences: if actually overlapping, the total overlap is smaller than 200 amino acids and the two sequences differ by less than five amino acid positions. This allowed us to concatenate 64,340 fragmented transcripts (out of the initial 219,099 transcripts) into 28,188 longer ones. We then removed 4,631 low-quality transcripts (i.e., that had >70% missing data and were shorter than 100 amino acids). Alignments were then checked again for apparent paralogy, this time for all species, and we only kept the 4,908 alignments that showed no paralogy. We also selected alignments containing at least 24 species (out of the 26 total species) to create a smaller and more complete phylogenomic data set of 1,306 alignments. These alignments were concatenated into supermatrices using ScaFos (Roure et al. 2007), available on https://github.com/AlexAliie/styeldidae; last accessed April 2018.

Functional annotation was done through the DAVID online service (Huang et al. 2009).

**Phylogenetic Analyses**

The assembled supermatrices as explained above were analyzed using several models of sequence evolution to assess the relative reliability of the different species to infer the phylogenies we might infer. The FULL supermatrix was analyzed under the Maximum Likelihood framework using the site-heterogeneous model of sequence evolution LGþG4 þ F implemented in IQ-TREE (Nguyen et al. 2015) and node support was estimated with 100 bootstrap replicates. The REDUCED supermatrix was analyzed with the exact same approach, as well as using the site-heterogeneous LG4X þ R4 þ F sequence evolution model (also in IQ-TREE). Lastly, we also analyzed the REDUCED supermatrix under a Bayesian framework using the site-heterogeneous CAT þ F81 þ G4 model implemented in PhyloBayes (Lartillot et al. 2013). We ran two independent MCMC chains until the two topologies obtained perfectly converged (maxdiff = 0 with a 50% burnin).

**In Situ Hybridization**

In situ hybridization on B. schlosseri was performed as in Ricci et al. (2016). In situ hybridization on P. zorritensis was performed on paraffin sections as follows. Samples at the desired stage were fixed in paraformaldehyde (PFA) 4%, DMSO 1%, DEPC 0.2%, RNase inhibitor (Sigma, R7397-30ML) 1/500 (=thereafter 1XPR), in Phosphate Buffer Saline (PBS 1×), then washed twice in PBS 1× and then dehydrated by graded methanol bath of 15 min (from 25% to 100% in PBS + 1XPR), then samples were immediately processed or stored at -20 °C. For paraffin embedding, samples were washed 2 × 10 min in butanol 1, then 1 h in butanol 1/paraffin at 60 °C, then overnight in paraffin at 60 °C. Paraffin blocks were allowed to solidify 8 h at room temperature and then were stored at 4 °C no more than two days. Paraffin sections were performed on Leica Microtome, at a thickness of 15 μm and attached on SuperFrost glass slides and allowed to dry overnight at 37 °C. All tools were previously washed with RNase decontamination solution. Then in situ hybridizations were performed as follows: Slides were dried at 55 °C for 30 min, followed by two baths of xylene of 10 min each, then rehydrated by graded ethanol baths and then washed 2 × 10 min in a solution PBST (PBS 1× + 0.1% Tween 20) + 1XPR, for 5 min. Slides were then treated with Proteinase K (2.5 ng/μl in PBST + RNase inhibitor) for 10 min at 37 °C, and immediately rinsed in PBST + 1XPR, and bath for 2 min in PBST. Then they were postfixed in PFA 4% + glutaraldehyde 0.2% in PBST + 1XPR for 30 min at room temperature, and rinsed in in PBST + 1XPR 3 × 5 min. Slides were prehybridized in buffer (Formamide deionized 65%; 5 × SSC; 1× denhardt’s solution; 0.1% tween 20; 5 mg/ml torula yeast RNA; 50 μg/ml heparine; 1% Dextran sulfate) for 1 h, after what probes at a final dilution of 1.5–1.7 ng/μl in 300 μl of hybridization buffer, for 44 or 68 h at 58 °C. Chambers of hybridization (FRAMESEAL BIORAD SLF 3001) were applied to the slides who accepted 300 μl of probe dilution. The slides were rinsed in 2 X SSC pH 4.5, 50% Formamide, 0.1% Tween 20 + 1XPR, 3 × 30 min at 58 °C, then rinsed in PBST + PR 2 × 5 min and in MABT (MAB 1× + 0.1% Tween 20) + RNase inhibitor 1× 2 × 5 min (each bath). Then the slides were blocked in MABT, 2% Blocking solution, 10% goat serum for 1 h at RT. 500 μl of antigid-AP (1/2,000 in MABT, 2% blocking solution, 10% goat serum) was deposited directly on the slides overnight at 4 °C. Slides were rinsed in MABT 2 × 5 min, then rinsed TMNT filtered (tris–HCl ph 9.5 100 mM, NaCl 100 mM, MgCl2, 50 mM, Levamisole 10 mM, 0.1% Tween 20) 2 × 5 min. Then NBT/ BCIP in TMNT was directly applied on the slides until the coloration developed. The reaction was stopped by 2 × 5 min rinse in PBST. Then slides were dehydrated in successive baths of ethanol (50%, 70%, 90%, 100%), 3 min each, and finally in xylene for 15 min and mounted in Permount medium.

**P. zorritensis Husbandry and Histology**

Colonies of P. zorritensis were collected in Taranto and La Spezia (Italy), fixed on glass slides and cultured in a circulating...
sea water system at 24 °C. The developing buds were fixed in 4% PFA in phosphate-buffered saline (PBS, pH 7.4) over night at 4 °C, dehydrated in a graded series of ethanol and butanol and then included in paraffin wax. Transversal section (5–6 μm thick) were cut using a hand operated microtome and stained with hematoxylin and eosin (H&E) following this protocol: stain for 3 min in Mayer’s hematoxylin (Hematoxylin Solution, Mayer’s Sigma–Aldrich), wash for 10 min in running tap water, stain for 15 s in Eosin Y solution (1 g Eosin Y powder in 20 ml MQ water and 80 ml 95% EtOH), rinse in 95% EtOH, dehydrated and mounted.

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments
We would like to thank Philippe Dru and Faisal Bekkouche for help with bioinformatics resources, Laurent Gilletta, Loann Gissat, Regis Lasbleiz, Sophie Collet, and Camille Lejeune for setting up the aqua-culture systems. Françoise Monniot for advice on taxonomy and sample identification. Sophie Booker, Elodie Bourrigaud, Laurent Lévêque, Ronan Garnier, Francesco Mastrototaro, and the Service Mer of Banyuls-sur-Mer, EMBRC-France for sample collection. We thank Martine Fioroni for providing literature and old manuscripts that were difficult to obtain. This work was supported by FAPESP Jovem Pesquisador grant (JP 2015/50164-5) to F.B., ANR grant (ANR-14-CE02-0019-01) to S.T., FAPESP postdoctoral fellowship Pesquisador grant (JP 2015/50164-5) to F.B., ANR grant (ANR-11-JSV2-007-01) to S.D.

Author Contributions
A.A., L.H., F.B., and S.T. conceived and designed the study, collected and identified biological samples and wrote the manuscript. A.A., L.H., and P.S. designed the assembly and clustering pipeline and performed sequence analyses. P.S. conducted orthology and phylogenomic analyses. A.A. and M.S. established P. zorritensis culture and conducted histological analyses. M.M.P. conducted NK4 in situ hybridization on B. schlosseri. A.A. and L.H. conducted DNA barcoding and phylogenetic analysis. S.L. performed NK4 in situ hybridization on P. zorritensis. E.D. and F.D. participated in the pipeline design and supported part of the transcriptome sequencing. P.L. and C.D. sequenced genomes and provided gene models. S.D. and K.K. provided RNA for several species and participated to species identification. All authors read and approved the final version of the manuscript.

References
Hughes RN, Cancino JM. 1985. An ecological overview of cloning in the tunicae


