De novo draft assembly of the *Botrylloides leachii* genome provides further insight into tunicate evolution.

Simon Blanchoud[1#], Kim Rutherford[2], Lisa Zondag[1], Neil Gemmell[2] and Megan J Wilson[1*]

1 Developmental Biology and Genomics Laboratory

2

Department of Anatomy, School of Biomedical Sciences, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

Current address: Department of Zoology, University of Fribourg, Switzerland

* Corresponding author:

Email: meganj.wilson@otago.ac.nz

Ph. +64 3 4704695

Fax: +64 479 7254

Keywords: chordate, regeneration, *Botrylloides leachii*, ascidian, tunicate, genome, evolution

Abstract (250 words)

Tunicates are marine invertebrates that compose the closest phylogenetic group to the vertebrates. This chordate subphylum contains a particularly diverse range of reproductive methods, regenerative abilities and life-history strategies. Consequently, tunicates provide an extraordinary perspective into the emergence and diversity of chordate traits. Currently published tunicate genomes include three Phlebobranchiae, one Thaliacean, one Larvacean and one Stolidobranchian. To gain further insights into the evolution of the tunicate phylum, we have sequenced the genome of the colonial Stolidobranchian *Botrylloides leachii*.

We have produced a high-quality (90% BUSCO genes) 159 Mb assembly, containing 82 % of the predicted 194 Mb. The *B. leachii* genome is much smaller than that of *Botryllus schlosseri (*725 Mb), but comparable to those of *C. intestinalis* and *M. oculata* (both 160 Mb). This difference is largely due to an increase in repetitive DNA content in *B. schlosseri*. By analyzing the structure and composition of the conserved homeobox gene clusters, we identified many examples of multiple cluster breaks and gene dispersion, suggesting that several lineage-specific genome rearrangements occurred during tunicate evolution.

In addition, we investigate molecular pathways commonly associated with regeneration and development. We found lineage-specific gene gain and loss within the Wnt, Notch and retinoic acid pathways. Such examples of genetic changes to key evolutionary conserved pathways may underlie some of the diverse regenerative abilities observed in the tunicate subphylum. These results, combined with the relatively recent separation from their last common ancestor (630 MYA), supports the widely held view that tunicate genomes are evolving particularly rapidly.

Introduction

Tunicates are a group of marine suspension-feeding hermaphrodites found worldwide in the inter- or sub-tidal region of the seas. This subphylum of invertebrates is phylogenetically located in the Chordata phylum, between the more basal Cephalochordata and the higher Vertebrata, of which they are considered the closest relatives (Fig. 1A; (Delsuc *et al.*, 2006)). These organisms include a wide range of reproductive methods, regenerative abilities, developmental strategies and life cycles (Lemaire *et al.*, 2008). Importantly, and despite a drastically different body plan during their adult life cycle, tunicates have organs as well as a tissue complexity incipient to that of vertebrates (Fig. 1A), including a heart, a notochord, an endostyle and a vascular system (Millar, 1971). In addition, this group of animals is undergoing faster genomic evolution compared to vertebrates through a higher nucleotide substitution rate in both their nuclear and mitochondrial genomes (Tsagkogeorga *et al.*, 2010, 2012; Rubinstein *et al.*, 2013; Berna and Alvarez-Valin, 2014). Therefore, this chordate subphylum provides an excellent opportunity to study the origin of vertebrates, the emergence of clade specific traits and the function of conserved molecular mechanisms. Biological features that can be investigated in tunicates include, among others, the evolution of colonialism, neoteny, sessilness, and budding. However, there are currently only seven Tunicata genomes publicly available, of which three have been well annotated. There is thus a paucity in the sampling of this very diverse subphylum.

Tunicates are separated into eight clades: Phlebobranchia (class Ascidiacea), Molgulida (class Ascidiacea), Copelata (class Appendicularia), Pyrosomida (class Thaliacea), Salpida (class Thaliacea), Doliolida (class Thaliacea), Aplousobranchia (class Ascidiacea), and Stolidobranchia (class Ascidiacea). Phlebobranchiae and Molgulida are clades of sessile solitary benthic ascidians. These animals reproduce sexually, releasing eggs through their atrial siphon for external fertilization, hence producing a motile larva. These larvae will

colonize novel environments, attach to a submersed substrate and undergo metamorphosis into a sessile filter-feeding adult. These ascidians are capable of regenerating a limited set of organs, including their oral siphon (Auger *et al.*, 2010) although regeneration capability reduces as they age (Jeffery, 2015). These clades are currently the most sampled ones with five published genomes (*Ciona intestinalis*, *Ciona savigny, Molgula oculata*, *Molgula occulta*, *Molgula occidentalis;* (Dehal *et al.*, 2002a; Small *et al.*, 2007; Stolfi *et al.*, 2014), and two yet unpublished species (*Phallusia mamilata*, *Phallusia fumigata*; (Brozovic *et al.*, 2016)). *C. intestinalis* and *O. dioika* where the first Tunicate species to have been sequenced and are the best characterized.

Copelata is a class of planktonic free-swimming organisms that possess chordate traits common to all tunicate larvae including a notochord, neural tube and pharyngeal slits. These social neotenic animals form dioecious communities where each individual lives inside a special external mucous structure, termed house, which concentrates and funnels their food. *Oikopleura dioika* is the sole example of the Copelatan to have its genome sequenced, showing exceptional compaction (70 Mb; (Seo *et al.*, 2001)). Whether these animals can undergo regeneration has not yet been assessed.

Pyrosomida, Salpida and Doliolida are clades of planktonic pelagic animals forming cylindrical free-floating compound colonies (Piette and Lemaire, 2015). These organisms can reproduce both sexually, through autogamy to initiate novel colonies, as well as asexually, through stolonial budding to increase the size of the colony. Owing to their peculiar life cycle and habitat, these tunicates have not been extensively studied, no genome has been sequenced and whether they can undergo regeneration remains unknown.

Aplousobranchia and Stolidobranchia consist of both solitary and colonial sessile benthic organisms. Colonial tunicates are capable of both sexual, through autogamy, and asexual reproduction, through a wide range of budding types (palleal, vascular, stolonial,

90 pyloric and strobilation; (Brown and Swalla, 2012)). In addition, these compound organisms
91 can undergo whole-body regeneration (WBR; reviewed in (Kürn *et al.*, 2011)). Colonial
92 ascidians are emerging as unique and increasingly popular model organisms for a variety of
93 studies including immunobiology, allorecognition, angiogenesis and WBR (Rinkevich *et al.*,
94 1995, 2013; Ballarin *et al.*, 2001; Manni *et al.*, 2007; Rinkevich, Douek, *et al.*, 2007; Gasparini
95 *et al.*, 2008; Franchi *et al.*, 2011; Lauzon *et al.*, 2013). While the genome of two solitary
96 Aplousobranchian are currently being assembled (*Halocynthia rorezi*, *Halocynthia aurantium*;
97 (Brozovic *et al.*, 2016)), colonial tunicates were only exemplified by *Botryllus schlosseri*.
98 Furthermore, this ascidian which a significant expantion of its genome size when compared to
99 the other available Tunicate genomes (725 Mb). To further investigate this fascinating
100 subphylum and assess whether genome expansion is a prerequisite for coloniality, we have
101 assembled and analysed the genome sequence of *Botrylloides leachii* (class Ascidiacea, order
102 Stolidobranchia; (Savigny, 1816)).

103 The viviparous colonial ascidian *B. leachii* can reproduce sexually through a tadpole
104 stage that allows the settlement of a novel colony onto a new substrate (Fig. 1B). Each colony
105 is composed of genetically identical adults (termed zooids) organized in ladder-like systems
106 and embedded in gelatinous matrix (tunic). While each adult has its own heart, they all share
107 a common vascular system embedded within the tunic. In the presence of sufficient food
108 supply, the size of the colony doubles approximately every 20 days through synchronized
109 asexual reproduction, known as palleal budding. During this process each adult produces two
110 daughter zooids that ultimately replace the mother, which is then resorbed by the colony. In
111 addition, upon loss of all zooids from the colony, *B. leachii* can undergo whole-body
112 regeneration and restore a single fully-functional adult in as little as 10 days from a small
113 piece of its vascular system (Rinkevich *et al.*, 1995). Furthermore, when facing unfavorable
114 environmental conditions, these colonial tunicates can enter into hibernation, whereby all

115 zooids undergo regression and are resorbed by the remaining vascular system. When a
116 favorable environment is restored, mature adults will be restored to reestablish the colony
117 (Burighel *et al.*, 1976).

118 We have assembled and annotated the first *de novo* draft genome of the *B. leachii*by
119 taking advantage of our recently published transcriptomes (Zondag *et al.*, 2016). Using this
120 genome, we have then undertaken a large-scale comparison of the four best-annotated
121 ascidian genomes *(B. schlosser*i, *C. intestinalis, M. oculat*a and *O. dioika*) to gain insights into
122 some of the diverse biological abilities that have evolved within the Tunicata.

123 # Results

124

125 ***Genome assembly and annotation***

126 To minimize contamination from marine algae and bacteria typically present in the
127 pharyngeal basket of feeding *B. leachii*, we isolated genomic DNA from embryos of a single
128 wild *B. leachii* colony. Genomic DNA was used to produce two libraries: one short-range
129 consisting of 19090212 fragments (300 bp) of which 100 bp were paired-end sequenced -
130 important for obtaining high coverage - and a second long-range mate pair with 31780788
131 fragments (1.5-15 kb size range, median ~3 kb) of which 250 bp were paired-end sequenced
132 – important for scaffolding the assembly. Following quality checks, low quality reads were
133 removed and sequencing adaptors were trimmed, thus resulting in a high-quality dataset of
134 86644308 paired-end and 12112004 single-end sequences (100% with a mean Phred score
135 >= 30, <1% with an adapter sequence, Fig. S1).

136 We then followed a reference-free genome characterization (Simpson, 2014); provided
137 with statistics from the human, fish (*Maylandia zebra*; (Bradnam *et al.*, 2013)), bird
138 (*Melopsittacus undulatus;* (Bradnam *et al.*, 2013)) and oyster (*Crassostrea gigas*, (Zhang *et al.*,
139 2012)) genomes for comparison; to estimate three properties of the *B. leachii* genome. First,
140 k-mer count statistics were used to estimate the genome size to be 194.2 Mb (194153277 bp).
141 This size is thus comparable to that of the solitary *C. intestinalis, C. savigny and M. oculata*
142 (160 Mb, 190 Mb and 160 Mb, respectively; (Dehal *et al.*, 2002a; Small *et al.*, 2007; Stolfi *et al.*,
143 2014), larger than the compacted 70 Mb genome of *O. dioika* but appreciably smaller than the
144 predicted 725 Mb genome of the related colonial ascidian *B. schlosseri*, of which 580 Mb have
145 been sequenced (Voskoboynik *et al.*, 2013). Second, by quantifying the structure of the de
146 Brujin graph obtained using the k-mer counts, the computational complexity of the assembly

147 was estimated (sequencing errors 1/213, allelic differences 1/233, genomic repeats 1/2439).
148 With a cumulative occurrence of 1/106, the *B. leachii* genome is similar to that of bird, more
149 variable than those of fish and human, but still quite less complex than the notably difficult
150 oyster genome (Fig. S1). Third, sequence coverage was estimated using the distribution of 51-
151 mers counts, showing a well-separated bimodal distribution with a true-genomic k-mers
152 maximum at 31x coverage, similar to the human genome but higher than both the fish and the
153 bird. Overall, these metrics suggest that *B. leachii* has a genome well suited for *de novo*
154 assembly and that our sequencing could result in a high quality assembly .

155 *De novo* assembly using Metassembler (Wences and Schatz, 2015) produced a genome
156 of 159132706 bp (estimated percentage of genome assembled is 82%), with an average
157 sequencing coverage of 66x (after adaptor trimming). The assembly is composed of 1778
158 scaffolds, with a N50 scaffold length of 209776 and a L50 scaffold count of 223. The 7783
159 contigs, with a N50 length of 48085, and a L50 count of 781, represent a total size of
160 146061259 (92%, Table 1). To evaluate the completeness of our assembly, we used the
161 Benchmarking Universal Single-Copy Orthologs (BUSCO; (Simão *et al.*, 2015)). This tool
162 provides a quantitative measure of genome completeness by verifying the presence of a set of
163 manually curated and highly conserved genes. Out of the 843 orthologues selected in
164 metazoans, 760 (90%) were found in our assembly of the *B. leachii* genome (File S1), a
165 relatively high score when compared to the BUSCO score of frequently used genome
166 assemblies such as *Homo sapiens* (89%, GCA_000001405.15) and *Mus musculus* (78%,
167 GCA_000001635.4). In addition, we took advantage of our previous assembly of the *B. leachii*
168 transcriptome (Zondag *et al.*, 2016) to further assess the quality of our genome. Using BLAT
169 (Kent, 2002), we were able to map 93 % of transcript sequences (48510/52004) onto our
170 assembly. Finally, we further estimated the quality of our assembly by mapping the
171 proteomes of the available tunicate genomes using tBLASTn (Camacho *et al.*, 2009): *C.*

172 *intestinalis* 71% (4233/14740), *C. savigny* 81% (3835/20155), *M. oculata* 77% (3828/16616)
173 and *B. schlosseri* 71% (13575/46519). Overall, these results indicate that our *de novo* genome
174 is largely complete and suitable for annotation.

175 *Ab initio* genome annotation was performed using MAKER2 (Holt and Yandell, 2011)
176 and predicted 15839 coding genes, of which 13507 could be classified using InterProScan
177 (Jones *et al.*, 2014). Comparing these predictions with our mapping of the transcriptome, we
178 found out that 83% of our aligned cDNA (40188/48510) mapped to a predicted gene locus
179 thus spanning 78 % of the annotated genes (12395/15839). In addition, a total of 4213 non-
180 coding sequences were predicted using Infernal (Nawrocki and Eddy, 2013), Snoscan (Lowe,
181 1999) and tRNAscan-SE (Lowe, 1999). Finally, repetitive elements were annotated using
182 RepeatMasker (Smit *et al.*, 2015) and a specie-specific library created using the
183 RepeatModeler module (Smit and Hubley, 2015). 18% of the genome was identified as
184 containing repetitive elements (Table 2), a majority (17%) of these being interspersed
185 repeats. This proportion is similar to that found in other tunicates including *C. intestinalis*
186 (17%), M. oculata (22 %) and *O. dioica* (15%), while being lower than that in *B. schlosseri*
187 (60%) and *C. savigny* (33 %).

188

189 ***Ancient gene linkages are fragmented in tunicate genomes***

190 Ancient gene linkages are spatially restricted and highly conserved sets of genes. These
191 clusters arose in a common ancestor and were preserved because of a common regulatory
192 mechanism such as cis-regulatory elements located within the cluster The homeobox-
193 containing *Hox* gene family, typically composed of 13 members in vertebrates (Hoegg and
194 Meyer, 2005), is among the best-studied examples of such ancient gene cluster and is critical
195 for the correct embryonic development (Pearson *et al.*, 2005). In particular, the linear
196 genomic arrangement within the *Hox* cluster reflects their spatial expression along the

197 anterior-posterior body axis (Pascual-Anaya *et al.*, 2013), which establishes regional identity
198 across this axis. The basal cephalochordate *B. floridae* has all 13 *hox* genes located in a single
199 stereotypical cluster, along with an additional 14th gene (Fig. 2B; (Takatori *et al.*, 2008)),
200 suggesting that the chordate ancestor also had an intact cluster. However in tunicates, this
201 clustering appears to be lost. In *C. intestinalis*, the nine identified *Hox* genes are distributed
202 across five scaffolds, with linkages preserved only between *Hox2, Hox3* and *Hox4; Hox5* and
203 *Hox6*; *Hox12* and *Hox13* (Fig. 2; (Spagnuolo *et al.*, 2003; Wada *et al.*, 2003)). In *O.dioica,* the
204 total number of *Hox* genes is further reduced to eight, split between 6 scaffolds, including a
205 duplication of *Hox9* (Fig. 2A; (Edvardsen *et al.*, 2005)). In *M. oculata* we could identify only six
206 *Hox* genes, divided between 4 scaffolds, with clustering retained for the *Hox10*, *Hox11* and
207 *Hox12* genes (Fig. 2). In Botryllidae genomes, the same seven *Hox* genes are conserved (Fig
208 2B), with a preserved linkage between *Hox10*, *Hox12* and *Hox13* in *B. leachii* and a three copies
209 of *Hox5* present in *B. schlosseri*. Altogether these genomic distributions of the *Hox* cluster
210 genes supports the hypothesis that reduction and separation of this ancient gene linkage
211 occurred at the base of tunicate lineage (Edvardsen *et al.*, 2005). In addition, *Hox9* appears to
212 be specifically retained in neotenic Tunicates while there is no pattern of consereved *Hox*
213 cluster genes specific to colonial ascidians.

214 A second ancient homeobox-containing gene linkage is the *NK* cluster. This cluster,
215 which was seemingly inherited from the last common ancestor of bilaterians (Luke *et al.*,
216 2003), consists of *Msx, Lbx*, *Tlx*, *NKx1, NKx3, NKx4* and *NKx5* (Fig. 3). In *B. floridae*, linkages
217 between *Msx, NKx4* and *NKx3*; as well as between *Lbx* and *Tlx* provide evidence of retained
218 ancestral clustering while *NKx5* was lost (Fig. 3; (Luke *et al.*, 2003)). However in vertebrates,
219 NKx5 is still present while only the gene linkages between *Lbx* and *Tlx* as well as between
220 *Nkx4* and *Nkx3* remain (Fig. 3; (Garcia-Fernàndez, 2005)). To further clarify the evolution of
221 this ancestral cluster in tunicates, we determined the structure of the NK cluster within five

222 ascidian genomes. In all these species, *NKx1* is absent and no evidence of clustering could be
223 found with all identified orthologues located on different scaffolds (Fig. 3). In *C. intestinalis*, *M.*
224 *oculata* and *O. dioica* only five members of this cluster remain, with the loss of either *Lbx* or
225 *Tlx* as well as of *NKx3* and the duplication of the orthologue of *NKx4* (Fig. 3). By contrast, in
226 the colonial tunicates *B. leachii* and *B. schlosseri*, *Tbx*, *Lbx* and *NKx3* are all present. In *B.*
227 *schlosseri*, *Msx1* is absent and *NKx4* duplicated. In the *B. leachii* genome, *NK1* is the only
228 ancestral cluster member to be missing and *Nk5* has been duplicated (Fig. 3). Altogether,
229 these results suggest that there has been a loss of *NKx5* in Cephalochordate, one of *NKx1* in
230 Tunicate and that the conjunction of *NKx3*, *Lbx* and *Tbx* is specific to colonial ascidians.

231 A third ancient linkage that we investigated is the pharyngeal cluster, a gene group
232 present in hemichordates, echinoderm and vertebrates genomes that is considered to be
233 Deuterosome specific (Simakov *et al.*, 2015). The cluster groups *foxhead domain protein*
234 (*FoxA*), *NKx2 (NKx2.2 and Nkx2.1)*, *Pax1/9*, *mitochondrial solute carrier family 25 member 21*
235 (*slc25A21)*, *mirror-image polydactyly 1 protein* (*mipol1*), *egl nine homolog 3* (*egln3*) and
236 *dehydrogenase/reductase member* 7 (*dhrs7*). Among these, s*lc25a21, Pax1/9*, *mipol1* and *FoxA*
237 pairs are also found in protostomes suggesting an even more ancient origin (Simakov *et al.*,
238 2015). The pharyngeal cluster is thought to have arisen due to the location of the regulatory
239 elements of *Pax1/9* and *FoxA* within the introns of *slc25A21* and *mipol1* (Santagati *et al.*, 2003;
240 Wang *et al.*, 2007), constraining the genes to remain in tight association with each other. In
241 the *B. floridae* genome, the entire cluster is located on the same scaffold, with the exception of
242 the *Nkx2.1* and *Nk2.2* gene pair located on separate scaffold. In *C. intestinalis*, only orthologs of
243 *FoxA*, slc25a29, *Pax1* and *Pax9* could be identified. Nevertheless, all of them are located on the
244 same chromosome (Fig. 4). In *O. dioica*, the cluster appears even further reduced as, while
245 orthologues of *FoxA*, *Pax1/9* and *Nkx2.2* genes were found on different scaffolds, only one
246 rather distant linkage (> 1 Mb) between *a Pax-like* gene and *slc25A21* is retained. Copies of

247 Pax1/9, slc25a29 and Nk2.2 genes were found in *B. schlosseri*, however all on different
248 scaffolds. In the *B. leachii* genome, *mipol1* is the sole missing gene from this cluster. However,
249 only the pairing of a *Pax-like* and *slc25A21* genes remains (Fig. 4). For both *B. schlosseri* and *M.*
250 *oculata*, there was no evidence of clustering between genes (Fig. 4) Altogether, these results
251 suggest that most of the Tunicates did not conserve the structure of this ancient linkage, but it
252 is unknown what consequences this would have to their expression and function.
253
254 ***Lineage-specific changes to cell-signaling pathways in Botryllidae genomes.***
255 To dissect more specifically the evolution of colonial ascidians, we examined the
256 genomes of *B. leachii* and *B. schlosseri*, looking for key components of signaling pathways
257 required for metazoan development and regeneration. Of particular interest, we focused on
258 the Wingless-related integration site (Wnt), Notch and Retinoic acid (RA) signaling pathways.
259
260 **Wnt pathway**
261 Wnt ligands are secreted glycoproteins that have roles in axis patterning,
262 morphogenesis and cell specification (Loh *et al.*, 2016). The ancestral gene cluster appears to
263 originate very early on during muti-cellular evolution and to be composed of eleven members
264 (Kusserow *et al.*, 2005; Guder *et al.*, 2006). The *Wnt* gene family expanded to 19 members in
265 the human genome, while independent gene loss has reduced this family reduced to 7 genes
266 in *Drosophila melanogaster* and *Caenorhabditis elegans* (Prud'homme *et al.*, 2002).
267 Consequently, we set out to investigate whether the *Wnt* gene family has either expanded or
268 reduced during Tunicata speciation.
269 Most strikingly, we found among tunicate genomes an increase in the number of *Wnt5a*
270 genes. *C. intestinalis* has a total of 11 *Wnt* genes, including a single *Wnt5a* gene (Fig. 5, Table
271 S2; (Hino *et al.*, 2003)). In the compact *O. dioica* genome, this number has reduced to 6 (Wnts

272 3, 4, 7, 11 and 16), none of which are *Wnt5a* orthologues (Table S2). *M. oculata* has only 7 Wnt
273 ligand genes , including three *Wnt5a-like* genes (Fig. 5, Table S2). In *B. schlosseri*, we identified
274 15 *Wnt* members, including seven *Wnt5a-like* genes on multiple scaffolds (Fig. 5, Table S2).
275 Finally, in the *B. leachii* genome, fourteen *Wnt* ligand genes were identified, including four
276 *Wnt5a* genes located on the same scaffold near *Wnt4* (Fig. 5). Overall, this suggests that an
277 expansion through gene duplication of the Wnt5 family occurred during tunicate evolution,
278 but was lost in some lineages.

279 To assess the functionality of the Wnt pathway in Tunicates, we set out to assess
280 whether its downstream effectors are themselves present in the available genomic data. The
281 downstream pathways activated by Wnt ligands are divided into canonical, non-canonical
282 calcium and non-canonical planar cell polarity. The Wnt5a ligand is associated with both of
283 the non-canonical pathways through binding of membrane receptors that include *frizzled*
284 (*Fzd4*), *receptor tyrosine kinase-like orphan receptor 1/2* (*Ror1/2*) and *atypical tyrosine kinase*
285 *receptor* (*Ryk*). Further downstream, disheveled (dsh), β-catenin (Cnntb), Axin, low-density
286 lipoprotein receptor-related protein 5/6 (LRP5/6) and nuclear factor of activated T-cells
287 (NFAT) are proteins essential for triggering intracellular responses to Wnt signaling
288 (MacDonald *et al.*, 2009). We identified orthologues for each of these signaling transduction
289 molecules in all Tunicata genomes (Table S2), with no evidence of further gene duplication
290 events. This supports the interpretation that signaling through the Wnt pathway is functional
291 in tunicates.

292

293 **Notch pathway**

294 Notch receptors are transmembrane proteins that are involved in cell-cell signaling
295 during development, morphogenesis and regeneration (Hamada *et al.*, 2015). Following
296 activation through the binding of the delta or jagged/serrate ligands, the intracellular domain

297 of Notch is cleaved and induces the expression of downstream target genes including the *hes*
298 *(hairy and enhancer of split*) gene family members (Guruharsha *et al.*, 2012). The presence of
299 both Notch and the Delta/Serrate/lag-2 (DSL) proteins in most metazoan genomes suggests
300 that their last common ancestor had a single copy of each gene (Gazave *et al.*, 2009). To
301 establish how this pathway has evolved in tunicates, we screened these genomes for the
302 Notch receptor using the conserved lin-Notch repeat (LNR) domain, and for genes encoding
303 probable Notch ligands such as genes from the DSL family .

304 In all examined genomes, only a single *Notch* receptor gene was identified while the
305 number of ligand genes varied (Table S3). *C. intestinalis* genome contains two *DSL* genes; *O.*
306 *dioica M. oculata* and *B. schlosseri* a single *DSL* one. By contrast, we found three DSL genes in
307 *B. leachii* (Table S3). To determine the relationships between these identified tunicate DSL-
308 like genes, a phylogeny was constructed along with other chordate DSL proteins. All three *B.*
309 *leachii* genes are Delta orthologues, two of them related to the *B. schlosseri* and *Cionidae.*copy;
310 the third one closer to the *M. oculata* and *H. roretzi* variant. The mouse, human and zebrafish
311 delta and delta-like (DLL) proteins form a discrete clade loosely related to the genes found in
312 Cephalochordate and Tunicate (Fig. 6, shaded box). Jagged proteins form a separate and less
313 conserved clade (Fig. 6). The tunicate DSL-like proteins present long phylogenetic branches,
314 suggestive of greater diversity, which is also observed in the protein alignment (Fig. S3). This
315 suggests that the tunicate DSL proteins are diverging rapidly from each other, indicative of
316 lineage specific evolution of DSL-like genes.

317

318 **Retinoic acid signaling**

319 Retinoic acid (RA) is an extracellular metabolite that is essential for chordate
320 embryonic development. RA is synthesized from retinol (vitamin A) by two successive
321 oxidation steps. In the first step, retinol dehydrogenase (RDH) transforms retinol into retinal.

322 Then RA is produced by aldehyde dehydrogenase (ALDH), a superfamily of enzymes with
323 essential roles in detoxification and metabolism (Jackson *et al.*, 2011). RA influences the
324 expression of downstream target genes by binding to the RA receptors, RAR and RXR (Fig. 7A
325 (Cunningham and Duester, 2015)). Finally, RA is metabolized by the enzymes cytochrome
326 P450 family 26 (Cyp26), a process key to restricting RA-induced responses to specific tissues
327 or cell types (Ross and Zolfaghari, 2011). Components of this pathway have been found in
328 non-chordate animals, suggesting a more ancient origin (Canestro *et al.*, 2006). This pathway
329 has previously been shown to be required for *B. leachii* WBR and *Ciona* development, yet
330 several genes required for RA signaling appear to be missing in *O. dioica* (Martí-Solans *et al.*,
331 2016).

332 Rdh10 is the major dehydrogenase associated with the first steps of RA production,
333 although the Rdh16 and RdhE2 enzymes can also substitute this function (Belyaeva *et al.*,
334 2009, 2015; Lee *et al.*, 2009). The *O. dioica* genome has no orthologues for either *Rdh10* or
335 *Rdh16* but it does have four genes that encode for RdhE2 proteins (Martí-Solans *et al.*, 2016).
336 *O. dioica* also lacks both an *Aldh1*-type gene as well as a *Cyp26* gene but has a single RXR-
337 orthologue (Table S4, (Martí-Solans *et al.*, 2016)). In contrast, the *C. intestinalis* genome,
338 contains single copies of *Rdh10*, *Rdh16* and *RdhE2* genes and a total of four *Aldh1* genes,
339 located on two chromosomes (Canestro et al., 2006). Consistent with *C. intestinalis*; *M.*
340 *oculata*, *B. leachii* and *B. schlosseri* genomes all have single copies of *Rdh10, Rdh16* and *RdhE2*
341 genes, as well as three *Aldh1a/b* genes on separate scaffolds (Table S4).

342 Three retinoic acid receptor genes were identified within the *B. leachii* genome, one of
343 which had been cloned previously (*g03013,* (Rinkevich, Paz, *et al.*, 2007). All three were also
344 found in *C. intestinalis*, *M. oculata* and *B. schlosseri* genomes(Table S4). While there is only one
345 potential *Cyp26* gene in *M. oculata*, four paralogues were identified in *B. leachii* and *B.*
346 *schlosseri*. A phylogenetic analysis showed that these 4 genes group with CYP26 proteins (Fig.

347 7B, Table S4). Altogether, these results show a loss of key RA-pathway genes in *O. dioica*

348 (*Rdh10, Rdh16, Cyp26* and *Aldh1a*) , while increased copy numbers in other tunicate genomes.

349 # Discussion

350

351 **Genomic diversity within the Solidobranchia**

352 The *B. leachii* genome, along with previous genomic analyses of other ascidian species,
353 support the widely held view that ascidian genomes are diverse and rapidly evolving, which is
354 particularly evident in the Solidobranchia group (Seo *et al.*, 2001; Dehal *et al.*, 2002b;
355 Tsagkogeorga *et al.*, 2010, 2012; Bock *et al.*, 2012; Rubinstein *et al.*, 2013; Voskoboynik *et al.*,
356 2013; Griggio *et al.*, 2014; Stolfi *et al.*, 2014). Nevertheless, botryllids are sufficiently similar
357 in external appearance and morphology for early researchers to have suggested that
358 *Botrylloides* could be a subgenus of *Botryllus* (Saito *et al.*, 2001; Nydam *et al.*, 2017).
359 Strikingly however, the *B. schlosseri* genome differs from that of *B. leachii*, as well as from
360 other sequenced tunicate genomes (Table 2). In particular, the comparison between the *B.*
361 *leachii* and *B. schlosseri* genome sizes (194 Mb vs 725 Mb), their fraction of repetitive
362 sequences (18 % vs 60 %; 65 % in (Voskoboynik *et al.*, 2013)) and their predicted gene
363 number (15'839 vs 27'000; (Voskoboynik *et al.*, 2013)) suggest a different genomic
364 architecture. Altogether, these comparisons indicate that the *B. schlosseri* genome has
365 undergone a significant increase in its genomic content, including retrotransposons (Table
366 S1). In particular, there are at least two additional families in the *B. schlosseri* hAT transposon
367 superfamily and counts of common hAT element, such as hAT-Charlie, differ dramatically (for
368 hAT-Charlie: 366 in *B. leachii* vs 46,661 in *B. schlosseri*). DNA methylation is a key suppressor
369 of transposon activity, changes to the methylation of transposable elements is a known driver
370 of increased tranposition (O'Neill *et al.*, 1998; Simmen *et al.*, 1999; Suzuki *et al.*, 2007;
371 Maumus and Quesneville, 2014). However, DNA methylation in tunicate species has been only
372 studied in *C. intestinalis,* and is described as mosaic, gene body methylation, whereas non-
373 coding regions, including transposons, remain unmethylated (Suzuki et al. 2007). However,

374 DNA methylation has not been studied in other tunicates, and it It is unknown how
375 retrotransposons are suppressed in tunicate genomes. Nevertheless, the observed increase in
376 transpositions could be a consequence of low non-coding DNA methylation, which may
377 contribute to the rapid genome evolution observed in tunicate species, even between closely
378 related species such as *B. schlosseri* and *B. leachii*.

379 Rapid genome evolution, and active transposable elements in particular, are proposed
380 to aid adaptation to new environments for invasive species (Stapley *et al.*, 2015). Differences
381 in the colonization ability of tunicates has been noted, not only between related species such
382 as *B. leachii* and *B. schlosseri* (Brunetti, 1974, 1976; Brunetti *et al.*, 1980), but even at the
383 molecular level within *B. schlosseri* populations (Bock *et al.*, 2012; Nydam *et al.*, 2017). It is
384 thus possible that the observed success in tunicate invasion (Zhan *et al.*, 2015) is supported
385 by their plasticity in genome characteristics like transposon diversity and gene number.

386 Ancient homeobox gene clusters whose structure has been retained over millions of
387 years of evolution in many organisms are fragmented in tunicate genomes. Because, the
388 expression of each *Hox* genes across the anterior-posterior axis relates to their location within
389 the *Hox* gene cluster (Pascual-Anaya *et al.*, 2013), cluster breaks are predicted to have
390 consequences for patterning processes. However, an adult body plan with correct spatial
391 orientation of its body axes during tissue development in ascidians also needs to be
392 established during sexual, asexual and WBR. Early patterning events in tunicate species have
393 only been characterized during sexual reproduction in *Ciona*. Early stages of development
394 (prior to gastrulation) follow a mosaic pattern of developmental axis formation, where
395 inheritance of maternally provided factors establishes the body axes (Nishida, 2005). *Hox*
396 gene knockdown experiments in *C. intestinalis* revealed that they have very limited roles, with
397 defects only observed in larval neuronal and tail development upon loss of *Ci-Hox11* and *Ci-*
398 *Hox12* function (Ikuta *et al.*, 2010). It thus appears that patterning events in *C. intestinalis* are

399 less dependent upon anterior-posterior spatial expression of *Hox* genes to establish regional
400 identity. Previously, in *B. schlosseri*, the entry point of the connective test vessel into the
401 developing bud determines the posterior end of the new zooid (Sabbadin *et al.*, 1975).
402 Therefore it is possible that ascidians incorporate environmental and physical cues to
403 compensate for the lost gene cluster during polarity establishment. A wider analysis
404 comprising multiple tunicate species will be necessary to investigate the exact consequences
405 of homeobox cluster dispersion and whether the compensatory mechanism observed in *C.*
406 *intestinalis* is the norm or an exception.
407

408 **Lineage-specific changes to evolutionarily conserved cell communication pathways**

409 Cell signaling pathways are critical for morphogenesis, development and adult
410 physiology. In particular, we have focused our analysis on three highly conserved pathways:
411 Wnt, Notch and Retinoic Acid signaling. Representatives of all twelve *Wnt* genes subfamilies
412 are found in metazoans, suggesting that they evolved before evolution of the bilaterians
413 (Janssen *et al.*, 2010). We identified members of each Wnt subfamily in tunicate genomes,
414 along with numerous examples of lineage-specific gene loss and/or duplication. The most
415 striking of these events was an increase in *Wnt5a* gene copy number in *B. leachii*, *B. schlosseri*
416 and *M. oculata* genomes. Indeed, most invertebrates genomes, including the basal chordate *B.*
417 *floridae*, contain a single *Wnt5* gene while most vertebrate genomes have two *Wnt5a*
418 paralogues, believed to be a result of whole genome duplication (Martin *et al.*, 2012).
419 However, in the analyzed tunicate genomes, up to 15 copies of this gene were identified,
420 potentially these additional genes may have been co-opted into novel roles and were retained
421 during tunicate evolution. Wnt5a ligands have numerous biological roles, including a
422 suppressive one during zebrafish regeneration (Stoick-Cooper *et al.*, 2007) and a promotive
423 one during amphioxus regeneration (Somorjai *et al.*, 2012). Furthermore, components of both

424 Wnt signaling pathways are differentially expressed during WBR (Zondag *et al.*, 2016).
425 Altogether, it is thus conceivable that *Wnt5a* gene number has expanded in colonial tunicates
426 to sustain WBR. A functional characterization of the role of these numerous copies of Wnt5a
427 would thus be highly interesting and potentially give evolutionary insights into chordate
428 regeneration.

429 All components of the Notch pathway are present in the genomes we investigated. Of
430 particular interest, the DSL Notch ligandappears to be rapidly evolving in the tunicates. This
431 indicates that tunicate DSL proteins are under less pressure, than vertebrate orthologous
432 proteins, to conserve their protein sequence. Given that the concentration and activity of the
433 DSL ligands are typically the rate limiting step in this key signaling pathway, it will be
434 interesting to assess whether the functional properties of tunicate proteins have adapted
435 accordingly.

436 Components of the RA signaling pathway have also been identified in all the tunicate
437 genomes. However, *Oikopleura* has seemingly lost a functional RA synthesis pathway, while
438 still forming a functional body plan. This suggests that either uniquely RA is not involved in
439 critical developmental events in this species, that the RA signaling function has been replaced
440 or that *O. dioica* utilizes an alternative synthesis approach. Conversely, lineage specific
441 increases in RA pathway gene numbers have been observed in *C. intestinalis* (Aldh1, (Sobreira
442 *et al.*, 2011)) and botryllids (*CYP26* genes).

443 RA, Notch and Wnt pathways play roles in regeneration and development in many
444 species, including Solidobranchian tunicates (Rinkevich, Paz, *et al.*, 2007; Rinkevich *et al.*,
445 2008; Zondag *et al.*, 2016) and *Cionidae (Hamada et al., 2015; Jeffery, 2015)*. The observed loss
446 of RA signaling genes may result in a poorer regeneration ability for *O. dioica*, however it's
447 regenerative has not been characterized. Given the unique chordate WBR potential developed
448 by colonial tunicates, it is conceivable that there is selective pressure on their genomes to

449 retain these pathways. In particular given the likely role these pathways also play in colony
450 reactivation following hibernation, as well as in asexual reproduction.

451 In addition, among tunicates there are significant differences in both life cycle,
452 reproduction and regeneration ability, even between closely related species of the same
453 family, that likely reflect an underlying diversity in genomic content. For instance, differences
454 in both asexual and sexual reproduction have been observed between within the Botryllidae
455 family (Berrill, 1941, 1947, 1951; Oka and Watanabe, 1957; Brunetti, 1974, 1976).
456 Furthermore, *B. schlosseri* can only undergo WBR during a short time frame of their asexual
457 reproductive cycle when the adults are reabsorbed by the colony (Voskoboynik *et al.*, 2007;
458 Kürn *et al.*, 2011) while *B. leachii* can undergo WBR throughout their adult life (Rinkevich,
459 Paz, *et al.*, 2007). Overall, this indicates that despite a generally similar appearance, the rapid
460 evolution of the Tunicata subphylum has provided diversity and innovations within its
461 species. It will thus be particularly interesting to investigate in future studies how such
462 genomic plasticity balances between adaptation to new challenges and constraint, preserving
463 common morphological features.

464

465 In conclusion, our assembly of the *B. leachii* genome provides an essential resource for
466 the study of this colonial ascidian as well as a crucial point of comparison to gain further
467 insights into the remarkable genetic diversity among tunicate species. In addition, the
468 genome of *B. leachii* will be most useful for dissecting WBR in chordates, in particular by
469 comparing it with that of *B. schlosseri* to understand how the initiation of WBR can be blocked
470 during specific periods of their life cycle. Furthermore, given the key phylogenetic position of
471 Tunicates with respect to vertebrates, the analysis of their genomes will provide important
472 insights in the emergence of chordate traits and the origin of vertebrates.

Methods

Sampling, library preparation and sequencing

B. leachii colonies were collected from Nelson harbour (latitude 41.26°S, longitude 173.28°E) in New Zealand. To reduce the likelihood of contamination, embryos were dissected out of a colony and pooled before carrying out DNA extraction using E.Z.N.A SP Plant DNA Mini Kit. A total of 2 µg each was sent to New Zealand Genomics Limited (NZGL) for library preparation and sequencing. A short read (TruSeq illumina HiSeq) generated 19090212 paired-end reads of 100 bp (average fragment size: 450bp, adaptor length: 120bp). A second library (Illumina Nextera MiSeq Mate Pair) not size-selected library (fragment size: 1.5-15 kb, median size: ~3 kb, adaptor length: 38bp).generated 31780788 paired-end sequences of 250 bp.

PreQC report was generated using the String Graph Assembler software package (Simpson, 2014) and quality metrics before assembly with both FastQC (Andrews, 2010) as well as MultiQC (Ewels *et al.*, 2016) (Fig. S1). These analyses revealed that 91 % of sequences had a mean Phred quality score >= 30, 96 % of bases a mean Phred quality score >= 30, and 39 % of sequences an adapter sequence (either Illumina or Nextera). Adaptor trimming was performed with NxTrim (O'Connell *et al.*, 2015) for the mate pair library, followed by Trimmomatic (Bolger *et al.*, 2014) with the following options: MINLEN:40 ILLUMINACLIP:2:30:12:1:true LEADING:3 TRAILING:3 MAXINFO:40:0.4 MINLEN:40 for both libraries. After trimming, 86644308 paired-end (85 %) and 12112004 (12 %) single-end sequences remained (100% with a mean Phred quality score >= 30, <1% with an adapter sequence).

Genome assembly

498 *De novo* assembly was performed in three consecutive iterations following a Meta-
499 assembly approach (Table S5). First, both libraries were assembled together in parallel, using
500 a k-mer size of 63 following the results from KmerGenie (Chikhi and Medvedev, 2014)
501 whenever available, by five assemblers: AbySS (Simpson *et al.*, 2009), Velvet (Zerbino and
502 Birney, 2008), SOAPdenovo2 (Luo *et al.*, 2012), ALLPATHS-LG (Gnerre *et al.*, 2011), MaSuRCA
503 (Zimin *et al.*, 2013). The MaSuRCA assembler was run twice, once running the adapter
504 filtering function (here termed "MaSuRCA-filtered"), the other without (termed simply
505 "MaSuRCA"). Their respective quality was then estimated using three different metrics: the
506 N50 length, the BUSCO core-genes completion (Simão *et al.*, 2015) and the Glimmer number of
507 predicted genes (Delcher *et al.*, 1999). Second, these drafts were combined by following each
508 ranking using Metassembler (Wences and Schatz, 2015), hence producing three new
509 assemblies (limiting the maximum insert size at 15 kb). Third, the *B. leachii* transcriptome
510 (Zondag *et al.*, 2016) was aligned to each meta-assembly using STAR (Dobin *et al.*, 2013),
511 which were then combined thrice more using Metassembler following their alignment
512 percentage and limiting the maximum insert size at 3 kb, 8 kb and 15 kb. Finally, the quality of
513 the meta-meta-assemblies was estimated using the BUSCO score and the best one (Table S5)
514 selected as the reference *de novo* assembly.
515

516 **Data access**

517 *B. leachiii*: ANISEED http://www.aniseed.cnrs.fr/fgb2/gbrowse/boleac_v3/
518 *C. intestinalis*: Ciona_intestinalis.KH.cds.all.fa, Ciona_intestinalis.joinedscaffold.fa Ciona
519 unmasked v2.0, https://www.aniseed.cnrs.fr/fgb2/gbrowse/ciona_intestinalis/
520 C. savigny: *C. savigni*_paired_scaffolds.fa,
521 https://www.aniseed.cnrs.fr/fgb2/gbrowse/cisavi_ens81/,
522 *M. oculata*: , https://www.aniseed.cnrs.fr/fgb2/gbrowse/moocul_elv12/,

523 *B. schlosseri*: botznik-transcripts.fa, botznik-chr.fa, botznik-ctg.fa,

524 https://www.aniseed.cnrs.fr/fgb2/gbrowse/boschl_botznik2013/,

525 *O. dioica*: Odioica_reference_v3.fa, Oikopleura_transcripts_reference_v1.0.fa,

526 http://www.genoscope.cns.fr/externe/GenomeBrowser/Oikopleura/.

527

528 ***Repeat region analysis***

529 A *de novo* repeat library was build for each tunicate genome using RepeatModeler

530 (Smit and Hubley, 2015). This utilizes the RECON tandem repeats finder from the RepeatScout

531 packages to identify species-specific repeats in a genome assembly. RepeatMasker (Smit *et al.*,

532 2015) was then used to mask those repeats.

533

534 **Gene annotation**

535 *Ab initio* genome annotation was performed using MAKER2 (Holt and Yandell, 2011)

536 with Augustus (Stanke and Waack, 2003) and SNAP (Korf, 2004) for gene prediction. In

537 addition, we used our previously published transcriptome (Zondag *et al.*, 2016) and a

538 concatenation of UniProtKB (UniProt Consortium, 2015), *C. intestinalis* and *B. schlosseri*

539 proteins into a custom proteome as evidence of gene product. Using the predicted genes,

540 Augustus and SNAP were then trained to the specificity of *B. leachii* genome. A second round

541 of predictions was then performed, followed by a second round of training. The final

542 annotation of the genome was obtained after running a third round of predictions, and the

543 provided trained Augustus and SNAP configurations after a third round of training.

544 The following databases were used for homology identification: InterProScan (Jones *et al.*,

545 2014) tRNA (Lowe and Eddy, 1997) snoRNA (Lowe, 1999) Infernal (Nawrocki and Eddy,

546 2013) Rfam (Nawrocki *et al.*, 2015).

547

548 **Analysis of specific gene families**

549 Genes and transcripts for each examined genome were identified by a tBLASTn search
550 with an e-value cut-off at 10^{-5} using the SequencerServer software (Priyam *et al.*, 2015). This
551 was followed by a reciprocal BLAST using SMART blast
552 (http://blast.ncbi.nlm.nih.gov/smartblast/smartBlast.cgi?CMD=Web) to confirm their
553 identity.

554 Delta serrate ligand conserved protein domain (PF01414) was used to identify the
555 corresponding proteins in tunicate genomes**.** To identify *Notch* receptor genes the conserved
556 LNR (lin-notch repeat) domain (PF00066) was used. ALDH-like genes were identified by
557 tBLASTn search (PF00171) and classified using SMART blast.

558

559 **Phylogenetics**

560 Sequences were aligned with ClustalX (Jeanmougin *et al.*, 1998) before using ProtTest
561 3 (Abascal *et al.*, 2005) to determine the best-fit model of evolution. The best-fit model for the
562 DSL phylogeny was WAG+I+G and, for CYP26 proteins, was LG+I+G.

563 Bayesian inference (BI) phylogenies were constructed by MrBayes (Ronquist and
564 Huelsenbeck, 2003) with a mixed model for 100,000 generations and summarized using a
565 Sump burnin of 200. Maximum Likelihood (ML) phylogenies were generated by PhyML
566 (Guindon *et al.*, 2010), using the appropriate model, estimating the amino acid frequences.

567 Accession numbers are provided in File S3 and sequence alignments are provided in
568 Figure S3. Analyses carried out with BI and ML produced identical tree topologies.
569 Trees were displayed using FigTree v1.4.2 [REF] (http://tree.bio.ed.ac.uk/software/figtree/).

570 # Acknowledgements

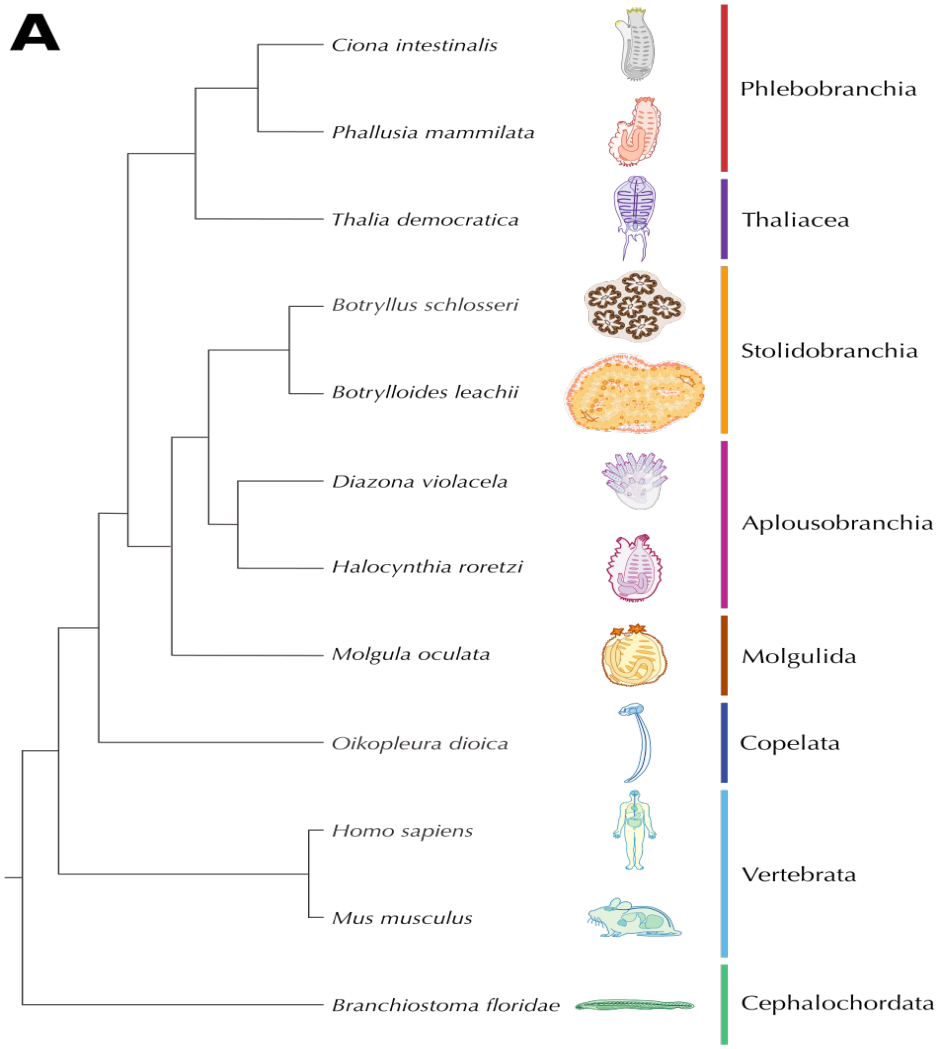
571 Funding support was provided to M.J.W. by the Otago BMS Deans Bequest and
572 Department of Anatomy. S.B. by the Swiss National Science Foundation (SNSF) [grant number
573 P2ELP3_158873]. We would like to thank Peter Maxwell and the New Zealand eScience
574 Infrastructure (NeSI); Christelle Dantec and ANISEED for help and advice during the
575 annotation process, as well as for the accompanying *B. leachii* genome browser.

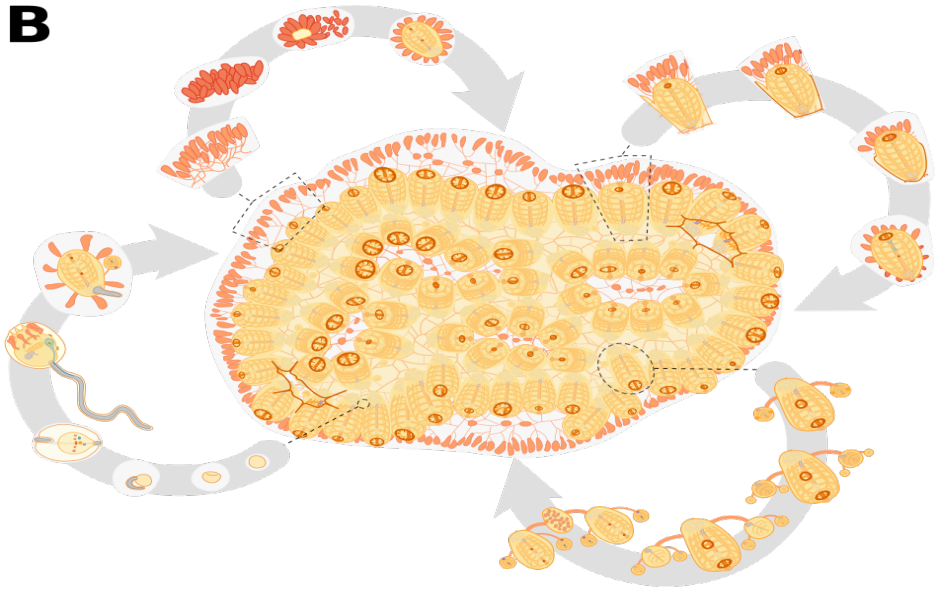
576

Figure Legends

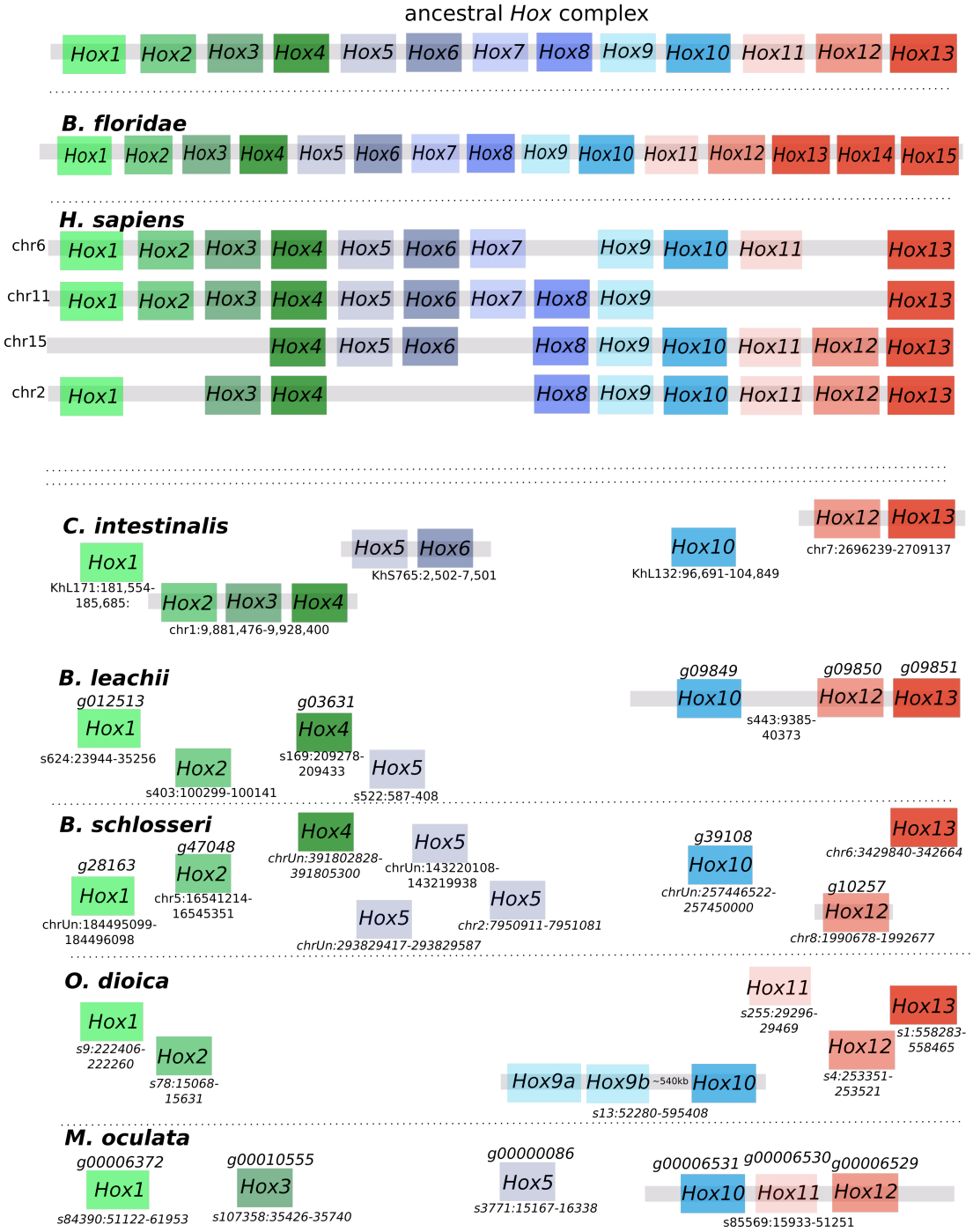
Figure 1 *B. leachii* phylogenetic position and life cycle.

A. Schematic showing phylogeny of tunicates with respect to the chordate clade. **B.** Life cycle of *B. leachii*. The colony expands and grows by asexual reproduction. During favorable conditions such as warmer water temperatures, members of the colonies start sexual reproduction. The embryos remain with the colony in brood pouches until release. Hatched larvae attach to nearby substrates and begin metamorphosis into a zooid.


A
Ciona intestinalis
Phallusia mammilata
Thalia democratica
Botryllus schlosseri
Botrylloides leachii
Diazona violacela
Halocynthia roretzi
Molgula oculata
Oikopleura dioica
Homo sapiens
Mus musculus
Branchiostoma floridae
Phlebobranchia
Thaliacea
Stolidobranchia
Aplousobranchia
Molgulida
Copelata
Vertebrata
Cephalochordata


B

589 **Figure 2. *Hox* genes are dispersed and reduced in number within tunicate genomes.**
590 Schematic depicting remaining *Hox* gene linkages in five tunicate genomes in comparison to
591 the ancestral *Hox* complex, which included thirteen genes. Orthologous genes are indicated by
592 common colours. Chromosome or scaffold number is shown, along with gene ID when
593 available for newly annotated genomes.



594

595

Figure 3. NK homeobox cluster genes are fragmented within tunicate genomes.

Schematic depicting the organization of the *NK homeobox* cluster genes among the studied chordate genomes. Double-parallel lines indicate > 1Mb distance between genes. Chromosome or scaffold number is shown, along with gene ID when available for newly annotated genomes. Orthologous genes are indicated by common colours.

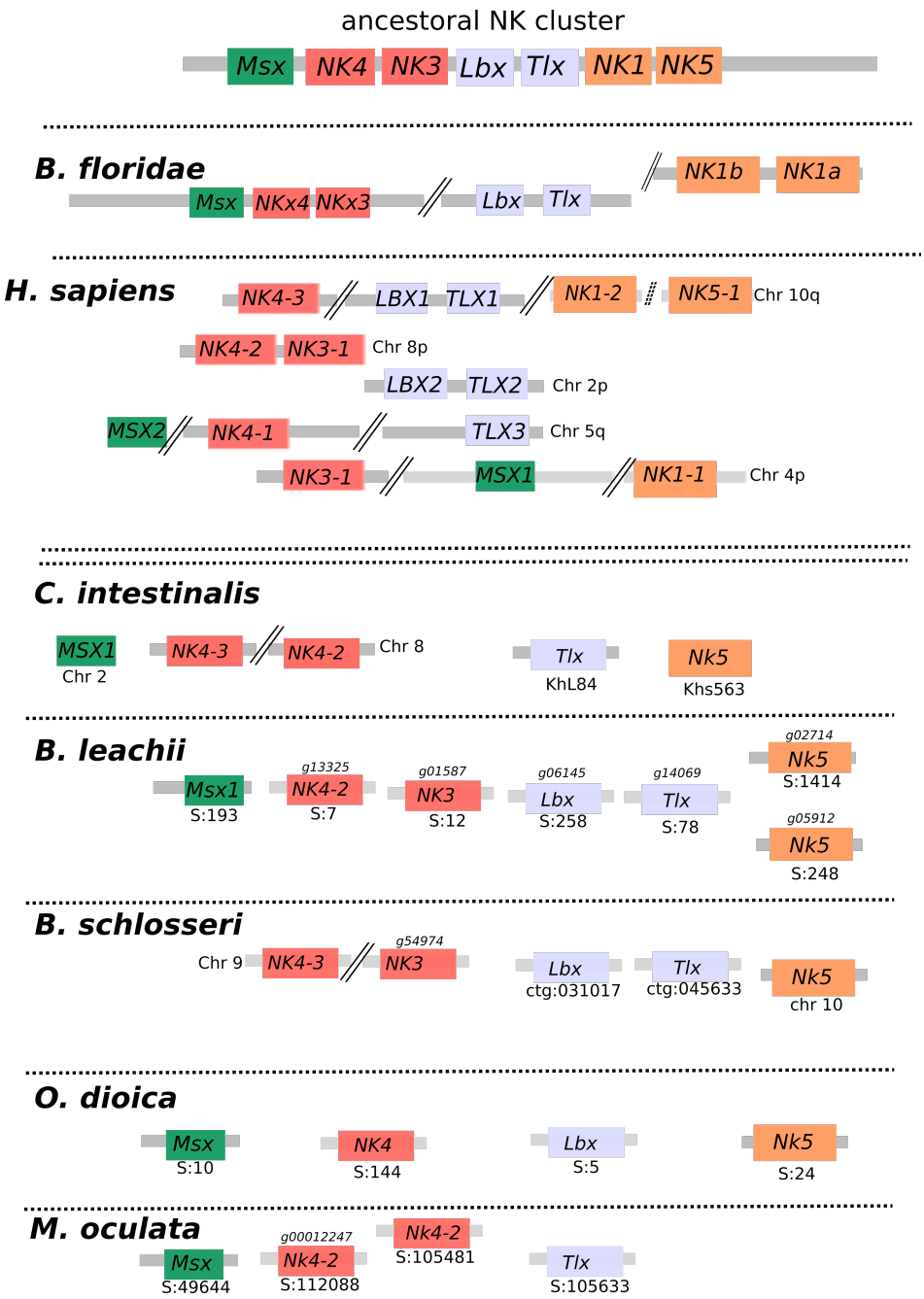
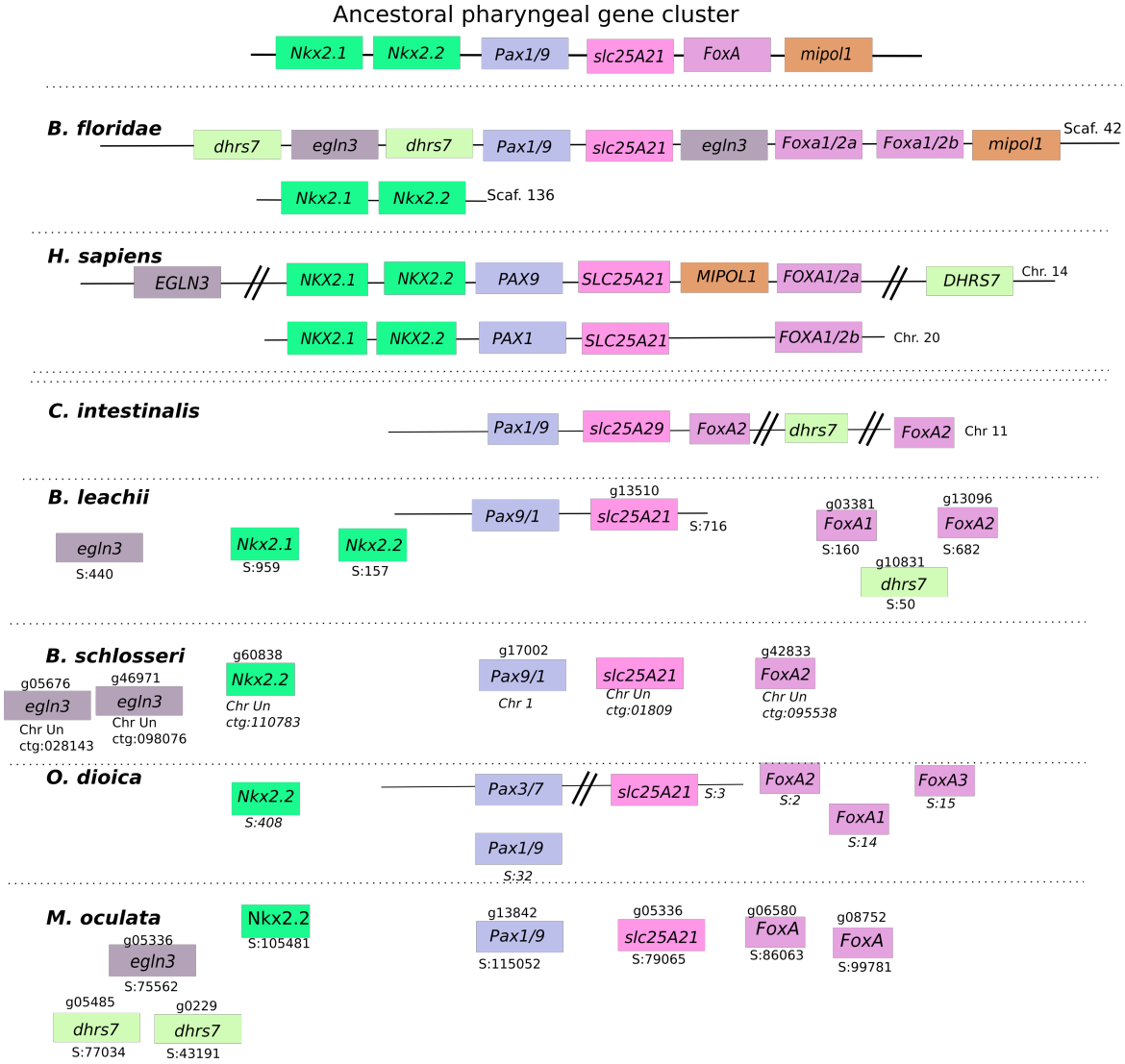


Figure 4. Ancestral gene linkages remain between a few pharyngeal cluster genes in tunicate genomes. Gene order of the six pharyngeal cluster genes, *Nkx2.1, Nk2.2, Pax1/9* and *FoxA* in chordate genomes. Double-parallel lines indicate > 1Mb distance between genes. Chromosome or scaffold number is shown, along with gene ID when available for newly annotated genomes. Orthologous genes are indicated by common colours.



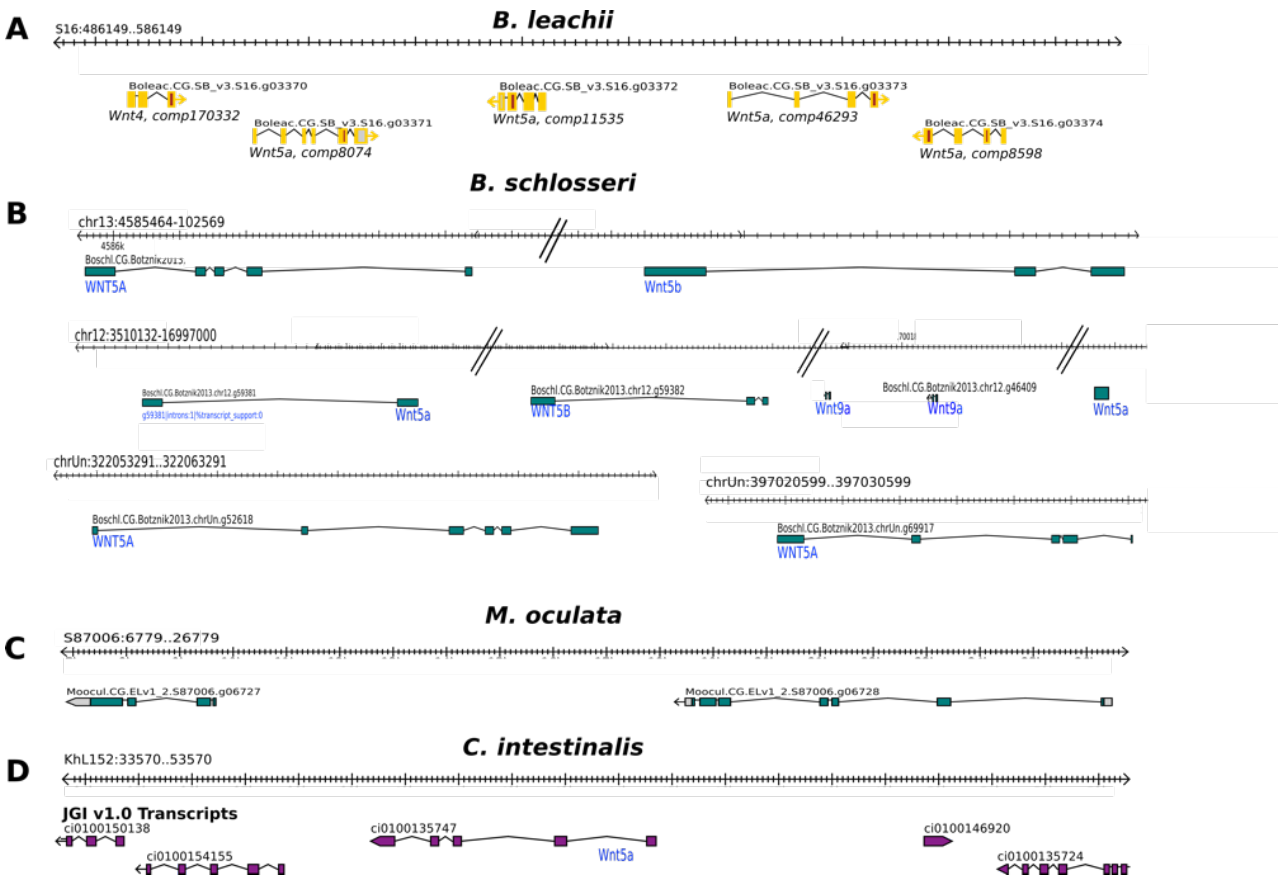
614 **Figure 5. Duplication of *Wnt5a* genes in Tunicate genomes.**

615 Schematic showing the genomic location of *B. leachii* (A), *B. schlosseri* (B), *M. oculata* (C) and *C.*

616 *intestinalis* (D) *Wnt5*-like genes within each genome. Double-parallel lines indicate > 1Mb

617 distance between genes.

618



619

620

621

622

623

624

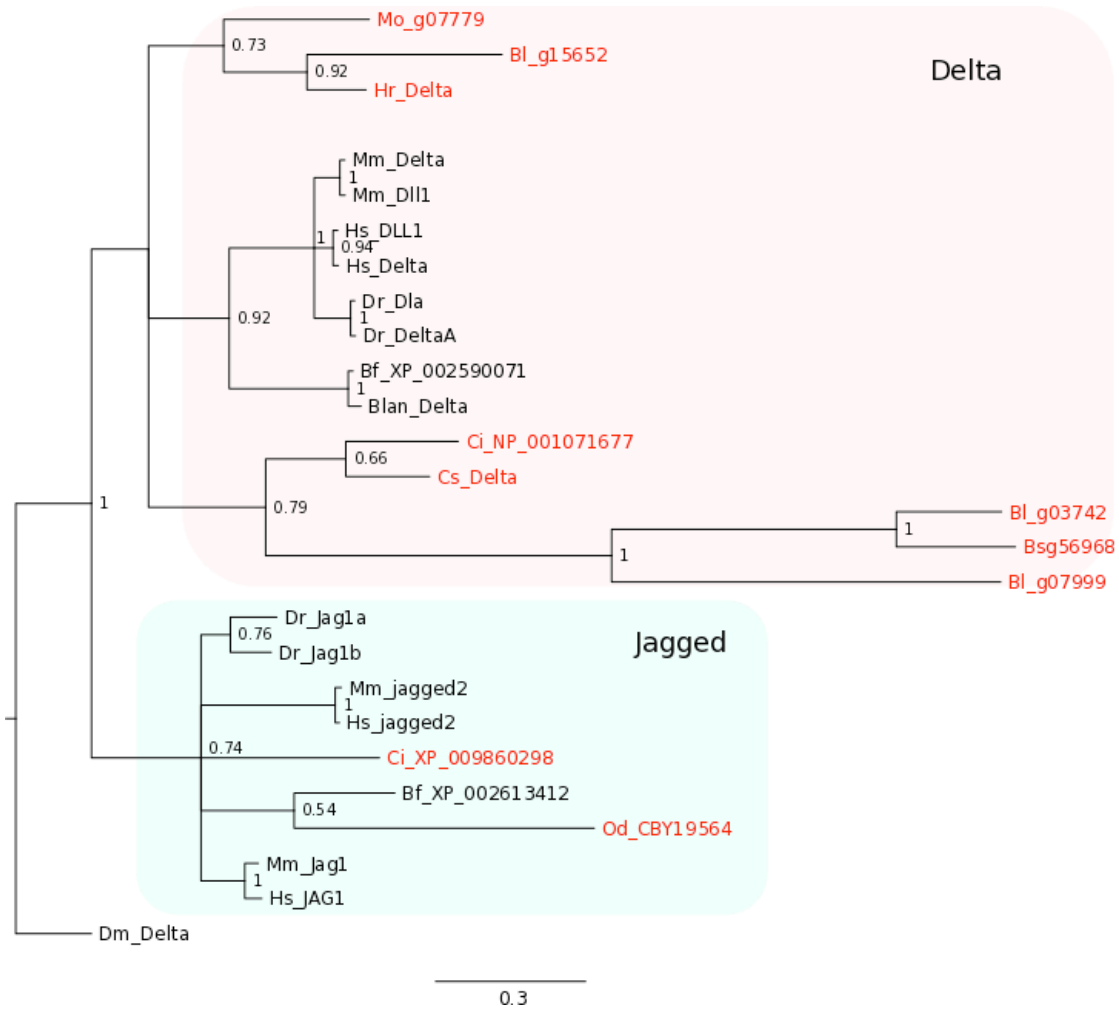
625

626

627 **Figure 6**. ***B. leachii* Notch pathway**

628 Bayesian phylogenetic tree depicting the relationship between tunicate and vertebrate DSL
629 proteins, using *Drosophila* Delta to root the tree. *B. leachii* proteins are shown in red and
630 shaded areas correspond to clade groupings. Branch support values (probabilities) are
631 indicated.

632 **Abbreviations: Bl, *Botrylliodes leachii*; BsS, *Botryllus schlosseri*; Od, *Oikopleura odioica*;**
633 **Hs, *Homo sapiens*; Ci, *Ciona intestinalis*, Dm, *Drosophila melanogaster*; Dr, *Danio rerio;***
634 **Mo, *M. oculata;* Hr *Halocynthia roretzi*; *Blan, Branchiostoma lanceolatum; Bf,***
635 ***Branchiostoma floridae.***



636

637

638

639

640

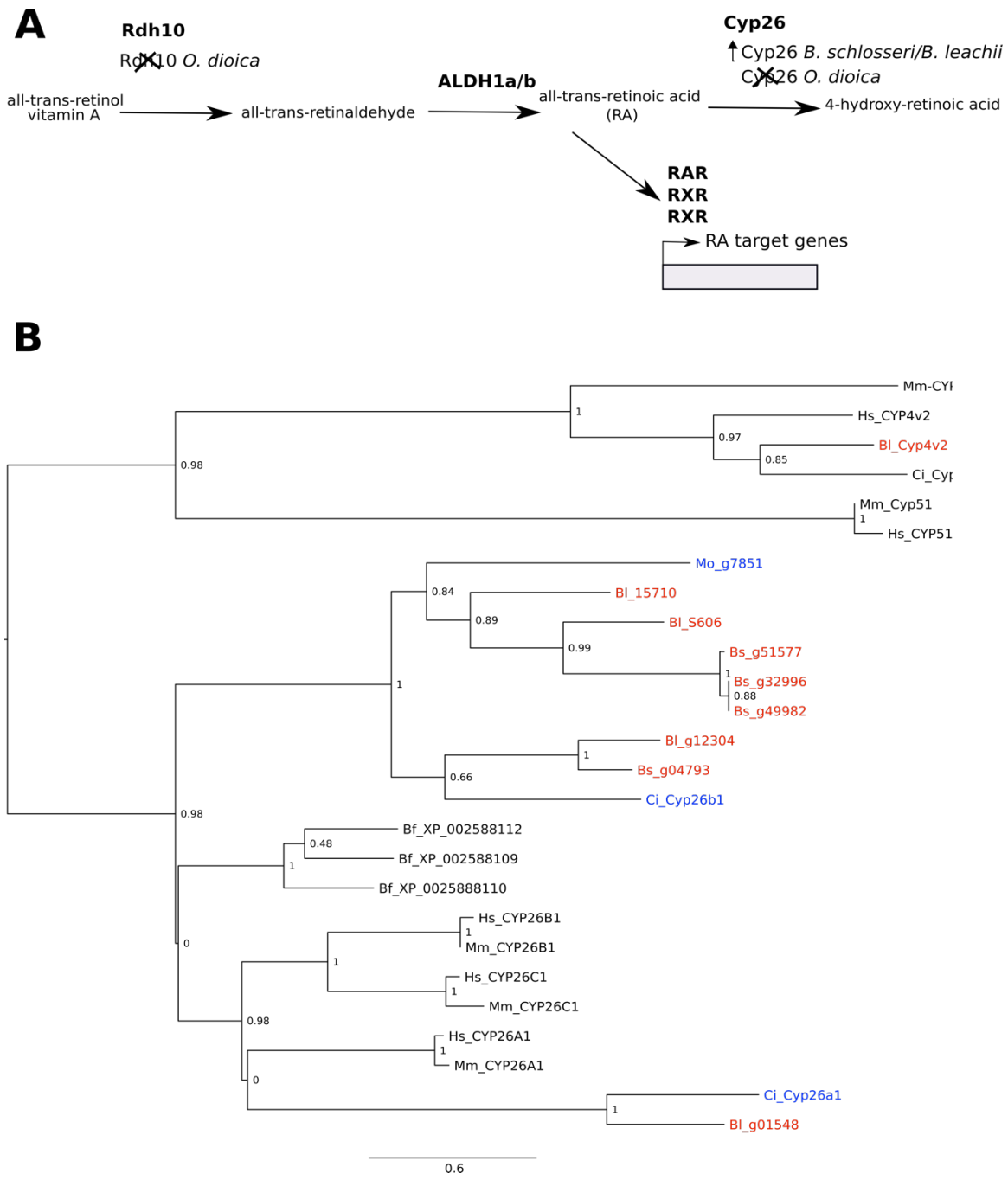
641 **Figure 7**. **Evolution of the RA pathway in tunicates**

642 **(A)** Overview of the RA synthesis and degradation pathway. In bold are the major proteins
643 that contribute to RA signaling during animal development. Indicated below these are
644 changes to the number of copies present in examined genomes. **(B)** ML phylogenetic tree
645 depicting the relationship between invertebrate and vertebrate CYP26 proteins using CYP4
646 and CYP51 proteins as an outgroup. *B. schlosseri* and *B. leachii* proteins are shown in red, *C.*
647 *intestinalis* and *M. oculata* are indicated in blue. No *Cyp26* gene has been identifed in the *O.*
648 *dioica* genome. Values for the approximate likelihood-ratio test (aLRT) are indicated.

649 **Abbreviations: Bl, *Botrylliodes leachii*; Bs, *Botryllus schlosseri*; Hs, *Homo sapiens*; Ci,**
650 ***Ciona intestinalis;* Mo*, M. oculata.***

651

652



653

654

655

656

657 # References

658 **Abascal, F., R. Zardoya and D. Posada**. **2005**. ProtTest: selection of best-fit models of protein
659 evolution. *Bioinformatics* **21**: 2104–2105.

660 **Andrews, S. 2010**. FastQC: a quality control tool for high throughput sequence data.

661 **Auger, H., Y. Sasakura, J.-S. Joly and W. R. Jeffery**. **2010**. Regeneration of oral siphon pigment
662 organs in the ascidian Ciona intestinalis. *Dev. Biol.* **339**: 374–389.

663 **Ballarin, L., A. Franchini, E. Ottaviani and A. Sabbadin**. **2001**. Morula cells as the major
664 immunomodulatory hemocytes in ascidians: Evidences from the colonial species *Botryllus*
665 *schlosseri*. *Biol. Bull.* **201**: 59–64.

666 **Belyaeva, O. V., C. Chang, M. C. Berlett and N. Y. Kedishvili**. **2015**. Evolutionary origins of
667 retinoid active short-chain dehydrogenases/reductases of SDR16C family. *Chem. Biol.*
668 *Interact.* **234**: 135–143.

669 **Belyaeva, O. V., S.-A. Lee, O. V. Kolupaev and N. Y. Kedishvili**. **2009**. Identification and
670 characterization of retinoid-active short-chain dehydrogenases/reductases in Drosophila
671 melanogaster. *Biochim. Biophys. Acta - Gen. Subj.* **1790**: 1266–1273.

672 **Berna, L. and F. Alvarez-Valin**. **2014**. Evolutionary genomics of fast evolving tunicates. *Genome*
673 *Biol. Evol.* **6**: 1724–1738.

674 **Berrill, N. J. 1941**. The development of the bud in *Botryllus*. *Biol. Bull.* **80**: 169.

675 **Berrill, N. J. 1947**. The developmental cycle of *Botrylloides*. *Q. J. Microsc. Sci.* **88**: 393–407.

676 **Berrill, N. J. 1951**. Regeneration and Budding in Tunicates. *Biol. Rev.* **26**: 456–475.

677 **Bock, D. G., H. J. MacIsaac and M. E. Cristescu**. **2012**. Multilocus genetic analyses differentiate
678 between widespread and spatially restricted cryptic species in a model ascidian. *Proc. R. Soc.*
679 *B Biol. Sci.* **279**: 2377–2385.

680 **Bolger, A. M., M. Lohse and B. Usadel**. **2014**. Trimmomatic: a flexible trimmer for Illumina
681 sequence data. *Bioinformatics* **30**: 2114–2120.

682 **Bradnam, K. R., J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J. A.**
683 **Chapman, G. Chapuis, R. Chikhi, *et al.* 2013**. Assemblathon 2: evaluating de novo methods
684 of genome assembly in three vertebrate species. *Gigascience* **2**: 10.

685 **Brown, F. D. and B. J. Swalla**. **2012**. Evolution and development of budding by stem cells:
686 Ascidian coloniality as a case study. *Dev. Biol.* **369**: 151–162. Elsevier.

687 **Brozovic, M., C. Martin, C. Dantec, D. Dauga, M. Mendez, P. Simion, M. Percher, B. Laporte,**
688 **C. Scornavacca, A. Di Gregorio, *et al.* 2016**. ANISEED 2015: A digital framework for the
689 comparative developmental biology of ascidians. *Nucleic Acids Res.* **44**: D808–D818.

690 **Brunetti, R. 1974**. *Observations on the Life Cycle of Botryllus Schlosseri (Pallas) (Ascidiacea) in*
691 *the Venetian Lagoon*.

692 **Brunetti, R. 1976**. Biological cycle of *Botrylloides leachi* (Savigny) (Ascidiacea) in the Venetian
693 lagoon. *Vie Milieu* **XXVI**: 105–122.

694 **Brunetti, R., L. Beghi, M. Bressan and M. Marin**. **1980**. Combined effects of temperature and
695 salinity on colonies of *Botryllus schlosseri* and *Botrylloides leachi* (Ascidiacea) from the
696 Venetian Lagoon. *Mar. Ecol. Prog. Ser.* **2**: 303–314.

697 **Burighel, P., R. Brunetti and G. Zaniolo**. **1976**. Hibernation of the colonial ascidian *Botrylloides*
698 *leachi* (Savigny): histological observations. *Ital. J. Zool.* **43**: 293–301.

699 **Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L.**
700 **Madden**. **2009**. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

701 **Canestro, C., J. H. Postlethwait, R. Gonzalez-Duarte and R. Albalat**. **2006**. Is retinoic acid
702 genetic machinery a chordate innovation? *Evol. Dev.* **8**: 394–406.

703 **Chikhi, R. and P. Medvedev**. **2014**. Informed and automated k-mer size selection for genome
704 assembly. *Bioinformatics* **30**: 31–37.

705 **Cunningham, T. J. and G. Duester**. **2015**. Mechanisms of retinoic acid signalling and its roles in
706 organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16**: 110–123.

707 **Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A.**
708 **Di Gregorio, M. Gelpke, D. M. Goodstein, *et al.* 2002a**. The Draft Genome of Ciona
709 intestinalis: Insights into Chordate and Vertebrate Origins. *Science (80-.).* **298**: 2157–2167.

710 **Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. W. De Tomaso, B. Davidson,**
711 **A. Di Gregorio, M. Gelpke, D. M. Goodstein, *et al.* 2002b**. The draft genome of *Ciona*
712 *intestinalis*: insights into chordate and vertebrate origins. *Science* **298**: 2157–67.

713 **Delcher, A. L., D. Harmon, S. Kasif, O. White and S. L. Salzberg**. **1999**. Improved microbial
714 gene identification with GLIMMER. *Nucleic Acids Res.* **27**: 4636–41.

715 **Delsuc, F., H. Brinkmann, D. Chourrout and H. Philippe**. **2006**. Tunicates and not
716 cephalochordates are the closest living relatives of vertebrates. *Nature* **439**: 965–968.

717 **Dobin, A., C. a Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson**
718 **and T. R. Gingeras**. **2013**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:
719 15–21.

720 **Edvardsen, R. B., H.-C. Seo, M. F. Jensen, A. Mialon, J. Mikhaleva, M. Bjordal, J. Cartry, R.**
721 **Reinhardt, J. Weissenbach, P. Wincker, *et al.* 2005**. Remodelling of the homeobox gene
722 complement in the tunicate Oikopleura dioica. *Curr. Biol.* **15**: R12–R13.

723 **Ewels, P., M. Magnusson, S. Lundin and M. Käller**. **2016**. MultiQC: Summarize analysis results
724 for multiple tools and samples in a single report. *Bioinformatics* btw354.

725 **Franchi, N., F. Schiavon, M. Carletto, F. Gasparini, G. Bertoloni, S. C. E. Tosatto and L.**
726 **Ballarin**. **2011**. Immune roles of a rhamnose-binding lectin in the colonial ascidian Botryllus
727 schlosseri. *Immunobiology* **216**: 725–736.

728 **Garcia-Fernàndez, J. 2005**. The genesis and evolution of homeobox gene clusters. *Nat. Rev.*
729 *Genet.* **6**: 881–92.

730 **Gasparini, F., P. Burighel, L. Manni and G. Zaniolo**. **2008**. Vascular regeneration and
731 angiogenic-like sprouting mechanism in a compound ascidian is similar to vertebrates. *Evol.*
732 *Dev.* **10**: 591–605.

733 **Gazave, E., P. Lapébie, G. S. Richards, F. Brunet, A. V Ereskovsky, B. M. Degnan, C.**
734 **Borchiellini, M. Vervoort and E. Renard**. **2009**. Origin and evolution of the Notch signalling
735 pathway: an overview from eukaryotic genomes. *BMC Evol. Biol.* **9**: 249.

736 **Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe,**
737 **G. Hall, T. P. Shea, S. Sykes, *et al.* 2011**. High-quality draft assemblies of mammalian
738 genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci.* **108**: 1513–1518.

739 **Griggio, F., A. Voskoboynik, F. Iannelli, F. Justy, M.-K. M.-K. Tilak, T. Xavier, G. Pesole, E.**
740 **J. P. Douzery, F. Mastrototaro and C. Gissi**. **2014**. Ascidian Mitogenomics: Comparison of
741 Evolutionary Rates in Closely Related Taxa Provides Evidence of Ongoing Speciation Events.
742 *Genome Biol. Evol.* **6**: 591–605.

743 **Guder, C., I. Philipp, T. Lengfeld, H. Watanabe, B. Hobmayer and T. W. Holstein**. **2006**. The
744 Wnt code: cnidarians signal the way. *Oncogene* **25**: 7450–7460.

745 **Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel**. **2010**. New
746 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the
747 Performance of PhyML 3.0. *Syst. Biol.* **59**: 307–321.

748 **Guruharsha, K. G., M. W. Kankel and S. Artavanis-Tsakonas**. **2012**. The Notch signalling
749 system: recent insights into the complexity of a conserved pathway. *Nat. Rev. Genet.* **13**: 654–
750 666.

751 **Hamada, M., S. Goricki, M. S. Byerly, N. Satoh and W. R. Jeffery**. **2015**. Evolution of the
752 chordate regeneration blastema: Differential gene expression and conserved role of notch
753 signaling during siphon regeneration in the ascidian Ciona. *Dev. Biol.* **405**: 304–315.

754 **Hino, K., Y. Satou, K. Yagi and N. Satoh**. **2003**. A genomewide survey of developmentally
755 relevant genes in Ciona intestinalis. *Dev. Genes Evol.* **213**: 264–272.

756 **Hoegg, S. and A. Meyer**. **2005**. Hox clusters as models for vertebrate genome evolution. *Trends*
757 *Genet.* **21**: 421–424.

758 **Holt, C. and M. Yandell**. **2011**. MAKER2: an annotation pipeline and genome-database
759 management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.

760 **Ikuta, T., N. Satoh and H. Saiga**. **2010**. Limited functions of Hox genes in the larval development
761 of the ascidian Ciona intestinalis. *Development* **137**: 1505–1513.

762 **Jackson, B., C. Brocker, D. C. Thompson, W. Black, K. Vasiliou, D. W. Nebert and V.**
763 **Vasiliou**. **2011**. Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Hum.*
764 *Genomics* **5**: 283–303.

765 **Janssen, R., M. Le Gouar, M. Pechmann, F. Poulin, R. Bolognesi, E. E. Schwager, C. Hopfen,**
766 **J. K. Colbourne, G. E. Budd, S. J. Brown, *et al.* 2010**. Conservation, loss, and redeployment
767 of Wnt ligands in protostomes: implications for understanding the evolution of segment
768 formation. *BMC Evol. Biol.* **10**: 374.

769 **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins and T. J. Gibson**. **1998**. Multiple
770 sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**: 403–5.

771 **Jeffery, W. R. 2015**. Regeneration, Stem Cells, and Aging in the Tunicate Ciona: Insights from the
772 Oral Siphon. *Int. Rev. Cell Mol. Biol.* **319**: 255–82.

773 **Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen,**
774 **A. Mitchell, G. Nuka, *et al.* 2014**. InterProScan 5: Genome-scale protein function
775 classification. *Bioinformatics* **30**: 1236–1240.

776 **Kent, W. J. 2002**. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**: 656–64.

777 **Korf, I. 2004**. Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59.

778 **Kürn, U., S. Rendulic, S. Tiozzo and R. J. Lauzon**. **2011**. Asexual propagation and regeneration
779 in colonial ascidians. *Biol. Bull.* **221**: 43–61.

780 **Kusserow, A., K. Pang, C. Sturm, M. Hrouda, J. Lentfer, H. A. Schmidt, U. Technau, A. von**
781 **Haeseler, B. Hobmayer, M. Q. Martindale, *et al.* 2005**. Unexpected complexity of the Wnt
782 gene family in a sea anemone. *Nature* **433**: 156–160.

783 **Lauzon, R. J., C. Brown, L. Kerr and S. Tiozzo**. **2013**. Phagocyte dynamics in a highly
784 regenerative urochordate: insights into development and host defense. *Dev. Biol.* **374**: 357–73.
785 Elsevier.

786 **Lee, S.-A., O. V. Belyaeva and N. Y. Kedishvili**. **2009**. Biochemical characterization of human
787 epidermal retinol dehydrogenase 2. *Chem. Biol. Interact.* **178**: 182–7.

788 **Lemaire, P., W. C. Smith and H. Nishida**. **2008**. Ascidians and the plasticity of the chordate
789 developmental program. *Curr. Biol.* **18**: R620-31.

790 **Loh, K. M., R. van Amerongen and R. Nusse**. **2016**. Generating Cellular Diversity and Spatial
791 Form: Wnt Signaling and the Evolution of Multicellular Animals. *Dev. Cell* **38**: 643–655.

792 **Lowe, T. M. 1999**. A Computational Screen for Methylation Guide snoRNAs in Yeast. *Science*
793 *(80-.).* **283**: 1168–1171.

794 **Lowe, T. M. and S. R. Eddy**. **1997**. tRNAscan-SE: A program for improved detection of transfer
795 RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.

Luke, G. N., L. F. C. Castro, K. McLay, C. Bird, A. Coulson and P. W. H. Holland. **2003**. Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci.* **100**: 5292–5295.

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**: 18.

MacDonald, B. T., K. Tamai and X. He. **2009**. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Dev. Cell* **17**: 9–26.

Manni, L., G. Zaniolo, F. Cima, P. Burighel and L. Ballarin. **2007**. *Botryllus schlosseri*: a model ascidian for the study of asexual reproduction. *Dev. Dyn.* **236**: 335–52.

Martí-Solans, J., O. V. Belyaeva, N. P. Torres-Aguila, N. Y. Kedishvili, R. Albalat and C. Cañestro. **2016**. Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. *Mol. Biol. Evol.* **33**: 2401–2416.

Martin, A., S. Maher, K. Summerhurst, D. Davidson and P. Murphy. **2012**. Differential deployment of paralogous Wnt genes in the mouse and chick embryo during development. *Evol. Dev.* **14**: 178–195.

Maumus, F. and H. Quesneville. **2014**. Ancestral repeats have shaped epigenome and genome composition for millions of years in Arabidopsis thaliana. *Nat. Commun.* **5**.

Millar, R. H. 1971. The biology of ascidians. *Adv. Mar. Biol.* **9**: 1–100.

Nawrocki, E. P., S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate, *et al.* 2015. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* **43**: D130–D137.

Nawrocki, E. P. and S. R. Eddy. **2013**. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.

Nishida, H. 2005. Specification of embryonic axis and mosaic development in ascidians. *Dev. Dyn.* **233**: 1177–1193.

Nydam, M. L., K. B. Giesbrecht and E. E. Stephenson. **2017**. Origin and Dispersal History of Two Colonial Ascidian Clades in the Botryllus schlosseri Species Complex. *PLoS One* **12**: e0169944.

O'Connell, J., O. Schulz-Trieglaff, E. Carlson, M. M. Hims, N. a. Gormley and a. J. Cox. **2015**. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* **31**: btv057.

O'Neill, R. J., M. J. O'Neill and J. A. Graves. **1998**. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393**: 68–72.

Oka, H. and H. Watanabe. **1957**. Vascular budding, a new type of budding in *Botryllus*. *Biol. Bull.* **112**: 225.

832 **Pascual-Anaya, J., S. D'Aniello, S. Kuratani and J. Garcia-Fernàndez**. **2013**. Evolution of Hox
833 gene clusters in deuterostomes. *BMC Dev. Biol.* **13**: 26.

834 **Pearson, J. C., D. Lemons and W. McGinnis**. **2005**. Modulating Hox gene functions during
835 animal body patterning. *Nat. Rev. Genet.* **6**: 893–904.

836 **Piette, J. and P. Lemaire**. **2015**. Thaliaceans, The Neglected Pelagic Relatives of Ascidians: A
837 Developmental and Evolutionary Enigma. *Q. Rev. Biol.* **90**: 117–145.

838 **Priyam, A., B. J. Woodcroft, V. Rai, A. Munagala, I. Moghul, F. Ter, M. A. Gibbins, H.
839 Moon, G. Leonard, W. Rumpf, *et al.* 2015**. Sequenceserver: a modern graphical user
840 interface for custom BLAST databases. *bioRxiv*.

841 **Prud'homme, B., N. Lartillot, G. Balavoine, A. Adoutte and M. Vervoort**. **2002**. Phylogenetic
842 analysis of the Wnt gene family. Insights from lophotrochozoan members. *Curr. Biol.* **12**:
843 1395.

844 **Rinkevich, B., Z. Shlemberg and L. Fishelson**. **1995**. Whole-body protochordate regeneration
845 from totipotent blood cells. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 7695–9.

846 **Rinkevich, Y., J. Douek, O. Haber, B. Rinkevich and R. Reshef**. **2007**. Urochordate whole body
847 regeneration inaugurates a diverse innate immune signaling profile. *Dev. Biol.* **312**: 131–46.

848 **Rinkevich, Y., G. Paz, B. Rinkevich and R. Reshef**. **2007**. Systemic bud induction and retinoic
849 acid signaling underlie whole body regeneration in the urochordate *Botrylloides leachi*. *PLoS
850 Biol.* **5**: e71.

851 **Rinkevich, Y., B. Rinkevich and R. Reshef**. **2008**. Cell signaling and transcription factor genes
852 expressed during whole body regeneration in a colonial chordate. *BMC Dev. Biol.* **8**: 100.

853 **Rinkevich, Y., A. Voskoboynik, A. Rosner, C. Rabinowitz, G. Paz, M. Oren, J. Douek, G.
854 Alfassi, E. Moiseeva, K. J. Ishizuka, *et al.* 2013**. Repeated, long-term cycling of putative
855 stem cells between niches in a basal chordate. *Dev. Cell* **24**: 76–88. Elsevier Inc.

856 **Ronquist, F. and J. P. Huelsenbeck**. **2003**. MrBayes 3: Bayesian phylogenetic inference under
857 mixed models. *Bioinformatics* **19**: 1572–4.

858 **Ross, A. C. and R. Zolfaghari**. **2011**. Cytochrome P450s in the regulation of cellular retinoic acid
859 metabolism. *Annu. Rev. Nutr.* **31**: 65–87.

860 **Rubinstein, N. D., T. Feldstein, N. Shenkar, F. Botero-Castro, F. Griggio, F. Mastrototaro, F.
861 Delsuc, E. J. P. Douzery, C. Gissi and D. Huchon**. **2013**. Deep sequencing of mixed total
862 DNA without barcodes allows efficient assembly of highly plastic Ascidian mitochondrial
863 genomes. *Genome Biol. Evol.* **5**: 1185–1199.

864 **Sabbadin, A., G. Zaniolo and F. Majone**. **1975**. Determination of polarity and bilateral
865 asymmetry in palleal and vascular buds of the ascidian Botryllus schlosseri. *Dev. Biol.* **46**: 79–
866 87.

867 **Saito, Y., M. Shirae, M. Okuyama and S. Cohen**. **2001**. Phylogeny of Botryllid Ascidians. In:
868 *The Biology of Ascidians*, pp. 315–320. Springer Japan, Tokyo.

869 **Santagati, F., K. Abe, V. Schmidt, T. Schmitt-John, M. Suzuki, K.-I. Yamamura and K. Imai**.
870 **2003**. Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region:
871 implication for evolutionary conserved synteny. *Genetics* **165**: 235–42.

872 **Savigny, J.-C. 1816**. *Mémoires sur les animaux sans vertèbres*. Dufour, G., Paris.

873 **Seo, H.-C. C., M. Kube, R. B. Edvardsen, M. F. Jensen, A. Beck, E. Spriet, G. Gorsky, E. M.**
874 **Thompson, H. Lehrach, R. Reinhardt, *et al.* 2001**. Miniature genome in the marine chordate
875 *Oikopleura dioica*. *Science* **294**: 2506.

876 **Simakov, O., T. Kawashima, F. Marlétaz, J. Jenkins, R. Koyanagi, T. Mitros, K. Hisata, J.**
877 **Bredeson, E. Shoguchi, F. Gyoja, *et al.* 2015**. Hemichordate genomes and deuterostome
878 origins. *Nature* **527**: 459–465.

879 **Simão, F. A., R. M. Waterhouse, P. Ioannidis and E. V Kriventseva**. **2015**. BUSCO : assessing
880 genome assembly and annotation completeness with single-copy orthologs. *Genome Anal.* 9–
881 10.

882 **Simmen, M. W., S. Leitgeb, J. Charlton, S. J. Jones, B. R. Harris, V. H. Clark and A. Bird**.
883 **1999**. Nonmethylated transposable elements and methylated genes in a chordate genome.
884 *Science* **283**: 1164–7.

885 **Simpson, J. T. 2014**. Exploring genome characteristics and sequence quality without a reference.
886 *Bioinformatics* **30**: 1228–1235.

887 **Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones and I. Birol**. **2009**.
888 ABySS: A parallel assembler for short read sequence data. *Genome Res.* **19**: 1117–1123.

889 **Small, K. S., M. Brudno, M. M. Hill and A. Sidow**. **2007**. A haplome alignment and reference
890 sequence of the highly polymorphic Ciona savignyi genome. *Genome Biol.* **8**: R41.

891 **Smit, A. and R. Hubley**. **2015**. RepeatModeler Open-1.0.

892 **Smit, A., R. Hubley and P. Green**. **2015**. RepeatMasker Open-4.0.

893 **Sobreira, T. J. P., F. Marletaz, M. Simoes-Costa, D. Schechtman, A. C. Pereira, F. Brunet, S.**
894 **Sweeney, A. Pani, J. Aronowicz, C. J. Lowe, *et al.* 2011**. Structural shifts of aldehyde
895 dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial
896 patterning in metazoans. *Proc. Natl. Acad. Sci.* **108**: 226–231.

897 **Somorjai, I. M. L., H. Escrivà and J. Garcia-Fernàndez**. **2012**. Amphioxus makes the cut—
898 Again. *Commun. Integr. Biol.* **5**: 499–502.

899 **Spagnuolo, A., F. Ristoratore, A. Di Gregorio, F. Aniello, M. Branno and R. Di Lauro**. **2003**.
900 Unusual number and genomic organization of Hox genes in the tunicate Ciona intestinalis.
901 *Gene* **309**: 71–9.

902 **Stanke, M. and S. Waack**. **2003**. Gene prediction with a hidden Markov model and a new intron
903 submodel. *Bioinformatics* **19**: 215–225.

904 **Stapley, J., A. W. Santure and S. R. Dennis**. **2015**. Transposable elements as agents of rapid
905 adaptation may explain the genetic paradox of invasive species. *Mol. Ecol.* **24**: 2241–2252.

906 **Stoick-Cooper, C. L., G. Weidinger, K. J. Riehle, C. Hubbert, M. B. Major, N. Fausto and R.**
907 **T. Moon**. **2007**. Distinct Wnt signaling pathways have opposing roles in appendage
908 regeneration. *Development* **134**: 479–89.

909 **Stolfi, A., E. K. Lowe, C. Racioppi, F. Ristoratore, C. T. Brown, B. J. Swalla and L.**
910 **Christiaen**. **2014**. Divergent mechanisms regulate conserved cardiopharyngeal development
911 and gene expression in distantly related ascidians. *Elife* **3**: e03728.

912 **Suzuki, M. M., A. R. W. Kerr, D. De Sousa and A. Bird**. **2007**. CpG methylation is targeted to
913 transcription units in an invertebrate genome. *Genome Res.* **17**: 625–631.

914 **Takatori, N., T. Butts, S. Candiani, M. Pestarino, D. E. K. Ferrier, H. Saiga and P. W. H.**
915 **Holland**. **2008**. Comprehensive survey and classification of homeobox genes in the genome of
916 amphioxus, Branchiostoma floridae. *Dev. Genes Evol.* **218**: 579–590.

917 **Tsagkogeorga, G., V. Cahais and N. Galtier**. **2012**. The Population Genomics of a Fast Evolver:
918 High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate
919 Ciona intestinalis. *Genome Biol. Evol.* **4**: 852–861.

920 **Tsagkogeorga, G., X. Turon, N. Galtier, E. J. P. Douzery and F. Delsuc**. **2010**. Accelerated
921 Evolutionary Rate of Housekeeping Genes in Tunicates. *J. Mol. Evol.* **71**: 153–167.

922 **UniProt Consortium**. **2015**. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**: D204–
923 D212.

924 **Voskoboynik, A., N. F. Neff, D. Sahoo, A. M. Newman, D. Pushkarev, W. Koh, B. Passarelli,**
925 **H. C. Fan, G. L. Mantalas, K. J. Palmeri, *et al.* 2013**. The genome sequence of the colonial
926 chordate, *Botryllus schlosseri*. *Elife* **2**: 1–24.

927 **Voskoboynik, A., N. Simon-Blecher, Y. Soen, B. Rinkevich, A. W. De Tomaso, K. J. Ishizuka**
928 **and I. L. Weissman**. **2007**. Striving for normality: whole body regeneration through a series
929 of abnormal generations. *FASEB J.* **21**: 1335–44.

930 **Wada, S., M. Tokuoka, E. Shoguchi, K. Kobayashi, A. Di Gregorio, A. Spagnuolo, M. Branno,**
931 **Y. Kohara, D. Rokhsar, M. Levine, *et al.* 2003**. A genomewide survey of developmentally
932 relevant genes in Ciona intestinalis. *Dev. Genes Evol.* **213**: 222–234.

933 **Wang, X.-P., M. Suomalainen, S. Felszeghy, L. C. Zelarayan, M. T. Alonso, M. V Plikus, R. L.**
934 **Maas, C.-M. Chuong, T. Schimmang and I. Thesleff**. **2007**. An integrated gene regulatory
935 network controls stem cell proliferation in teeth. *PLoS Biol.* **5**: e159.

936 **Wences, A. H. and M. C. Schatz**. **2015**. Metassembler: merging and optimizing de novo genome
937 assemblies. *Genome Biol.* **16**: 207. Genome Biology.

Zerbino, D. R. and E. Birney. **2008**. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**: 821–9.

Zhan, A., E. Briski, D. G. Bock, S. Ghabooli and H. J. MacIsaac. **2015**. Ascidians as models for studying invasion success. *Mar. Biol.* **162**.

Zhang, G., X. Fang, X. Guo, L. Li, R. Luo, F. Xu, P. Yang, L. Zhang, X. Wang, H. Qi, *et al.* **2012**. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49–54.

Zimin, A. V., G. Marcais, D. Puiu, M. Roberts, S. L. Salzberg and J. A. Yorke. **2013**. The MaSuRCA genome assembler. *Bioinformatics* **29**: 2669–2677.

Zondag, L. E., K. Rutherford, N. J. Gemmell and M. J. Wilson. **2016**. Uncovering the pathways underlying whole body regeneration in a chordate model, *Botrylloides leachi* using de novo transcriptome analysis. *BMC Genomics* **17**: 114. BMC Genomics.