UNIVERSITAT DE BARCELONA

Oikopleura dioica com a model animal per investigar l'impacte de les pèrdues gèniques en l'Evo-Devo: les vies de senyalització de l'àcid retinoic i Wnt com a cas d'estudi

Josep Martí Solans

Facultat de Biologia
Departament de Genètica, Microbiologia i Estadística
Programa de doctorat de Genètica

Oikopleura dioica com a model animal per investigar l'impacte de les pèrdues gèniques en l'Evo-Devo: les vies de senyalització de l'àcid retinoic i Wnt com a cas d'estudi

Memòria presentada per
Josep Martí-Solans

Per optar al grau de
Doctor per la Universitat de Barcelona

Tesis Doctoral realitzada sota la direcció dels Dr. Ricard Albalat i Dr. Cristian Cañestro del Departament de Genètica, Microbiologia i Estadística de la Facultat de Biologia de la Universitat de Barcelona

Dr. Ricard Albalat
(co-director i tutor)

Dr. Cristian Cañestro
(co-director)

Josep Martí-Solans
(autor)

7 de Juny de 2018

Agraïments

No puc començar una llista d'agraïments sense donar les gràcies als meus directors de Tesi. Gràcies Cristian. Gràcies Ricard. Us vull donar les gràcies molt sincerament, primer, per haver-me donat la oportunitat de formar-me al vostre costat, i després, per haver estat alguna cosa més que uns investigadors que buscaven un estudiant. Us heu preocupat de que generes un esperit crític i m'heu cuidat dins i fora del laboratori, us estic molt agraït.

Mirant enrere, recordo tota la gent amb la que he compartit laboratori. Gràcies Hèctor pel teu caos i la teva espontaneïtat. Encara que a vegades eren desconcertants, en guardo un bon record. Gracias Marcos por transmitirme tu pasión por la ciencia, recuerdo cuando volvías de las clases de genética del desarrollo alucinado con lo que te habían contado, agradezco mucho que lo descargases en mí, me ayudó mucho. Alfonso gracias por coger el testigo. I la resta dels que en algun moment han format part del *Oikoteam* -David, Arnau, Jordi, Núria, Míriam, l'Anna(e), Sebas, Sara, Natalia, Paula, Enya, Marc, Vittoria-, gràcies pels bons moments que heu decidit compartir amb mi.

M'agradaria donar les gràcies també als membres de l'Stazione Zoologica de Nàpols. Gracies Salvatore, per portar-nos a buscar Oikos amb el teu veler, gràcies a l'Ana i a la Giovanna, per la acollida en els vostres laboratoris y gràcies Alberto per facilitar les coses.

També m'agradaria donar les gràcies als membres del laboratori d'en Hiroki Nishida. Mai havia vist una gent tan complaent i que es preocupes tan pel benestar dels altres com vosaltres. Em vau posar les coses molt fàcils. In particular, I would like to thank Hikawa, Ayumi, Shi Yu and Tomura-kun for the fantastic moments that we shared. Without you, guys, I would not have endured neither two months! Arigatō!!! Vull agrair, també, en Takeshi Onuma la supervisió dels meus experiments mentre jo era allà i per ensenyar-me com s'han d'injectar les femelles d'Oiko.

Abans de la part més sentimental, et vull donar les gràcies a tu, lector. Gràcies per obrir la meva Tesi. Espero que et sigui útil.

Per anar acabant, vull donar les gracies als meus pares, germà, iaia i tieta. És meravellós el suport incommensurable que et poden arribar a transmetre les

persones que t'estimen, ja sigui en forma de sofregit per les mandonguilles o en forma de roba planxa, que t'havies deixat a la secadora perquè has sortit corren cap al laboratori. Dels meus pares he après que són les petites coses de la vida les que et fan feliç, i han estat les petites victòries les que m'han permés seguir amb la Tesi.

A tu Jordi et vull donar les gràcies, perquè, ha estat ara, amb el doctorat que m'he adonat del que costa ser el primer. Moltíssimes gràcies per haver lluitat per obrir-me tantes portes.

A ti Alba, una de las personas con la que más he hablado de ciencia, gracias por estar allí, gracias por no dejarme caer.

Index

Pròleg

Quantes potes ha de tenir una taula perquè sigui el més estable i ferma possible, quantes més millor? Quants gens ha de tenir un animal, una planta o un microbi per sobreviure, quants més millor? De la mateixa forma que l'experiència demostra que la taula més estable només té tres potes, i no quatre com la majoria de taules, estudis recents suggereixen que malgrat haver associat intuïtivament l'Evolució dels essers vius a un increment en el nombre de gens, perdre gens pot ser també una important font de variabilitat gènica capaç de proporcionar capacitats adaptatives als organismes. En aquest projecte de tesis doctoral hem volgut reptar la intuïció i estudiar que passa quan es perden els gens, que els hi passa als que es queden, i com els esdeveniments de pèrdua o retenció de gens poden haver impactat en l'evolució dels mecanismes del desenvolupament embrionari (EvoDevo) dels organismes.

Introducció

L'EvoDevo

L'Evolució ha donat lloc a una extraordinària biodiversitat d'organismes amb una gran varietat de formes i estructures. Les formes i les estructures dels éssers vius resulten dels processos ontogenètics i, per tant, per entendre els mecanismes evolutius generadors de biodiversitat cal comprendre l'evolució del desenvolupament embrionari. D'aquesta forma va sorgir a finals del segle XX l'Evo-Devo, una nova disciplina de la Biologia que per tal d'entendre l'evolució dels éssers vius integrava coneixements de la biologia del desenvolupament, la genètica evolutiva, l'anatomia comparada, la genòmica i genètica molecular, i la filogènia i la sistemàtica. Els primers estudis d'Evo-Devo als anys vuitanta van posar de manifest que el desenvolupament embrionari de moltes estructures, òrgans i teixits d'espècies evolutivament allunyades i morfològicament diferents depenia d'un conjunt de gens similars. És a dir, que malgrat els organismes podien mostrar una gran "diversitat fenotípica", compartien una caixa d'eines gèniques del desenvolupament ("*genetic toolkit*") molt similar, i per tant compartien una "unitat genotípica" destacable (Jacob, 1977; King and Wilson, 1975). Aquest fet, conegut com la paradoxa de l'Evo-Devo, ha estat il·lustrat per la conservació dels gens Hox (Homeobox) i del gen *Pax6* (paired box gene 6), encarregats d'establir, respectivament, les identitats dels territoris al llarg de l'eix antero-posterior en tots els bilaterals i de controlar el desenvolupament de l'ull en mamífers, insectes i cefalòpodes (Halder et al., 1995; McGinnis et al., 1984).

Les caixes d'eines genètiques, malgrat estar globalment conservades, no són immutables i han sofert canvis al llarg de l'evolució. Aquests canvis poden afectar la funció gènica modificant l'expressió d'un determinat gen, la proteïna que codifica o el nombre de còpies present en el genoma. Per exemple, canvis en les regions reguladores d'un gen o en els gens reguladors que controlen la transcripció d'aquest, poden alterar la seva expressió i causar modificacions de color en els organismes o fins i tot la formació d'apèndixs (Gompel et al., 2005; Shapiro et al., 2006) (fig. 1A). De forma similar, mutacions en la regió

codificant d'un gen poden alterar la funció de la proteïna que codifica i generar canvis fenotípics importants. Així per exemple, proteïnes activadores de la transcripció poden passar a ser repressores per canvis en la seva seqüència

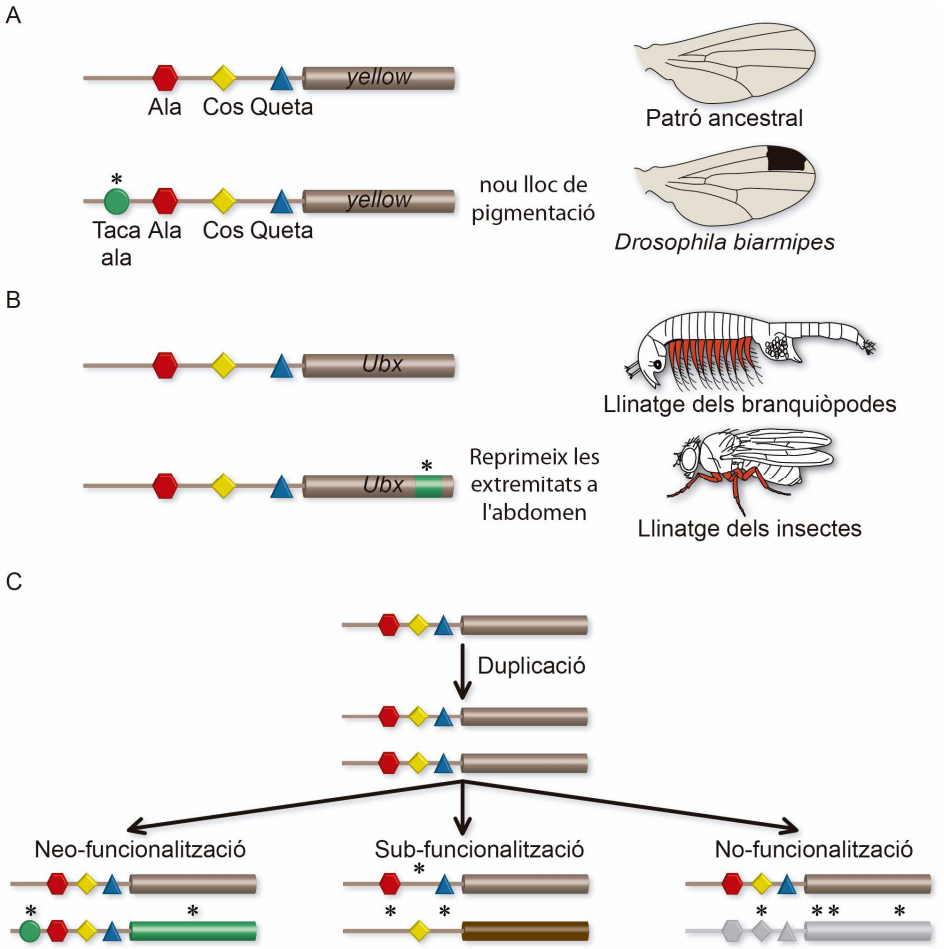


Fig. 1. **Canvis en les caixes d'eines genètiques augmenten la diversitat fenotípica.** (A) Una ampliació en el nombre d'elements *cis*-reguladors (cercle, hexàgon, rombe i triangle) per qualsevol mecanisme (transposició, reordenació, duplicació, mutació puntual) pot ampliar el nombre de teixits en què el gen està actiu, mentre conserva la funció original -p.e. l'aparició d'un nou lloc de pigmentació a causa de l'aparició d'un nou element *cis*-regulador (cercle) en el gen *yellow* en la família Drosophilidae-. Basat en Gompel et al., 2005. (B) Mutacions en la regió codificant d'un gen poden canviar la funció d'aquest –p.e. mutacions en els llocs de fosforilació serina/treonina de la proteïna Ultrabithorax (Ubx) durant l'evolució del llinatge dels insectes van causar que es deixes d'inhibir la funció repressora l'Ubx, contribuint així al pla corporal dels hexàpodes. Basat en Ronshaugen et al., 2002. (C) La duplicació de gens seguida de mutacions (asteriscs) en la regió reguladora o codificant dels gens paràlegs produeix gens que es poden expressar de diferent forma o proteïnes amb funcions diferents, mentre es manté la funció original (neofuncionalització). D'altre banda, la duplicació de gens pot anar seguida d'una pèrdua complementaria de subfuncions en els gens paràlegs de manera que es conservin permanentment la parella de gens duplicats (subfuncionalització). Encara que el més comú després d'una duplicació és que una de les còpies acumulin mutacions que produeixin una proteïna no funcional (no-funcionalització).

aminoacídica i impedir la formació de certes estructures (Galant and Carroll, 2002; Ronshaugen et al., 2002) (fig. 1B). Finalment, la duplicació d'un gen i el conseqüent increment del nombre de còpies pot afavorir l'aparició d'innovacions funcionals durant l'evolució. De fet, des de que el 1970 Susumo Ohno va publicar el seu llibre "Evolution by gen duplication", les duplicacions gèniques han estat probablement els processos evolutius generadors d'innovacions funcionals més estudiats. Quan un gen es duplica, i per tant existeix en més d'una còpia, una de les còpies pot alliberar-se de les obligacions funcionals imposades per la selecció natural. Llavors, mentre una de les còpies manté la funció ancestral, el duplicat pot acumular mutacions que el portin a adquirir noves funcions en un procés de "neo-funcionalització" (fig. 1C). De forma alternativa, és possible que la funció ancestral es "sub-funcionalitzi" entre les dues còpies seguint un model de Duplicació-Degeneració-Complementació (DDC) (Force, 1999). En aquest cas, les dues còpies es mantenen perquè mutacions complementaries entre els duplicats fa que els dos gens siguin necessaris per proporcionar les funcions ancestrals (fig. 1C).

Les pèrdues gèniques durant l'evolució

Malgrat un gen duplicat pot neo- o sub-funcionalitzar, el més probable segons el model clàssic de Susumo Ohno és que aquest degeneri per acumulació de mutacions, perdi la seva funció i esdevingui un pseudogen en un procés conegut com "no-funcionalització" o "pseudogenització (Ohno, 1970) (fig. 1C). En el cas de gens duplicats, aquest procés havia estat tradicionalment associat a la pèrdua de gens redundants sense conseqüències funcionals i, per tant, sense transcendència evolutiva. A més, fins fa poc, provar que un gen s'havia perdut era formalment difícil ja que es fonamentava en la demostració d'un resultat negatiu. L'explosió de "projectes genoma" en un ampli ventall d'espècies i, per tant, la possibilitat d'identificar tots els gens d'un organisme i de fer estudis de genòmica comparada al llarg de l'escala evolutiva ha canviat aquesta situació. Per exemple, la seqüenciació del genoma del cnidari *Nematostella vectensis* (Putnam et al., 2007) va revelar que la majoria dels gens dels vertebrats ja estaven presents en el ancestre comú dels bilaterals i, per tant, que durant l'evolució dels protòstoms ecdisozous (p.e. *Drosophila melanogaster* i *Caenorhabditis elegans*) s'haurien perdut fins el 17% de les famílies gèniques. La presència a *N. vectensis* de molts gens que es creien innovacions dels vertebrats desterrava, a més, la percepció que els organismes més complexes tenien un nombre i una varietat de gens més elevada (Szathmary, 2001), i potenciava la idea que el genoma de l'ancestre dels metazous era genèticament complex. Des d'aquesta complexitat, els gens

s'haurien perdut de forma esbiaixada en els diversos llinatges, de manera que en alguns grups d'ecdisozous (p.e. insectes i nematodes) haurien acumulat un elevat nombre de pèrdues gèniques, mentre que en altres grups com els lofotrocozous (p.e. mol·luscs i anèl·lids) o els deuteròstoms (p.e. vertebrats) no s'haurien donat tantes pèrdues. Per altre banda, més enllà dels trets generals dels grans grups d'animals, cada llinatge tindria la seva pròpia tendència a perdre o retenir gens, tal com es visualitza al comparar les pèrdues per una mateixa família de gens entre especies diferents distribuïdes al llarg de l'escala evolutiva (p.e. Wnt). En resum, l'elevada freqüència i el biaix detectat en la pèrdua de gens entre diferents organismes suggereix que la pèrdua gènica pot haver estat una força evolutiva rellevant capaç de generar diversitat genètica que ha afavorit la divergència de les espècies durant l'evolució dels essers vius.

Prescindir de gens

Constatada la notable prevalença de la pèrdua gènica durant l'evolució dels organismes, s'ha d'entendre sota quines circumstàncies els gens es poden perdre. La intuïció ens diu que la possibilitat de que un gen es perdi dependrà de com d'essencial o prescindible sigui, és a dir, de si la seva absència és crítica o no per a la supervivència de l'organisme. Com de prescindible és un gen és, per tant, inversament proporcional a la importància del gen, i aquesta importància s'ha pogut mesurar al laboratori a través de valorar la *fitness* dels individus genoanul·lats (*knockouts)* per el gen en qüestió. Sorprenentment, experiments a gran escala de genoanul·lació sistemàtica de gens individuals en bacteris han demostrat que vora el 90% dels gens bacterians son prescindibles ja que la seva anul·lació no produeix una reducció en la *fitness*, al menys en les condicions de laboratori utilitzades (Baba et al., 2006; de Berardinis et al., 2008; Korona, 2011). El mateix passa amb el 80% dels gens de llevat (Kamath et al., 2003; Kim et al., 2010) o amb el 85% i el 65% dels gens de *D. melanogaster* i *C.elegans*, respectivament (Dietzl et al., 2007; Kamath et al., 2003; Sönnichsen et al., 2005). Aquests valors sorprenentment alts de gens aparentment prescindibles en els genomes de tots els organismes és coneix com la "paradoxa dels *knockouts*" (Papp et al., 2011) i s'ha intentat explicar en base a dos factors que afecten a l'actitud prescindible d'un gen: la robustesa mutacional i la variabilitat ambiental.

Robustesa mutacional

La robustesa mutacional és la capacitat que tenen els sistemes biològics de mantenir inalterat un determinat fenotip malgrat es produeixin mutacions. En un sistema mutacionalment robust, un gen pot ser prescindible perquè la seva mutació no comporti un canvi fenotípic seleccionable gràcies a l'existència de

mecanismes compensatoris. Aquests mecanismes poden ser deguts a la presència de gens funcionalment redundants que provinguin d'un procés de duplicació (i.e. gens paràlegs) o de convergència evolutiva (gens anàlegs), o a l'existència de vies alternatives per fer la mateixa funció resultat de l'organització dels gens en xarxa. Per tant, la paradoxa dels *knockouts* pot explicar-se perquè molts gens són 'individualment' prescindibles ja que formen part d'un sistema robust. Així doncs, s'ha pogut demostrar en experiments de genoanul·lació múltiple en llevat que el nombre de gens prescindibles cau dramàticament, des del 80% a només el 25%, quan múltiples gens són mutats a la vegada (Deutscher et al., 2006).

Variabilitat ambiental.

A més de la robustesa mutacional, un altre factor que explica la paradoxa dels *knockouts* és variabilitat ambiental. Molts gens poden ser prescindibles en les condiciones ambientals del laboratori, però deixar de ser-ho si les condicions ambientals canvien. De fet, el catàleg actual de gens d'un organisme és fruit de les pressions selectives d'una àmplia gamma d'ambients que han actuat al llarg de milions d'anys d'evolució. De nou, l'estudi de milers de mutants de llevat analitzats en més de mil condicions ambientals diferents ha demostrat que mutacions en el 97% dels gens tenen efectes en el creixement del llevat en alguna de les condicions experimentals testades (Hillenmeyer et al., 2008; Musso et al., 2008).

Per tant, la possibilitat que un gen es perdi deriva del seu grau de prescidibilitat, el qual depèn de la robustesa mutacional o de les condicions ambientals a les que s'enfronti l'organisme. En qualsevol cas, un gen és prescindible quan la seva pèrdua no comporta una disminució de la *fitness* dels individus que el perden, és dir, quan la seva no-funcionalització és evolutivament neutra (o quasi neutra) o adaptativa (fig. 2).

Evolució per pèrdua de gens

És per tant fonamental entendre com les mutacions que causen la no-funcionalització dels gens poden ser adaptatives o neutres, ja que la seva naturalesa determina la probabilitat que tenen de fixar-se dins de la població i, en darrer terme, l'impacte que les pèrdues gèniques poden haver tingut en l'evolució de les espècies. Segons la hipòtesis del "menys és més" (*less is more hypothesis*), la pèrdua gènica pot ser adaptativa i en aquest cas la seva probabilitat de fixar-se depèn del seu coeficient de selecció. Per contra, en situacions d'evolució regressiva és probable que la pèrdua sigui neutra (o quasi neutra) i a les hores la possibilitat que es fixi depèn de la mida poblacional de

l'espècie la qual determina la força de la deriva genètica en el procés (fig.2).

La hipòtesis del menys és més.

La hipòtesi del "menys es més" proposa que la no-funcionalització dels gens és una resposta adaptativa freqüent especialment per canvis selectius causats per variacions dràstiques de les condicions ambientals. S'han descrit nombrosos exemples en microorganismes que sustenten aquesta hipòtesis, però també casos en plantes i animals. De fet, un dels casos més estudiats s'ha trobat en humans. La deleció de 32 parells de bases en el gen que codifica pel receptor de quimiocines CCR5 causa la no-funcionalització del receptor conferint als individus homozigots resistència a la infecció del virus de la immunodeficiència humana (VIH) (Dean et al., 1996; Samson et al., 1996). L'alta freqüència i la distribució de l'al·lel no-funcional del gen CCR5 indiquen que s'ha estat seleccionat intensament a Europa, encara que, aquesta selecció pugui no ser deguda al VIH sinó a altres virus ja que el SIDA es pensa que és una malaltia humana moderna (Galvani and Novembre, 2005; Saxena, 2009).

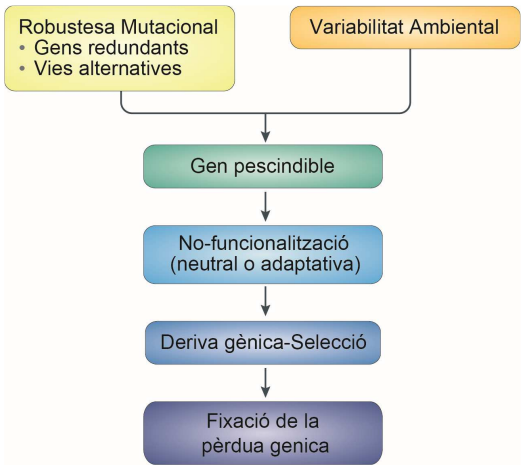


Fig. 2. **Marc conceptual per les pèrdues gèniques.** La pèrdua d'un gen depèn de com de prescindible és el gen, que al seu torn depèn de com la *fitness* es veu afectada per la no-funcionalització d'aquest. En un sistema mutacionalment robust les mutacions tindran menys impacte en la *fitness* convertint, així, els gens en més prescindibles. Per altre banda, variacions en l'ambient poden fer que gens essencials en unes condicions esdevinguin prescindibles en altres. La no-funcionalització d'un gen prescindible pot ser neutre (o gairebé neutre) quan el gen no és necessari i llavors la seva fixació en la població dependrà de la deriva gènica o pot ser adaptativa si la pèrdua de la funció és avantatjós en la nova condició i llavors la seva fixació serà seleccionada. Extret de Albalat and Cañestro, 2016.

Evolució regressiva.

El concepte de "evolució regressiva" es refereix al fet que els organismes perden característiques supèrflues o en desús durant la seva evolució. Molts exemples de pèrdua de gens selectivament neutres s'han descrit associats a processos d'evolució regressiva. Un dels casos més documentats és la pèrdua de gens en espècies que s'han adaptat a viure en la foscor de les coves, on la capacitat de veure i la pigmentació del cos han esdevingut supèrflues. Les poblacions de l'espècie del peix cavernícola *Astyanax mexicanus* que han colonitzat ambients

sense llum, per exemple, han acumulat mutacions no-funcionalitzadores diferents en el gen *Oca2* (*ocular and cutaneous albinism 2*), que és el gen més mutat en casos d'albinisme en humans (Protas et al., 2006).

El debat de si la majoria de pèrdues gèniques fixades en les espècies són adaptatives o neutres no s'ha resolt i s'emmarca en la clàssica discussió entre seleccionistes i neutralistes sobre la naturalesa neutra o adaptativa de la variació genètica en les poblacions. Tanmateix, és possible argumentar que per a organismes unicel·lulars de vida lliure amb mides poblacionals molts grans, la fixació de pèrdua de gens podria ser majoritàriament adaptativa i dirigida per la selecció natural, mentre que per espècies unicel·lulars paràsites o simbionts que pateixen freqüents colls d'ampolla, la majoria de les fixacions serien neutres i conseqüència de la deriva genètica. Pels organismes pluricel·lulars amb mides poblacionals molt més petites que les dels microorganismes, és possible que la deriva genètica hagi estat la força motriu principal per a la fixació de la pèrdua de gens en aquestes espècies.

Els patrons de pèrdues gèniques

Estudis comparatius a gran escala han mostrat que les pèrdues gèniques no han afectat a tots els gens per igual sinó que presenten biaixos a causa de la seva localització (biaix per la posició genòmica) o de la proteïna que codifiquen (biaix per la funció gènica).

Biaix per la posició genòmica.

S'han constatat biaixos en les pèrdues de gens que acompanyen al procés de diploidització, és a dir, durant el procés de retorn al estat diploide després d'una duplicació sencera del genoma (*whole-genome duplication*; WGD) les pèrdues de gens poden no donar-se de forma aleatòria sinó que és pot donar un biax en funció d'en quin dels cromosomes duplicats es troben els paràlegs. En alguns casos, s'han observat pèrdues asimètriques que generen regions amb un alt grau de sintènia conservada només en un dels cromosomes (Makino and McLysaght, 2012). En altres, s'han detectat pèrdues recíproques dels duplicats en espècies que van divergir després de una WGD, les quals podrien haver contribuït a la isolació reproductiva de les espècies ja que reduirien la viabilitat dels híbrids. Finalment, un dels exemples més extrems de biaix de pèrdues associat a la posició genòmica el trobem en l'evolució dels cromosomes sexuals (p.e. X i Y en humans) durant la qual la majoria de gens que una vegada compartien la parella d'autosomes dels que deriven han estat perduts majoritàriament en un d'ells (p.e. el cromosoma Y).

Biaix per la funció gènica.

Al classificar els gens que s'han perdut segons les seves categories funcionals (*Gene Ontology*; GO) s'han observat biaixos en les pèrdues de gens que pertanyen a determinades categories entre diferents organismes. Per exemple, gens de categories GO tals com "modificació proteica", "metabolisme proteic" "catabolisme" o "activitat peptidasa" són més propensos a perdre's en peixos que en vertebrats terrestres (Blomme et al., 2006). Aquests biaixos estarien possiblement relacionats amb els diferents estils de vida de les diferents espècies (p.e. espècies aquàtiques *versus* terrestres).

Un cas particularment interessant de biaix funcional és el de la co-eliminació. En aquest cas, el biaix consisteix en l'eliminació de conjunts de gens relacionats, bé perquè codifiquen proteïnes que s'organitzen formant una via funcional o perquè les proteïnes s'associen en complexos. Si per alguna raó, una d'aquestes

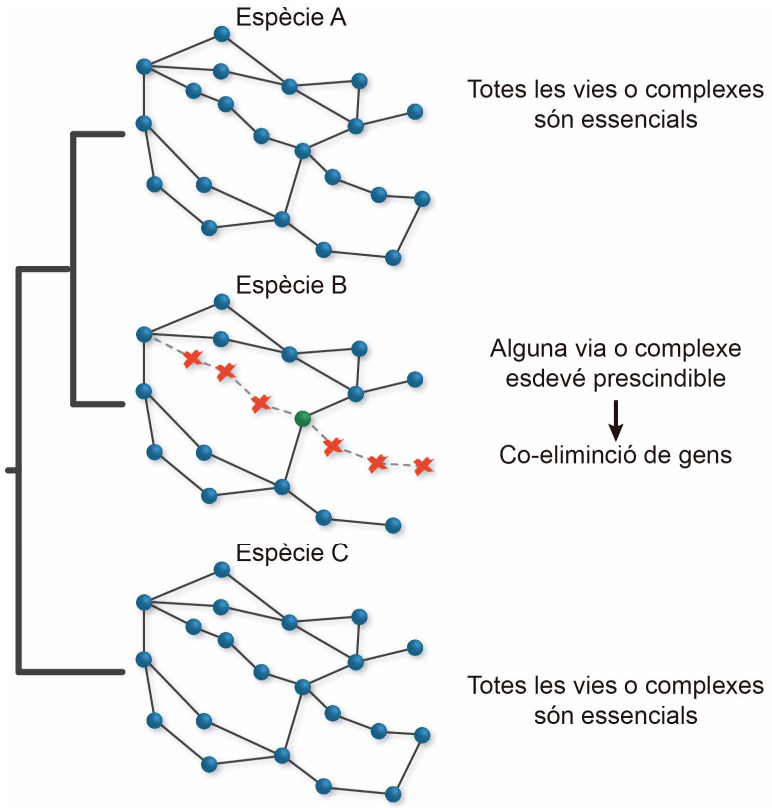


Fig 3. **Co-eliminació de gens.** La comparació dels membres d'una xarxa gènica (esferes blaves) entre dues especies amb l'ajut d'una xarxa gènica ancestral permet la identificació de patrons de gens que s'eliminen junts (co-eliminació, creus vermelles) revelant mòduls funcionals dins de les xarxes de gèniques (línia discontinua). Per altre banda, la identificació de gens que han sobreviscut al desballestament d'una via permet reconèixer els gens "hub" (esfera verda) que no es poden perdre perquè actuen en múltiples vies.

vies o complexos esdevé innecessari, tots els gens que en formen part podran perdre's, excepte aquells que intervinguin en altres funcions. S'ha observat, per exemple, la co-eliminació de gens del complex eIF3/signalosome a *S. cerevisiae*, de gens gàstrics a l'ornitoric i molts peixos teleostis, dels gens que codifiquen proteïnes estructurals crucials per a la formació de l'esmalt i la dentina en les aus, o del conjunt de gens involucrats en el sistema de reparació del DNA per unió d'extrems no homòlegs (*non-homologous DNA end joining*, NHEJ) en l'urocordat *Oikopleura dioica* (Aravind et al., 2000; Denoeud et al., 2010; Ordoñez et al., 2008; Sire et al., 2008). De fet, la identificació de patrons de co-eliminació de gens pot ser una estratègia útil per a reconèixer els components que formen els complexos proteics o per identificar els elements de les vies funcionals dins de les intricades xarxes de gèniques dels éssers vius. (fig. 3).

Oikopleura dioica *com a model animal per a l'anàlisi de les pèrdues gèniques*

Per estudiar l'impacte que les pèrdues gèniques han tingut en l'EvoDevo és important disposar d'un model animal de referència que compleixi una sèrie de requisits. El model a escollir ha de tenir unes característiques biològiques que facilitin els estudis de desenvolupament embrionari i disposar, a ser possible, de tècniques de biologia molecular que permetin les investigacions funcionals en el laboratori. A més, el model ha d'ocupar una posició filogenètica estratègica que permeti identificar esdeveniments de pèrdues gèniques i interpretar les possibles conseqüències evolutives d'aquestes. Finalment, el model ha de ser procliu a perdre gens, amb un genoma plàstic i d'una complexitat genètica reduïda. El nostre grup ha triat el cordat *Oikopleura dioica* Fol, 1872 com a model animal per a estudiar les pèrdues gèniques, a més de poder aportar dades rellevants per entendre l'evolució del nostre propi fílum, els cordats.

En aquest apartat s'exposaran les qualitats que fan d'*O. dioica* un model animal interessant per a estudis en el camp de L'EVO-DEVO. Concretament, es revisaran tres aspectes d'aquesta espècie: (i) les seves característiques biològiques que faciliten el seu cultiu i experimentació en el laboratori, (ii) la seva posició filogenètica com a membre del grup dels animals més proper als vertebrats que permet anàlisis de genòmica i anatomia comparades i, especialment, (iii) l'extraordinari grau de plasticitat del seu genoma que s'ha traduït en un elevat número de pèrdues gèniques durant la seva evolució.

*Les característiques biològiques d'*O. dioica

Distribució geogràfica i trets ecològics

A l'hora d'establir un model animals, les espècies cosmopolites són les que estan més ven posicionades per a convertir-se en animals model. Això ho sabia Thomas H. Morgan quan va deixar uns plàtans madurar en una finestra del seu laboratori per aconseguir l'animal model que utilitzaria per investigar i entendre les bases de l'herència en els animals, *D. melanogaster* (Kohler, 1994). L'elecció d'*O. dioica* com a animal model respecte aquesta circumstància.

O. dioica és un animal del fílum dels cordats, concretament de la classe de les apendiculàries, de vida lliure que forma part del plàncton i s'alimenta de microalgues. *O. dioica* és una espècie semi-cosmopolita i per tant es pot recol·lectar tant en els oceans Atlàntic, Pacífic i Índic, com en els mars

Mediterrani i Roig, però no en els oceans amb latituds més extremes, Àrtic i Antàrtic (Fenaux et al., 1998). A més, *O. dioica* és una espècie que viu en la zona nerítica dels oceans, en profunditats que van des de la superfície fins els 500 metres, per tant no es necessari endinsar-se mar en dins per trobar aquest organisme. S'ha detectat que en certes localitzacions la presència de les diferents espècies apendiculàries es dóna de forma estacional (Essenberg, 1922; Fenaux et al., 1998; Uye and Ichino, 1995). Per exemple, a l'oest del Mediterrani *O. dioica* mostra els seus màxims d'abundància a la primavera, de fet entre les Oikopleuridae és l'espècie més abundant arribant a ser el 22% del total de les apendiculàries en aquesta època de l'any (Fenaux et al., 1998), a la tardor *O. dioica* torna a mostrar un pic en abundància, encara que aquesta vegada és menor (Raduan et al., 1985). Així doncs, la abundància i localització d'*O. dioica* fan que aquesta sigui de relativament fàcil recol·lecció en tots els indrets marins del món.

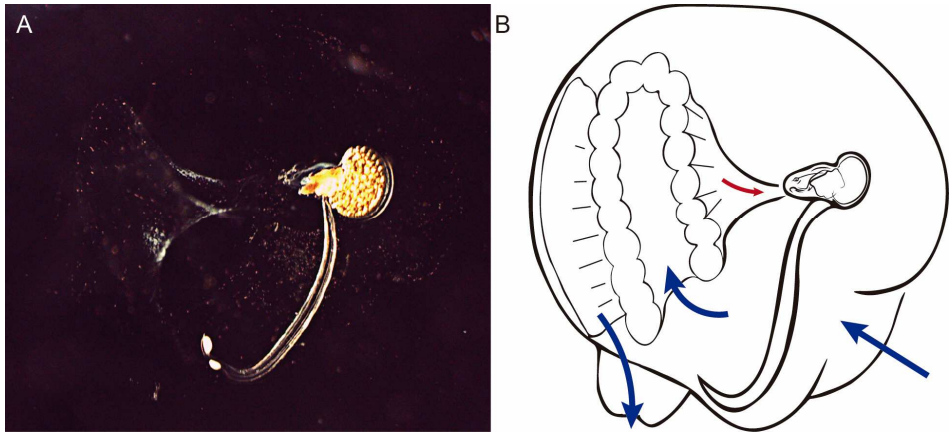


Fig. 4. **Cicle de vida d'*Oikopleura dioica*.** (A)Femella d'*O. dioica* dins d'una casa completa inflada que utilitza com a dispositiu filtrador i acumulador de l'aliment. El filtre concentrador d'aliment és visible a la part esquerra de la imatge, metre que l'animal és situa a la dreta amb la boca orientada cap a l'esquerra, la gonada cap a la dreta i la punta de la cua cap a la part baixa de la imatge. (B) Representació esquemàtica de la casa *O. dioica* modificada després de Flood and Deibel, 1998 i Thompson et al., 2001. El flux d'aigua a través de la casa s'indica mitjançant fletxes blaves i la concentració de partícules alimentàries cap a la boca mitjançant una fletxa vermella.

O. dioica, com la resta d'apendiculàries, viu dins d'una "casa" (d'aquí l'arrel del seu nom *oiko*, que en grec significa casa). Aquesta casa, composta principalment de mucopolisacàrids, és secretada per l'apendiculària que la utilitza com dispositiu filtrador capaç d'atrapar partícules que abasten diversos ordres de magnitud (0,1-50 μm) que inclou el picoplàncton (Fenaux, 1986; Körner, 1952; Thompson et al., 2001) (fig. 4). Una vegada la casa està bruta, les apendiculàries l'abandonen i en secreten una de nova, un procés que es pot

donar de 4 a 10 vegades al dia depenent de la temperatura de l'aigua en que visquin els animals (Flood and Deibel, 1998). Aquesta estratègia que a primera vista podria semblar energèticament massa exigent (Sato et al., 2001), en acabat podria ser una adaptació a ambients amb poc aliment (Acuña, 2001). Per respondre a les condicions ambientals desfavorables *O. dioica* ha evolucionat seguit l'estratègia dels manipuladors de la posta (*clutch manipulators*), és a dir, en una situació d'increment d'aliment *O. dioica* augmenta la quantitat d'oòcits que produeix en comptes d'escurçar el seu temps de generació (*time manipulators*)(Troedsson et al., 2002). Aquesta característica es pot utilitzar al laboratori per aconseguir postes més grans quan hi ha la necessitat experimental.

Cicle de vida, desenvolupament embrionari i estadi de larva

Una de les característiques més apreciades entre els biòlegs experimentals i que comparteixen dos dels organismes model més usats, *D. melanogaster* i *C. elegans*, és el fet que l'animal model tinguin temps de generació curt, p. e. 9 dies en *D. melanogaster* i 3 dies en *C. elegans*. *O. dioica* també té un temps de generació molt curt que és fix a una temperatura determinada (Lombard et al., 2009)(fig. 5). A 19ºC de temperatura, només passen 3,5 hores des de que els oòcits són fecundats externament per l'esperma fins que es completa l'embriogènesi amb la desclosa d'una nova larva. El desenvolupament larvari dura 6 hores i acaba quan l'animal canvia l'orientació de la cua 120º (*tailshift*). Uns 10 minuts després d'aquest canvi, l'animal juvenil infla la primera casa i comença a menjar. Durant els següents cinc dies, els individus d'*O. dioica* creixen fins arribar als 1,5-4 mil·límetres d'un animal adult, és llavors quan les gònades completen la seva maduració i alliberen els gàmetes que donaran lloc a la següent generació per mitjà de fecundació externa tancant el cilce de vida. *O. dioica* és una espècie semèlpara, i per tant, els individus moren després d'alliberar els gàmetes. En el laboratori s'ha pogut controlar la reproducció de *O. dioica* mitjançant la fertilització artificial dels oòcits (Fenaux, 1976; Holland et al., 1988; Nishino and Morisawa, 1998). Gràcies a aquesta pràctica s'han pogut descriure els processos que porten a l'oòcit a esdevenir un juvenil. Per començar, es va poder determinar que quan la femella allibera els oòcits aquets estan aturats a la primera metafase meiòtica, així doncs amb la fecundació dels oòcits apareixen els dos corpuscles polars que indiquen la finalització de la meiosis (Fenaux, 1998a; Fujii et al., 2008; Ganot et al., 2007) (fig. 6). Llavors, a partir de la fecundació s'inicia el desenvolupament embrionari que és invariant i determinat. Així doncs, s'ha pogut establir, també, el mapa de destí o llinatge cel·lular des de l'estadi d'una cèl·lula fins l'estadi de *tailbud* de forma equivalent al que s'ha fet amb altres models animals com el nematode *C.*

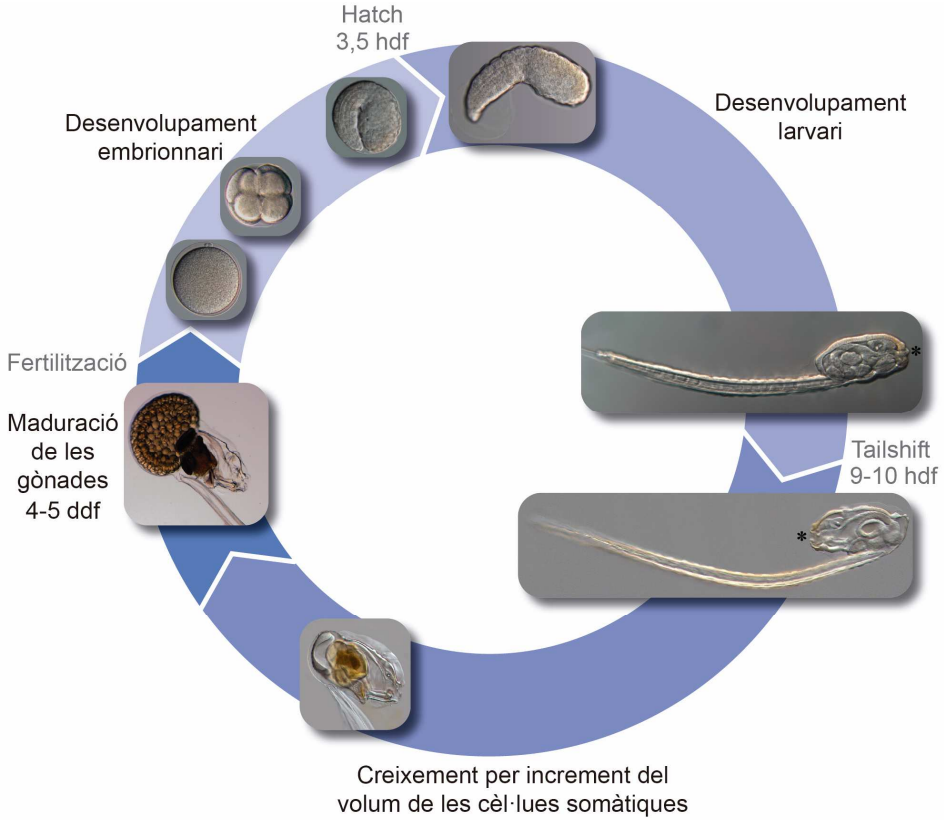


Fig. 5. **El cicle de vida *O. dioica* a 19˚C**. Es mostra la duració del desenvolupament embrionari que s'inicia amb la fertilització dels oòcits i s'acaba a les 3,5 hores després de la fertilització (hdf) amb la desclosa (*hatch*) de la larva; el desenvolupament larvari que finalitza 6 hores després del *hatch* amb el canvi d'orientació de la cua del animal (*tailshift*). Els asteriscs mostren que la posició de la boca respecta la punta de la cua a canviat després del *tailshift*. També es mostra la fase de creixement adult on apartir del quart dia després de la fertilització (ddp) comença la maduració de les gònades dels animals.

elegans o l'ascidi *Ciona intestinalis* (Fujii et al., 2008; Nishida, 1987; Nishida and Stach, 2014; Stach and Anselmi, 2015; Stach et al., 2008; Sulston et al., 1983). Als 35 minuts després de la fecundació (mdf), a 13ºC, té lloc la primera divisió cel·lular, i es generen dues cèl·lules iguals corresponents al cantó dret i esquerre de l'animal. Als 50 mdf té lloc una segona divisió de forma simètrica, meridional i perpendicular a la primera, i als 65 mdf té lloc una tercera divisió, paral·lela al pla equatorial i asimètrica, donant lloc a 8 cèl·lules, 4 de grans en el pol animal i a 4 de petites en el pol vegetal de l'embrió. Finalment, als 85 mdf, té lloc la quarta divisió que dóna lloc a una mòrula amb 16 cèl·lules de mides diferents. La gastrulació comença amb la cinquena divisió, a l'estadi de 32 cèl·lules a les 2 hores després de la fertilització (hdf), i en la següent divisió, 64 cèl·lules, com a conseqüència de la gastrulació, totes les cèl·lules del pol vegetal ja s'han internalitzat en l'embrió, alhora que, les cèl·lules del pol animal

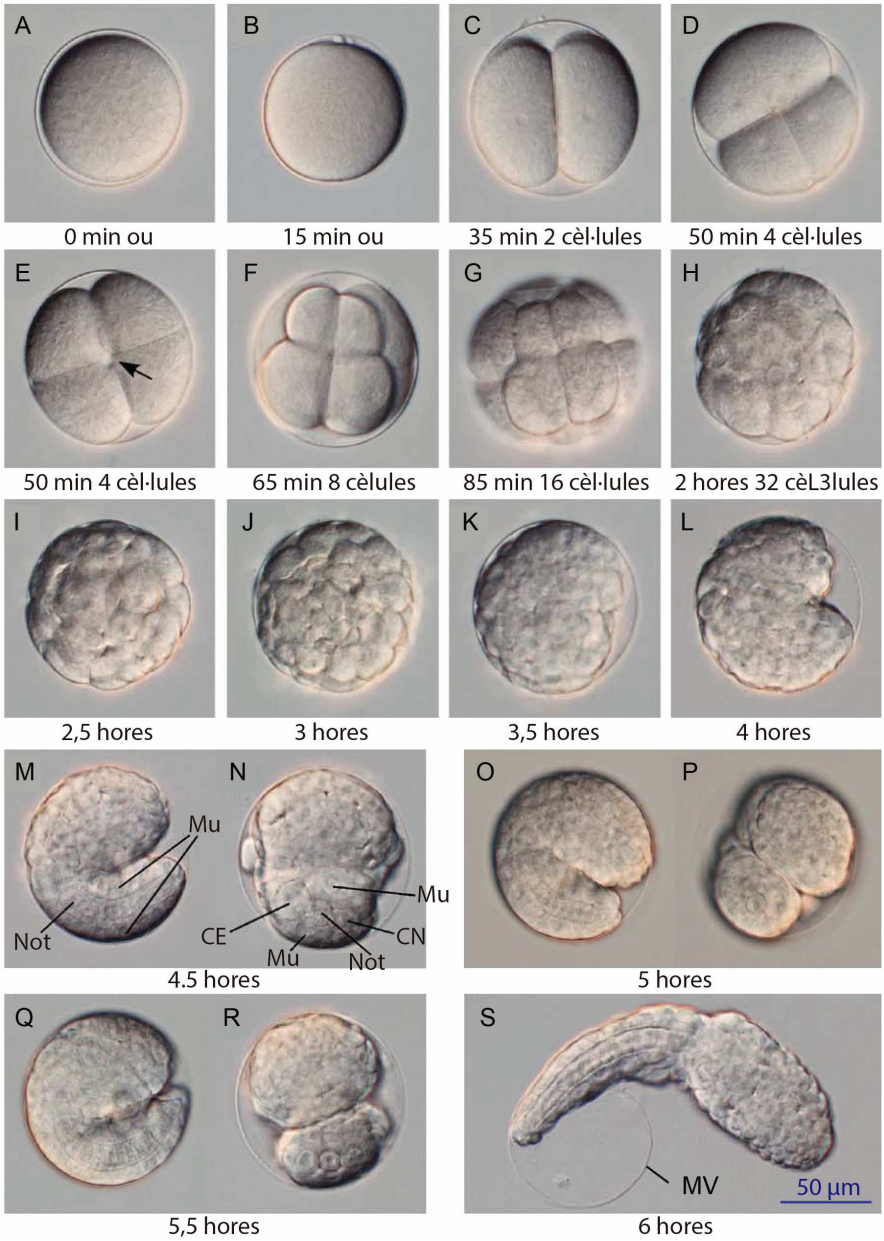


Fig. 6**. Desenvolupament embrionari d'*O. dioica***. Els temps de desenvolupament a 13˚C s'indiquen a sota de cada foto. (A) un ou no fertilitzat; té uns 80 µm de diàmetre. (B) Ou fertilitzat, s'observa l'emissió del segon cos polar. (C) Estadi de dues cèl·lules. (D-E) Dos exemples d'embrions de quatre cèl·lules. La fletxa en (E) denota el contacte entre dos blastòmers oposats en el pol animal o vegetal. (F-H) Exemples d'embrions de 8-, 16- i 32 cèl·lules, respectivament. (I-L) Estadis des de gàstrula tardana fins als estadis de *tailbud*, fotografies cada 30 minuts. (M-R) Vistes lateral dreta i ventrals en cada etapa durant elongació de la cua. En les seccions òptiques de la cua, el múscul (Mu), el notocordi (Not), el cordó nerviós (CN) i el cordó endodèrmic (CE) són visibles. La cua ha girat al 90˚ en el sentit contrari a les agulles del rellotge i per tant el cordó nerviós es situa al costat costat esquerre de l'animal. (S) *Tailbud* acabat de descloure de la membrana vitel·lina (MV). Extret de Fujii et al., 2008.

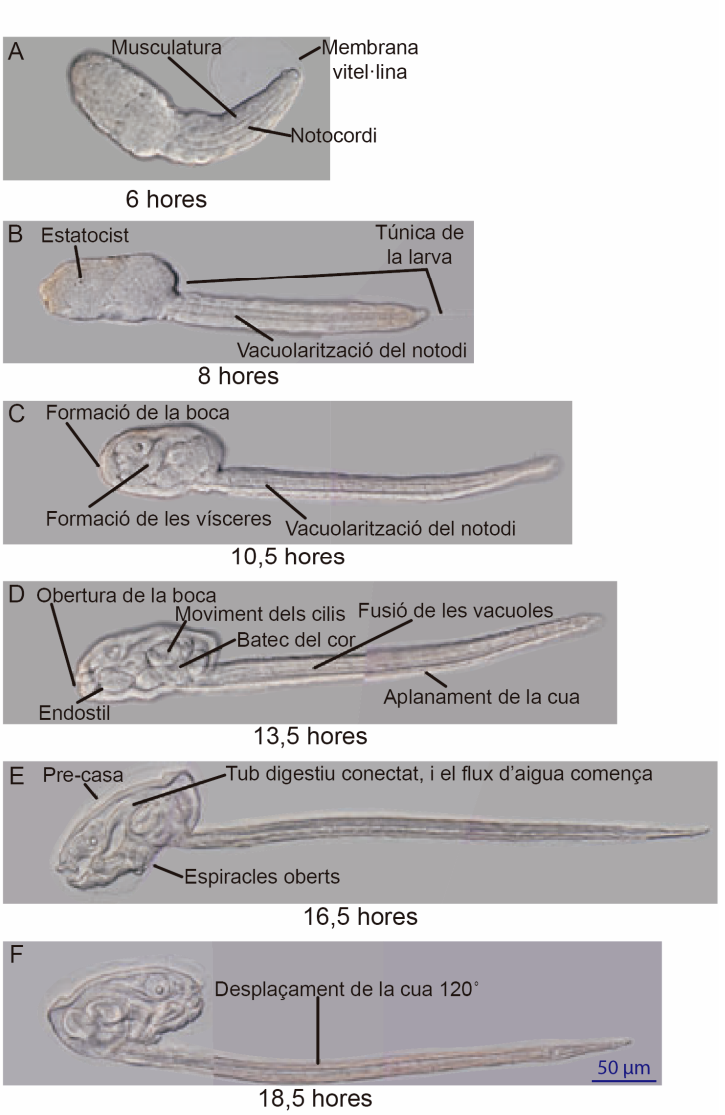


Fig. 7. **Desenvolupament larvari d'*O. dioica*.** Els temps de desenvolupament a 13˚C s'indiquen a sota de cada foto. El desenvolupament larval es subdivideix en cinc etapes de la I a la V, imatges A-E, respectivament. (F) El desenvolupament larvari acaba després del desplaçament de la cua, l'etapa VI, donant lloc a un juvenil. (B-E) Vistes laterals. (F) Vista lateral dreta després del desplaçament de la cua. Extret de Nishida, 2008.

han cobert tot l'embrió. Durant l'estadi de 64 cèl·lules, 8 cèl·lules de la regió anterior s'alineen en dues files de quatre formant una matriu que s'internalitza, aquest procés es considera la neurulació. La diferenciació del tronc i la cua comença 4 hdf, quan es comença a formar el *tailbub*. En aquest estadi, s'allarga la cua i l'embrió es doblega ventralment. Les cèl·lules del notocordi s'alineen en fila i comencen a ser evidents a les 4,5 hdf. Finalment, a les 6 hdf s'acaba l'embriogènesi, amb la desclosa (*hatch*) de la larva (Delsman, 1910; Fenaux, 1998a; Fujii et al., 2008; Stach et al., 2008).

Finalment, el desenvolupament larvari es divideix en sis etapes que comencen després de l'eclosió i duren 13h a 13ºC (fig. 7). En la fase I, just després de l'eclosió, l'animal s'estira i en pocs minuts comença a desplaçar-se propulsat pels primers moviments de la cua. A la etapa II, els límits dels òrgans comencen a aparèixer, i a l'etapa III, els òrgans ja es poden identificar. En l'etapa IV els òrgans són clarament distingibles: la boca comença a obrir-se, els cilis del tub digestiu i dels espiracles (fenedures branquials) comencen a moure's, les glàndules bucals es veuen a cada costat de l'endostil i el cor comença a batre; a la cua les vacuoles del notocordi es fusionen unes amb les altres, mentre la cua s'aplana per formar les aletes laterals. A l'etapa V, un flux d'aigua corre des de la boca als espiracle ja completament oberts, mentre tots els òrgans mostren un lumen continu i l'epidermis del tronc (l'oikoblast) comença a secretar el material mucilaginós de la casa (pre-casa o rudiment). En aquesta etapa els moviments natatoris ja són molt vigorosos. En la última etapa, la VI, després d'uns segons d'intens moviments, la cua canvia l'orientació 120º, i la punta distal es posiciona en la mateixa direcció que la boca per convertir-se en un animal juvenil, en el procés anomenat *tail shift* (Delsman, 1910; Fenaux, 1998a; Galt and Fenaux, 1990). El procés de *tail shift* s'ha comparat a la metamorfosis d'altres urocordats, però amb la significativa diferència que *O. dioica* desenvolupa els seus òrgans definitius, que es mantenen en l'adult durant el desenvolupament de la larva, al contrari del que passa en els ascidis que els òrgans definitius es formen després de la metamorfosi (Galt and Fenaux, 1990).

Morfologia i anatomia de l'animal adult

Un dels motius més importants per utilitzar models animals és que molts estudis en humans, especialment genètics, són difícils de portar a terme degut a la complexitat de la biologia humana i a les qüestions ètiques que implicarien. *O. dioica* mostra un pla corporal típic dels cordats amb un notocordi i un cordó nerviós dorsal. Així, en contraposició a altres organismes model evolutivament més llunyans, com ara mosques o cucs, l'estreta relació entre *O. dioica* i els humans permet establir assignacions inequívoques d'homologies d'òrgan, incloent el cervell, la pituïtària, la tiroide, el sistema digestiu, el cor o la cua post anal, entre altres.

El cos dels adults d'*O. dioica* esta format per un tronc (120 a 850 μm), en el que es troben la majoria dels òrgans de l'animal, i una cua (1,2 a 3,2 mm), que mou per a desplaçar-se o per generar corrents d'aigua i captar aliment (fig. 8). El cos d'*O. dioica* és transparent, de forma que les estructures internes poden ser observades amb l'animal sencer simplement canviant el pla focal del microscopi. Així doncs, l'anatomia d'*O. dioica* s'ha revisat en detall en

diversos treballs i a continuació se'n fa una síntesis (Fenaux, 1998b; Nishida, 2008).

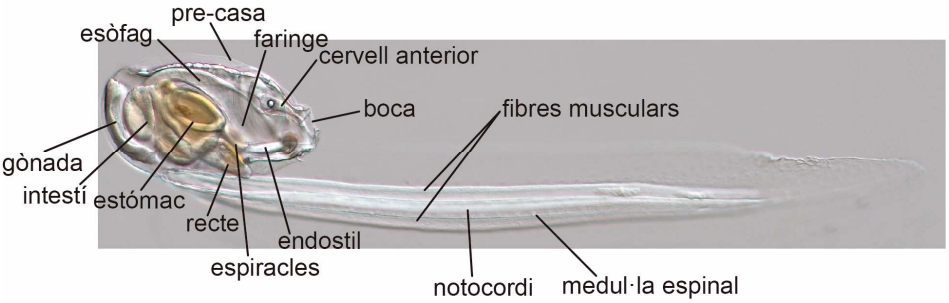


Fig. 8. **Anatomia d'*O. dioica***. Micrografia de contrast d'interferències diferencials (DIC) d'un individu d'*O.dioica* adult (Visió lateral dreta).

El sistema nerviós central (CNS). Està format per un cordó nerviós que corre dorsalment per sobre de la notocorda anomenat medul·la espinal o *spinal cord*, i que s'engruixeix en dos ganglis, el cervell anterior a la part davantera del tronc, i el gangli caudal a la part proximal de la cua, els dos ganglis contenen neurones que produeixen àcid gamma-aminobutíric (GABA) (Cañestro et al., 2005; Søviknes et al., 2005) (fig. 8). El cervell inclou un estatocist, que detecta la gravetat mitjançant un sistema d'estatolit i cèl·lules ciliades que connecten amb el cervell (Holmberg, 1984). El cervell sencer està format aproximadament per 70 cèl·lules i té funcions de rebre i integrar informació sensorial (Olsson et al., 1990; Søviknes et al., 2005). Del cervell es projecten axons a cèl·lules sensitives presents a la boca, la faringe i òrgan ventral. L'òrgan ventral consta de 30 cèl·lules i s'ha proposat que és homòleg al placode olfactiu i pituïtària de vertebrats (Bassham, 2005; Bollner et al., 1986). El cervell està connectat amb el gangli caudal per un cordó nerviós, el cordó nerviós del tronc o trunk nerve cord, que forma un angle recte i entra per la cua. Per altre banda, el gangli caudal junt amb la medul·la espinal, que conten d'unes 30 neurones són els responsables d'innervar les cèl·lules musculars de la cua, mitjançant motoneurones colinèrgiques; i coordinar els moviments de la cua (Søviknes and Glover, 2007).

L'epidermis. L'epidermis de d'*O. dioica* és un epiteli monoestratificat sense cap capa de teixit mesodèrmic per sota. L'epiteli que recobreix el tronc, oikoblast o epiteli oikoplàstic, s'encarrega de la secreció de la casa d'*O. dioica* (fig. 8). Així doncs es tracta d'un epiteli amb un nombre fix de cèl·lules organitzat en diferents camps definits per la forma de les cèl·lules i la morfologia dels nuclis d'aquestes. El patró de les cèl·lules del oikoblast té una correlació directa amb les diferents estructures arquitectòniques de la casa

(Kishi et al., 2017; Spada et al., 2001; Thompson et al., 2001).

La regió faringo-branquial i endostil. La cavitat faríngia, connecta la boca, situada a la part anterior del tronc, amb el sistema digestiu (fig. 9). Així doncs, les partícules filtrades per la casa entren per la boca i passen a la faringe gràcies al flux d'aigua generat pels espiracles, unes obertures branquials amb cilis a l'interior i que connecten directament a l'exterior del cos. Les bandes perifaríngies, dues bandes simètriques de cilis que creuen les parets laterals de la faringe per acabar trobant-se al sostre d'aquesta, s'encarreguen d'estendre per la faringea un moc que atrapa les partícules filtrades que han entrat per la boca (Fenaux, 1998b). Aquest moc el secreta l'endostil, un òrgan homòleg a la tiroide de vertebrats (Cañestro et al., 2008), que és troba situat ventralment a la faringe (Olsson, 1963).

El sistema digestiu. El tub digestiu d'*O. dioica* consta d'un esòfag dorsal, un estómac bilobulat (esquerra i dret) en forma de sac i un intestí corbat dividit en intestí vertical, mig i distal (o recte) (Burighel and Brena, 2001) (fig.9). Els diferents compartiments del tub digestiu tenen funcions diferents, encara que l'absorció de líquid, ions i petites molècules es dóna al llarg de tot el tracte digestiu. La major activitat digestiva es dóna al lòbul gàstric esquerra que presenta una gran varietat d'activitats d'enzims digestius (α-amilasa, fosfatasa àcida, esterases no específiques, 5'-nucleodidasa i aminopeptidasa M)(Cima et al., 2002). En el lòbul gàstric dret s'emmagatzema el material digerit en gotes de lípids, mentre que les cèl·lules epitelials d'aquest compartiment secreten una membrana peritròfica que envolta el menjar. Els pel·lets fecals es formen en l'intestí vertical on també s'emmagatzemen nutrients. Els pel·lets fecals avancen fins el recte on es segueix produint l'absorció i s'emmagatzemen material proteic dins de gränuls de les cèl·lules ciliades granulars abans que pel·lets siguin eliminats per l'anus (Burighel and Brena, 2001; Cima et al., 2002).

El notocordi. El notocordi és una estructura sinapomòrfica del fílum dels cordats (Satoh et al., 2012). En les larves d'*O. dioica*, el notocordi esta format per una fila de cèl·lules, o pila de monedes, que s'estenen al llarg de la cua. Durant el desenvolupament larvari és produeix el procés de vacuolació de les cèl·lules del notocordi (fig. 7), de manera el notocordi passa de ser un tub massís a ser un tub d'epiteli (Søviknes and Glover, 2008). Així doncs, el notocordi dels adults d'*O. dioica* és un tub de parets primes format per cèl·lules aplanades que segreguen proteïnes riques en sofre al medi extracel·lular que donen rigidesa al notocordi, però també prou flexibilitat per a facilita la natació i altres moviments (Olsson, 1965).

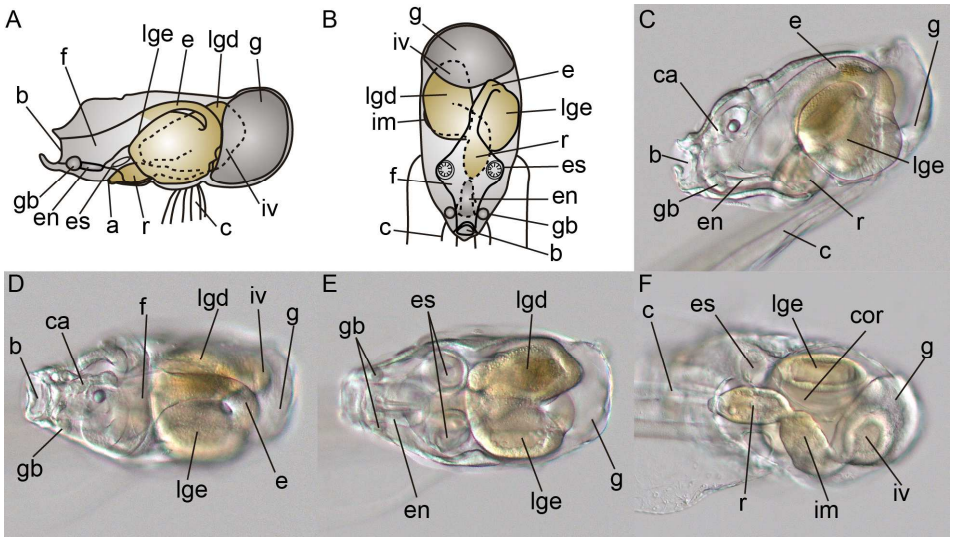


Fig. 9. **Sistema digestiu d'*O. dioica*.** Esquema del tronc i anatomia del tracte digestiu d'*O. dioica* des de l'esquerra (A) i dorsalment (B). Extret de Burighel and Brena, 2001. (C-F) Micrografies de contrast d'interferències diferencials (DIC) del tronc d'un individu d'*O. dioica* adult en visió lateral esquerra (C), dorsal a dos plans focals diferents (D-E) i ventral (F). Anus (a), boca (b), cervell anterior (ca), cor (cor), cua (c), endostil (en), esòfag (e), espiracle (es), faringe (f), glàndula bucal (gb), gònada (g), intestí mig(im), intestí vertical (iv), lòbul gàstric dret (lgd), lòbul gàstric esquerre (lge), recte (r).

El sistema circulatori. Per l'espai que hi ha entre l'ectoderm i els òrgans interns circula una hemolimfa, mancada de cèl·lules, impulsada pels moviments de la cua i el cor que esta situat ventralment entre el lòbul esquerra de l'estómac i l'intestí (fig.9F). El cor, que pot invertir el sentit de la circulació, esta format per una paret muscular ondulant, l'esquerra, i una membrana llisa no muscular, la dreta (Onuma et al., 2017a; Salensky, 1903).

Musculatura de la cua. La musculatura de la cua dels oikopleurids adults està composta per dues bandes, una dorsal i l'altre ventral, de 10 cèl·lules musculars que s'estenen al llarg de l'eix anteroposterior (Nishino et al., 2000) (fig. 8). En les larves acabades de descloure només hi ha 8 parells de cèl·lules i durant del desenvolupament larvari apareixen els dos parells que falten, encara que el seu origen no es coneix (Nishino and Satoh, 2001). Respectes a les fibres musculars, aquestes només es troben a la part interior de la cèl·lula muscular, mentre que en l'altre costat es concentren els mitocondris (Nishino et al., 2000).

El sistema reproductor. O. dioica, tal i com el seu nom indica, presenta individus de diferents sexes, i és l'única espècie de tot el subfílum dels urocordats que presenta aquesta característica (Nishida, 2008). Les gònades que se situen en la part posterior del tronc, són l'únic caràcter que permet distingir entre mascles i femelles d'*O. dioica.* A mesura que l'animal creix, les gònades

adquireixen la seva mida entera i maduren (fig. 10A). És llavors quan es

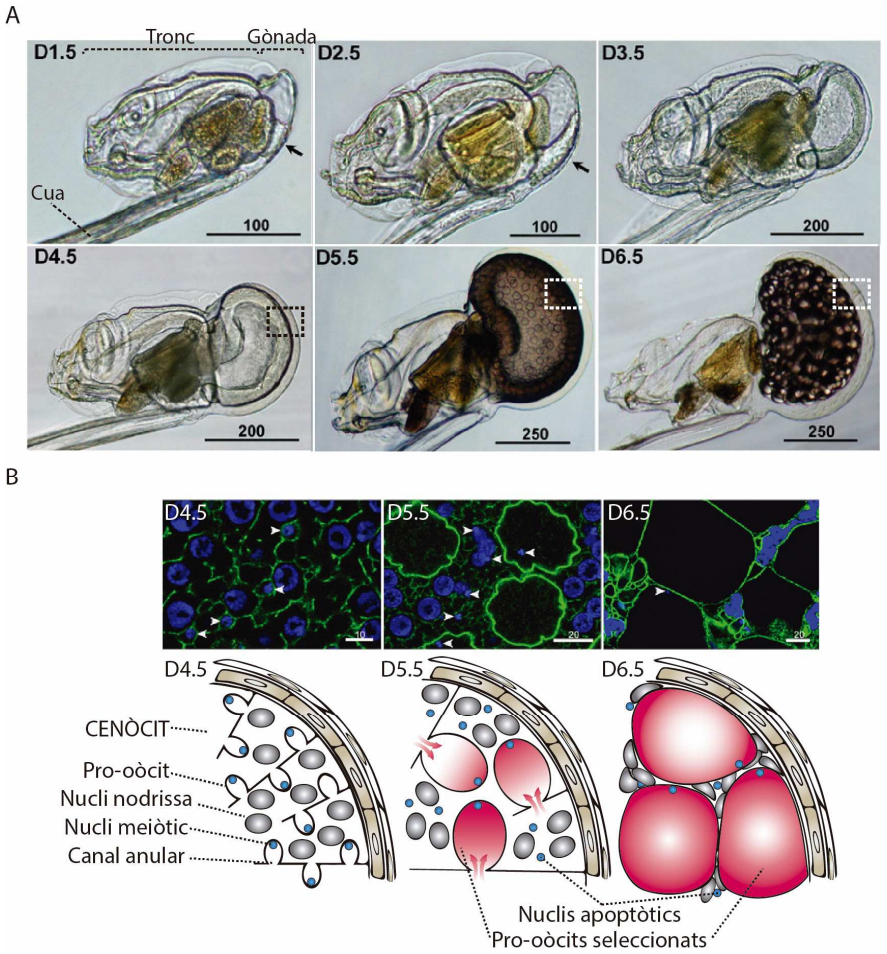


Fig. 10. **Sistema reproductor d'*O. dioica***. (A) Micrografies de camp clar de femelles d'O. dioica des del dia 1,5 al dia 6,5, desenvolupades a 15˚C. Les femelles estan orientades amb la part anterior (boca) a l'esquerra, la posterior (gònada) a la dreta i la part ventral (cua) a baix. Després de la metamorfosi al dia 0.5, la gònada externa transparent apareix posterior al tronc (fletxes als dies 1,5 i 2,5) i creix contínuament. Al dia 5,5, els oòcits en desenvolupament són visibles i al dia 6.5 els oòcits ja madurs omplen tot el volum de l'ovari abans de la posta. Extret de Ganot et al., 2007. (B) Imatges confocals en detall (quadrat en A) del cenòcit en diferents etapes de l'oogènesi d'O. dioica (panell superior): F-actina, verda; DNA, blau. Esquemes d'anatomia cel·lular a D4,5, D5,5 i D6,5 (panell inferior). A dia 3 la meitat dels milers de nuclis germinals entra en meiosis mentre que l'altre meitat augmenta la ploïdia a través de múltiples rondes d'endoreplicació formant el que s'anomena nuclis nodrissa (esferes grises). El dia 4,5 es caracteritza per un creixement massiu del cenòcit. Mentre que una xarxa interna complexa de F-actina defineix pseudo-compartiments (pro-oòcits) que envolten cada nucli meiòtic (esferes blaves) que s'obren al citoplasma general a través d'un canal de anular. El dia 5,5 tots els nuclis meiòtics entren a la metafase I. Paral·lelament, un subconjunt de pro-oòcits creix simultàniament adquirint el citoplasma del cenòcit a través del canal anular. Només un subconjunt de nuclis meiòtics continua la meiosi. Al final del cicle de vida (dia 6,5), els nuclis meiòtics que no han crescut i els nuclis nodrissa entren en apoptosi, deixant una gònada plena de oòcits en metafase sincronitzats. Extret de Ganot et al., 2008.

produeix la gametogènesis que és ràpida i només dura unes hores (Ganot et al., 2007; Nishino and Morisawa, 1998). L'espermatogènesi es dóna en el testicle, un sol síncit que interconnecta un nombre substancial de nuclis, els quals tots esdevindran espermatozoides (Martinucci et al., 2005; Onuma et al., 2017a). Per altre banda, l'ovari també és un sol síncit (anomenat cenòcit) i conté un nombre considerable de nuclis meiòtics i nuclis nodrissa. Cada nucli meiòtic està inclòs dins d'un pro-ovòcit interconnectat per un canal anular (ring canal) al citoplasma comú amb els seus nuclis nodrissa poliploides. Durant la ovogènesis, alguns dels pro-ovòcits començaran a créixer incorporant el citoplasma comú de l'ovari i esdevindran els ous que la femella fresarà, mentre que la resta de pro-oòcits no progressaran i degeneraran (Ganot et al., 2006; Ganot et al., 2007; Ganot et al., 2008)(fig. 10B). Així doncs, s'ha aprofitat que la estructura interna del ovari d'*O. dioica* permet que qualsevol líquid injectat en l'ovari difongui a gran part de la gònada per introduir àcids nucleics exògens en els oòcits d'*O. dioica* i així aconseguir la genoanul·lació de gens diana (Omotezako et al., 2013; Omotezako et al., 2015; Omotezako et al., 2017).

Sistemes d'estabulació i tècniques de biologia molecular per a *O. dioica*

Un bon model animal ha de disposar de sistemes de cultiu que permetin la seva estabulació i ens asseguri la disponibilitat il·limitada de material biològic pels experiments, així com de tècniques de biologia molecular que permetin portar a terme investigacions funcionals en el laboratori.

Respecte l'estabulació d'*O. dioica* en el laboratori, malgrat que fa més de 40 anys que s'han descrit sistemes de cultiu que permetien mantenir *O. dioica* diverses generacions en el laboratori (Fenaux and Gorsky, 1979; Fenaux and Gorsky, 1985; López-Urrutia and Acuña, 1999; Paffenhöfer, 1973; Sato et al., 1999), al inici d'aquest projecte de tesis tan sols dos laboratoris en el món tenien la capacitat de mantenir *O. dioica* durant tot l'any, un a l'institut SARS a Bergen (Noruega) (Bouquet et al., 2009; Chioda et al., 2002) i l'altre a l'Universitat d'Osaka (Japó) (Nishida, 2008).

Respecte a les tècniques d'anàlisi funcional, a més de les metodologies més clàssiques com les hibridacions *in situ* (Bassham and Postlethwait, 2000; Henriet et al., 2015; Onuma et al., 2017b) o les immunohistoquímiques (Onuma et al., 2017b; Subramaniam et al., 2014), un bon model ha de disposar d'eines d'enginyeria genètica que permetin alterar *in vivo* les funcions dels seus gens. Per alterar la funció dels gens en *O. dioica* s'han utilitzat diverses estratègies que van des de tractaments farmacològics amb productes que alteren l'expressió dels gens (Cañestro and Postlethwait, 2007) fins a l'edició del genoma mitjançant la tecnologia CRISPR (Deng and Chourrout, 2017), passant per la

microinjecció, tant en ous sense fertilitzar com en els ovaris de les femelles, de morfolinos (Sagane et al., 2010), RNA d'interferència (Mikhaleva et al., 2015; Omotezako et al., 2013) o fins i tot DNA d'interferència (DNAi) (Omotezako et al., 2015; Omotezako et al., 2017). Són aquestes dues tècniques de genoanul·lació, RNAi i DNAi, les més consolidades a l'hora de manipular l'expressió gènica a *O. dioica*.

El context filogenètic

O. dioica pertany al subfílum dels urocordats (o tunicats), el grup germà dels vertebrats, els quals juntament amb els cefalocordats constitueixen el nostre propi fílum, els cordats (fig. 11)(Delsuc et al., 2006; Tsagkogeorga et al., 2009). El subfílum dels urocordats està compost només per animals marins i, tradicionalment, s'ha dividit en tres classes, la classe Ascidacea, la Thaliacea i la Appendicularia (fig. 11). La classe Ascidiacea és la classe més gran en quant al nombre d'espècies i presenta adults sèssils que viuen en la zona bèntica. Les altres dues classes, Thaliacea i Appendicularia, constitueixen una gran part del zooplàncton gelatinós i les trobem errant al llarg de la columna d'aigua. Mentre que la classe Thaliacea esta composta per salpes, doliòlids i pirosomes, la classe Appendicularia (o Larvacea) només la compon un grup, les apendiculàries. Les apendiculàries també reben el nom de larvacis ja què retenen la morfologia larval junt amb les sinapomorfes dels cordats (el notocordi i tub nerviós dorsal) al llarg de tot el cicle de vida, aquesta característica permet la comparació amb els vertebrats més fàcilment, a diferència de la resta d'urocordats que pateixen una metamorfosi dràstica dificultant la comparació d'estructures amb els vertebrats (Holland, 2016).

A nivell filogenòmic, els urocordats tenen especial interès ja que van divergir abans de les dues rondes de duplicació completa del genoma (WGD) que es van produir a principis del llinatge de vertebrats i, per tant, la robustesa mutacional en els urocordats, i per inclusió en *O. dioica*, és molt menor que en els vertebrats. Aquest fet, junt amb la disponibilitat de genomes completament seqüenciats de diversos cordats -3 espècies cefalocordats, 11 espècies d'urocordats i més de 100 vertebrats- proporciona un marc evolutiu ideal per a la identificació d'esdeveniments de pèrdua de gens a *O. dioica* mitjançant genòmica comparativa amb altres cordats (fig.11).

Característiques gèniques i genòmiques

Plasticitat genòmica i complexitat genètica

La seqüenciació del genoma d'*O. dioica* va posar de manifest que aquest havia

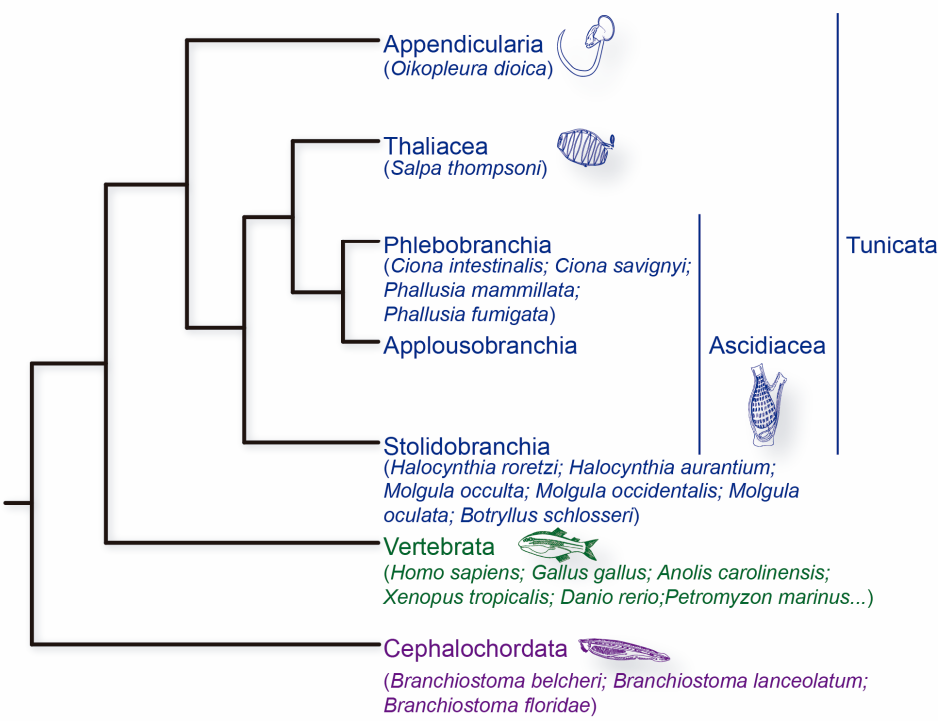


Fig. 11. **Filogènia dels cordats**. Arbre filogenètic reconstituït mitjançant l'aproximació Bayesiana que evita els artefactes per atracció de branques llargues (*long-branch attraction*). Tingueu en compte que els tunicats i els vertebrats són grups germans, així com els taliacis i phlebobranquis dins dels ascidis. Entre parèntesis alguna de les especies seqüenciades per aquell subfílum. Modificat de Lemaire and Piette, 2015.

patit una extrema compactació. De fet, amb només 70 megabases (Mb) el genoma d'*O. dioica* és un dels genomes més petits entre tots els metazous (Denoeud et al., 2010), superat només per genomes extremadament reduïts d'animals paràsits -p. e. el nematode paràsit de plantes *Pratylenchus coffeae,* 19,7 Mb, el nematode paràsit de invertebrats marins *Intoshia linei,* 43.2 Mb o el cnidari paràsit de anèl·lids i vertebrats *Kudoa iwatai*, 22.5 Mb, (Burke et al., 2015; Chang et al., 2015; Mikhailov et al., 2017)- (fig. 12). La reducció i compactació del genoma d'*O. dioica* és pot explicar per quatre factors. En primer lloc, en el genoma d'*O. dioica* hi hagut una pèrdua significativa d'elements genètics transponibles (TEs) (Denoeud et al., 2010; Volff et al., 2004). En segon lloc, *O. dioica* ha sofert una disminució en la mida de les regions intergèniques, sent inferiors a 1 kb en el 53% dels gens i arribant a l'extrem de desaparèixer en els quasi 5000 gens organitzats en menys de 1800 operons (Denoeud et al., 2010). En tercer lloc, a nivell intragènic, *O. dioica* ha patit una extrema reducció en la mida dels introns assolint mides menors de 50 pb en el 62% dels gens (Denoeud et al., 2010; Seo, 2001). A més, *O. dioica* mostra una elevada velocitat en recanvi de les posicions intròniques, el que ha portat a que 76% de les posicions intròniques no estiguin conservades respecte

els gens homòlegs d'altres organismes (Berna and Alvarez-Valin, 2014; Denoeud et al., 2010; Edvardsen et al., 2004). Finalment, la seqüenciació del genoma d'*O. dioica* va evidenciar la impossibilitat de detectar molts grups de gens presents en cordats ancestrals, la qual cosa suggeria una pèrdua de gens elevada en aquesta espècie que contribuiria a la compactació del genoma d'*O. dioica*. Cal mencionar, però, que *O. dioica* també ha sofert nombroses duplicacions gèniques –p.e. entre els 15 grups de gens homeòtics, *O. dioica* hauria guanyat 19 gens (per a un total de 36 gens) al comparar-la amb l'ancestre comú dels tunicats (Edvardsen et al., 2005)- (Chavali et al., 2011; Denoeud et al., 2010). Així doncs, el balanç de nombrosos guanys i pèrdues gèniques, ha portat a que avui en dia *O. dioica* tingui aproximadament 18000 gens, un número similar al d'altres urocordats (e.g. *C. intestinalis* 15300) i tan sols lleugerament inferior al d'altres cordats com els cefalocordats (e.g. *B. floridae* 22000) o els vertebrats (e.g. *F. rubripes* 18300 o *H. sapiens* 20300) (fig. 12).

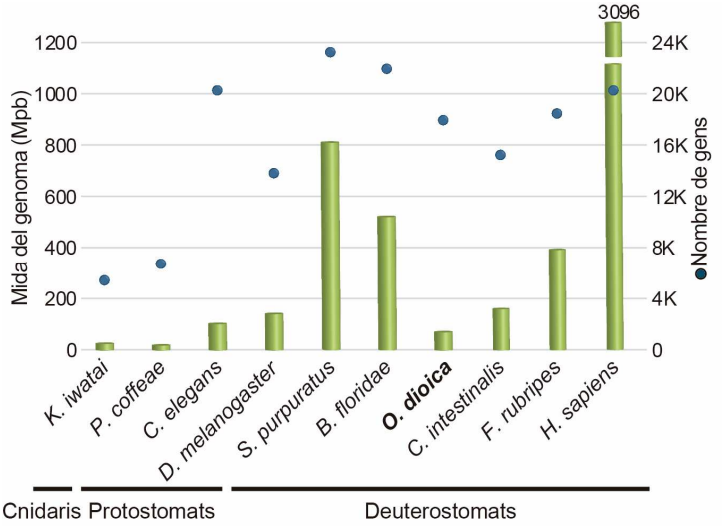


Fig. 12. **Comparació de la mida del genoma** (barres verdes) **i el nombre de gens** (cercles blaus) **entre els metazous.** O. dioica presenta el genoma més petit, 70 Mb, d'entre els genomes dels deuteròstoms descrits fins el moment. Com a referència, entre els cordats, la mida del genoma de l'ascidi Ciona és de 160 Mb; la de l'amfiox Branchiostoma floridae, 520 Mb; la del peix fugu (Fugu rubripes), 393 Mb; i la de l'humà, 3096 Mb; fora dels cordats la del eriço de mar (Strongylocentrotus purpuratus) 814 Mb i entre els protòstoms, la mida del genoma de Drosophila melanogaster és de 143,7 Mb; la de Caenorhabditis elegans, 100,3 Mb mentre que la mida del genoma metazou més petit és la del nematode paràsit de plantes Pratylenchus coffeae amb 19,7 Mb seguit d'aprop pel cnidari mixozou Kudoa iwatai amb 22,5 Mb (Berna and Alvarez-Valin, 2014; Burke et al., 2015; Chang et al., 2015; Sodergren et al., 2006; Yates et al., 2016). Per detalls en quan el nombre gens mirar el text.

Pèrdues gèniques

Una de les observacions més interessants sobre el genoma de *O. dioica* és que diversos gens presents en genomes de vertebrats, i fins i tot en altres animals més ancestrals, semblen estar absents al genoma d'*O. dioica* (Denoeud et al., 2010). Aquest fet suggereix una pèrdua de gens elevada en *O. dioica*. No obstant això, cal mencionar que la identificació d'homòlegs en el genoma d'*O. dioica* és difícil, ja que en ser una espècie que evoluciona molt ràpid –així ho demostren les branques sorprenentment llargues en els arbres els filogenètics (Delsuc et al., 2006; Tsagkogeorga et al., 2010)- les semblances entre els gens d'*O. dioica* i els seus ortòlegs de vertebrats s'hagin perdut.

Entre els nombrosos grups de gens que s'han perdut a *O. dioica*, cal nombrar en primer lloc la pèrdua de tots els gens del sistema de reparació de l'ADN per la unió d'extrems no homòlegs (NHEJ), ja que probablement és un dels motius que han contribuït a la elevada pèrdua de gens d'aquest organisme (Denoeud et al., 2010). Altres exemples a mencionar de pèrdua de gens que s'han donat en aquesta espècie són aquells que afecten la maquinària epigenètica, el sistema immune, diverses famílies de caspases, els sistema de defensa contra xenobiòtics per senyalització AhR o el repertori de microRNAs (per més detall i referències consultar Taula 1). Entre els gens clau del desenvolupament, *O. dioica* ha perdut més del 30% (26 de 81) dels gens amb homeobox ancestrals en els bilaterals (Edvardsen et al., 2005) (Taula 1), el clúster dels gens Hox s'ha desintegrat i s'ha vist reduït a tres gens Hox anteriors (Hox1, Hox2 i Hox4) i sis de posteriors (Hox9A, Hox9B, Hox10, Hox11, Hox12 i Hox13) i cap dels centrals (Seo et al., 2004). En quan a les vies de senyalització que actuen durant el desenvolupament, és d'especial atenció, el cas de la pèrdua dels principals gens que participen en la senyalització per àcid retinoic. (all-trans-RA o RA) (Cañestro and Postlethwait, 2007; Cañestro et al., 2006). La rellevància que té l'RA en el desenvolupament embrionari i l'absència d'alguns components clau en el metabolisme i senyalització d'aquest morfogen a *O.dioica* fan d'aquest retinoide i les vies de senyalització que hi interactuen candidats idonis per analitzar l'impacte de les pèrdues gèniques en l'Evo-Devo.

Llista de gens absents en el genoma d'*O. dioica*

Categoria	Gens absents	Referencies
Sistema de reparació de l'ADN	POLB; APEX2; LIG3; MSH3; ATM; CHEK2; APTX; NBN; RAD52; XRCC5 (Ku80)*; XRCC6 (Ku70)*; LIG4*; XRCC4*; NHEJ1*; DNA-PKc*; DCLRE1C*	Denoeud et al., 2010
Maquinària epigenètica	Dnmt1; Dnmt3; Mbd4/MeCP2 GCN5/PCAF; HAT1	Albalat et al., 2012; Cañestro et al., 2007 Navratilova et al., 2017
Sistema immune	NLRs; RLHs; MyD88-like; SARM1-like; TIRAP-like; TICAMP2-like	Denoeud et al., 2010
Família de les Caspases	CSP1/4/5 (CARD); CSP2/9 (CARD); CSP8/10 (DED/DED); CSP6	Weill et al., 2005
Sistema de senyalització AhR	AhR; AhRR	Yadetie et al., 2012
microRNAs	miR-9; miR-29; miR-33; miR-34; miR-96; miR-126; miR-133; miR-135; miR-153; miR-182; miR-183; miR-184; miR-196; miR-200; miR-216; miR-217; miR-218; miR-367	Fu et al., 2008; Wang et al., 2017
Gens diana de Brachyury	entactin; fibrinogen-like; multidom; myomegalin; Noto1; Noto2; Noto3; Noto4; Noto5; Noto6; Noto7; Noto8; Noto11; Noto12; Noto13; Noto14; Noto16; perlecan; PTP; SLC; SWIPI; tropomyosin-like; tune; UBE	Kugler et al., 2011
Gens homeobox	Gbx; Nk3; TGIF; POU VI; Lhx6/7; Vax; Cux; SATB; ZFH1; Sax; Xlox; Mox; Hlx; Bsh; Chox10; Otp; Prx; Goosecoid; NK6; Prox; Tlx	Edvardsen et al., 2005
Gens Hox	Hox3; Hox5; Hox6; Hox7; Hox8	Seo et al., 2004
Senyalització per àcid retinoic	Aldh1a; CYP26; RAR	Cañestro and Postlethwait, 2007

*gens del sistema de reparació de l'ADN per la unió extrems no homòlegs (NHEJ).

L'àcid retinoic

Durant el desenvolupament embrionari dels vertebrats, l'àcid retinoic (all-trans-RA o RA) actua com a morfogen regulant el patró axial al llarg dels eixos anterior-posterior i dorso-ventral, controlant el desenvolupament de les extremitats, la neurogènesi i l'organogènesi (Cunningham and Duester, 2015; Duester, 2013; Rhinn and Dolle, 2012). En l'adult, l'RA participa en la homeòstasi d'òrgans, en la regeneració de teixits i d'òrgans, i en la prevenció del creixement neoplàsic i de les malalties neurodegeneratives (Altucci and Gronemeyer, 2001; De Luca, 1991; Maden, 2007; Maden and Hind, 2003). Per exercir la seva acció, l'RA s'uneix al receptor nuclear RAR (*retinoic acid receptor*) que forma un heterodímer amb un segon receptor nuclear, el promiscu RXR (*retinoid X receptor*), per a reconèixer les seqüències RARE (*retinoic acid response element*) de les regions reguladores dels gens diana (p.e. *Hox1*). L'heterodímer RAR-RXR controla l'expressió gènica gràcies a la interacció amb complexos co-activadors com membres de la família de proteïnes p160 que indueixen l'acetilació de les histones i l'activació de la transcripció (Gutierrez-Mazariegos et al., 2014), o en la forma apo, sense lligant, amb co-repressors transcripcionals, tals com NCor o SMRT, per a reprimir la transcripció de gens.

L'RA es sintetitza a partir dels retinoides i carotenoides presents a la dieta provinents de parts d'animals riques en retinoides com el fetge o en carotenoides com el teixit adipos, a més els carotenoides també els trobem en els vegetals amb intensos colors grocs, taronges o vermells, ja que els carotenoides són un grup important dels pigments vegetals liposolubles. El principal retinoide per a la síntesis de l'RA és el all-trans-retinol (vitamina A), i deficiències en aquesta vitamina (VAD, *vitamin A deficiency*) estan relacionades amb trastorns a la pell i la visió (Hale, 1933; Wolbach and Howe, 1925), mentre que l'excés d'aquesta vitamina produeix una baixa densitat òssia i malformacions en l'embrió (Collins and Mao, 1999; Melhus et al., 1998). L'hidroxil present en un dels extrems de la vitamina A pot ser doblement oxidat, primer a un grup aldehid (retinal) i després a un grup carboxílic per a sintetitzar RA, o pot esterificar amb un àcid –freqüentment àcid palmític– per formar èsters de retinil que és com els retinoides s'emmagatzemen en els organismes. L'RA pot ser també produït a partir de carotenoides, principalment a partir del trencament simètric de β-carotè en dues molècules de retinal que són posteriorment oxidades a RA. Un cop l'RA ha exercit la seva acció, és degradat

a formes biològicament inactives com ara el 4-hidroxi-RA o el 4-oxo-RA. És el metabolisme de l'RA –la seva síntesi i degradació– el que defineix en darrer terme els nivells i la localització de l'RA en les diferents parts del cos dels animals i el que determina, per tant, la seva acció com a morfogen en els organismes.

La xarxa gènica del metabolisme de l'RA (RA-MGN)

En vertebrats, el metabolisme dels retinoides és un procés que implica la participació de moltes proteïnes (fig. 13), no només enzims sinó també proteïnes d'unió (*binding proteins*) i de transport. A l'intestí per exemple, el retinol, el principal retinoide de la dieta junt amb els seus esters, s'uneix a les Crbp (*cellular retinol binding protein*) per estabilitzar-lo en el medi aquos i transportar-lo a través del citoplasma fins al reticle endoplasmàtic, on l'enzim

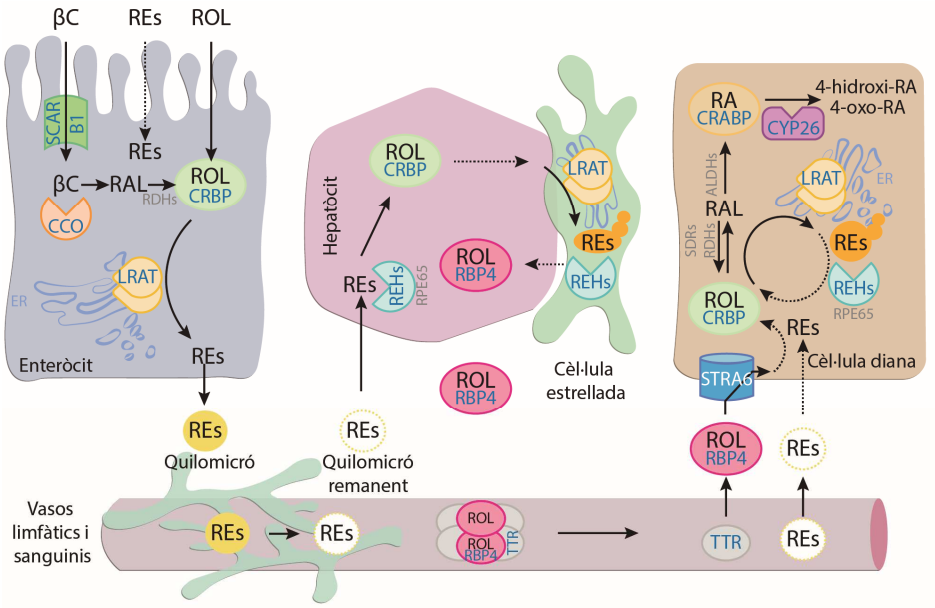


Fig. 13. **Xarxa gènica del metabolisme de l'RA en vertebrats.** El metabolisme dels retinoides implica a moltes proteïnes encarregades de la seva absorció, transport i emmagetzament. Per una explicació detalla de les proteïnes implicades mireu el text. Les abreviatures usades en el diagrama són: 11cRAL, 11-cis-retinal; ADHs, *alcohol dehydrogenases*; ALDHs, *aldehyde dehydrogenases*; βC, β-carotè; BCO1, *β-carotene-15,15-dioxygenase*; CRABPs, *cellular retinoic acid binding proteins*; CRBP1, *cellular retinol binding protein 1*; CRBP2, *cellular retinol binding protein 2*; ER, reticle endoplasmàtic; LRAT, *lecithin:retinol acyltransferase*; RA, all-trans_retinoic acid; RAL, *all-trans-retinaldehyde*; REHs, *retinyl ester hydrolases*; REs, *retinyl esters*; RESs, *retinyl esterases*; RPE65, *retinoid isomerase*; ROL, all-trans_retinol; RBP, retinol binding protein; SCARB1, *scavenger receptor class B, type I*; SDRs, *short chain dehydrogenases/reductases*; STRA6, *stimulated by retinoic acid 6* i TTR, *transthyretin*. Basat en Chelstowska et al., 2016.

Lrat (*lecithin-retinol acyltransferase*) catalitza la seva esterificació (Wongsiriroj et al., 2008). Paral·lelament, el receptor Scarb1 (*scavenger receptor class B Type 1*) facilita l'absorció dels carotenoides en les cèl·lules de la mucosa intestinal (Widjaja-Adhi et al., 2015) on enzims de la família Cco tallen el β-carotè per produir retinal. Si el trencament és simètric, catalitzat per la Bco1 (*β-carotene oxygenase 1*), es generen directament dues molècules de retinal, mentre que si el trencament és asimètric, catalitzat per la Bco2 (*β-carotene oxygenase 2*), el retinal es produeix indirectament pel processament dels subproductes per la via de la β-oxidació dels àcid grassos dins dels mitocondris (Wang et al., 1996). En qualsevol cas, i encara en l'intestí, les retinal reductases (p. e. les Rdh11-14, Dhrs4) converteixen el retinal a retinol, i aquest s'exportarà al fetge en forma de èster de retinil (Belyaeva and Kedishvili, 2002; Belyaeva et al., 2005; Belyaeva et al., 2008; Kedishvili et al., 2002; Lei et al., 2003).

Els èsters de retinil del fetge constitueixen el principal reservori de retinoides i són mobilitzats per respondre a les necessitats fisiològiques de l'organisme. Els èsters són hidrolitzats per diferents REHs (*retinyl ester hydrolases*, incloent la Rpe65 de la família de les Cco), i el retinol produït és transportat fins els teixits diana a través del torrent sanguini formant un complex amb la Rbp4 (*retinol binding protein 4*) i la TTR (*transthyretin*). Quan arriba a les cèl·lules diana, el receptor Stra6 (*stimulated by retinoic acid gene 6*) en gestiona l'entrada (Kawaguchi et al., 2015) i un cop dins, les Crbp s'uneixen al retinol per estabilitzar-lo i controlar-ne la seva oxidació fins a RA. La primera oxidació a retinal és una reacció reversible catalitzada per diverses SDR-Rdh (*short-chain retinol dehydrogenases*), entre les quals trobem la Rhd10 (*retinol dehydrogenase 10*), la Rdh16 (*retinol dehydrogenase 16*, també coneguda com a Rdh4 o RoDH4), o la RdhE2 (*epidermal retinol dehydrogenase 2*). Les Aldh1a (*aldehyde dehydrogenase 1a*), també conegudes com Raldh (*retialdehyde dehydrogenase*), o altres Aldh com les Aldh8a1, oxiden irreversiblement el retinal a RA que s'uneix a la Crabp (*cellular retinoic acid binding protein*) per protegir-lo i estabilitzar-lo. Finalment, l'RA pot ser inactivat per les Cyp26 (*cytochrome P450, family 26*) que catalitzen la seva oxidació a metabòlits biològicament inactius com el 4-hidroxi-RA o el 4-oxo-RA (Albalat, 2009). En resum, la quantitat i la localització de l'RA venen determinats pel grau i el lloc d'expressió dels gens que codifiquen els enzims i les proteïnes del seu metabolisme. Aquest conjunt de gens que constitueixen una xarxa gènica complexa anomenada RA-MGN (RA-*metabolic gene network*) establiria de forma precisa els nivells fisiològics i la distribució espai-temporal del RA durant el cicle de vida dels organismes.

La via de l'RA en l'evolució dels animals

En els darrers anys, la perspectiva de l'RA com una via de senyalització important pels animals ha canviat substancialment. Inicialment proposada com una innovació dels cordats, la espècies d'invertebrats ha portat a concloure que el seu origen és molt més antic, molt probablement abans de la diversificació dels bilaterals (Albalat and Cañestro, 2009; Cañestro et al., 2006; Simões-costa et al., 2008). La funció ancestral de la via és però controvertida i s'ha proposat que originalment l'RA podria estar relacionat amb processos de diferenciació neuronal més que en l'establiment del patró axial, el qual podria ser una funció sorgida durant l'evolució dels cordats (Carter et al., 2015; Dmetrichuk et al., 2006; Estephane and Anctil, 2010; Handberg-Thorsager et al., 2018). En qualsevol cas, el paper de l'RA en la formació del pla corporal de tots els cordats semblava ben establert ja que s'havia demostrat per espècies dels tres subfila, en el cefalocordat amphioxus, en l'urocordat ascidi, i en nombroses especies de vertebrats, des de peixos fins a mamífers (Diez del Corral et al., 2003; Escriva et al., 2002; Grandel et al., 2002; Nagatomo and Fujiwara, 2003).

De forma inesperada, l'anàlisi del genoma de l'urocordat *O. dioica* va revelar l'absència de parts importants de la RA-MGN (i.e. Aldh1a i Cyp26) així com d'un dels dos receptors nuclears, l'RAR, en aquesta espècie, la qual cosa obria la possibilitat que la regionalització del eix anteroposterior durant el desenvolupament d'*O. dioica* es fes en absència d'RA, malgrat que com a urocordat conservava el pla corporal arquetípic del fílum (Cañestro and Postlethwait, 2007; Cañestro et al., 2006). Aquest descobriment va portar a formulació de la paradoxa 'inversa' de l'Evo-Devo, és dir, que organismes aparentment molt similars –unitat fenotípica– podien no compartir el *genetic toolkit* –diversitat genotípica– per produir aquesta unitat (Cañestro and Postlethwait, 2007). En el cas d'*O. dioica*, la diversitat genotípica era conseqüència de la pèrdua de gens i per tant aquest cas obria la porta a estudiar com les pèrdues gèniques podien haver impactat l'evolució dels mecanismes del desenvolupament d'un cordat. Per això, però, era necessari determinar l'abast del desballestament de la RA-MGN revelant patrons de co-eliminació, i demostrar que la RA-MGN no era un sistema mutacionalment robust, és a dir, que no havia gens o vies alternatives capaços de portar a terme la mateixa funció.

La via Wnt

Gran quantitat d'estudis vinculen la funció de l'RA durant el desenvolupament en vertebrats amb l'acció d'altres vies de senyalització tals com la via Wnt, Fgf, Nodal, Bmp o Shh (Sonic hedgehog) (Engberg et al.,

2010; Liu et al., 2012; Metallo et al., 2008; Stavridis et al., 2010). Aquests conjunts de vies són essencials, per exemple, pel desenvolupament del cor, l'ull i les extremitats, per l'establiment del patró en el sistema nerviós central (CNS), en la diferenciació dels arcs faringis i en la regulació de la somitogènises, on l'efecte antagònic entre l'RA i les vies de senyalització Wnt i Ffg controla el desenvolupament de la part posterior del cos dels vertebrats Aquesta estratègia basada en gradients d'RA oposats als de Fgf i Wnt, no només la usen els vertebrats, sinó que els ascidis també la utilitzen encara que per establir el patró anteroposterior del sistema nerviós perifèric de la cua (Pasini et al., 2012). És possible argumentar, per tant, que el desballestament de la via de l'RA pot haver tingut un impacte important en el funcionament d'aquestes altres vies i, per tant, en l'evolució dels mecanismes del desenvolupament dels cordats. D'entre totes aquestes vies la dels Wnt és especialment interessant perquè s'ha proposat com a possible senyal principal en l'elongació posterior dels embrions en l'ancestre comú de tots els cordats (Bertrand et al., 2015).

La via de senyalització dels Wnt

La via de Wnt és un sistema de comunicació extracel·lular de curt abast que a través de diversos mecanismes de transducció de senyal regula la transcripció gènica de les cèl·lules diana sobre les que actua. Tots els organismes tenen múltiples gens Wnt (p.e. 19 gens en humans) que codifiquen proteïnes d'entre 350 i 400 aa, les quals són secretades després de ser glicosilades i acilades. Quan una cèl·lula secreta un Wnt, aquest actua com a lligand del receptor Frizzled i activa l'efector citoplasmàtic Dishevelled. Juntament amb aquests tres components, moltes altres proteïnes estan implicades en la senyalització per Wnt, ja sigui per regular la interacció Wnt-Frizzled o per transferir el senyal al nucli. Les diverses vies que transfereixen el senyal Wnt al nucli s'han classificat, històricament, en la via "canònica" i via "no canònica" (Croce and McClay, 2009) (fig. 14), i malgrat que l'ús que en fa un determinat Wnt d'aquetes vies proporciona una certa especificitat a la seva funció, les diferents vies no són mútuament excloents ja que si bé un Wnt concret szovint actua utilitzant una de les vies, depenent de les circumstàncies, pot fer servir també l'altra (Mikels and Nusse, 2006) (revisat en (Kestler and Kuhl, 2008)). La via canònica, també anomenada via de destí cel·lular (*cell fate*), depèn de l'estabilització i el transport de β-catenina al nucli on s'uneix a factors de transcripció (p.e. Tcf/Lef) que regulen l'expressió de gens diana involucrats majoritàriament en l'especificació i/o manteniment del destí cel·lular (fig. 14A). La via no canònica, nomenada via de polaritat cel·lular (*cell polarity*), actua independentment de la

β-catenina i és considerablement més diversa en referència als efectors intracel·lulars –p.e. JNK (*c-Jun N -terminal kinases*) o Ca^{2+}– que utilitza per regular la dinàmica del citoesquelet i l'adhesió cel·lular, i controlar la polarització espacial de la cèl·lula i l'alineació de la divisió, la migració o l'elongació cel·lular al llarg d'un mateix eix (Loh et al., 2016) (fig. 14B).

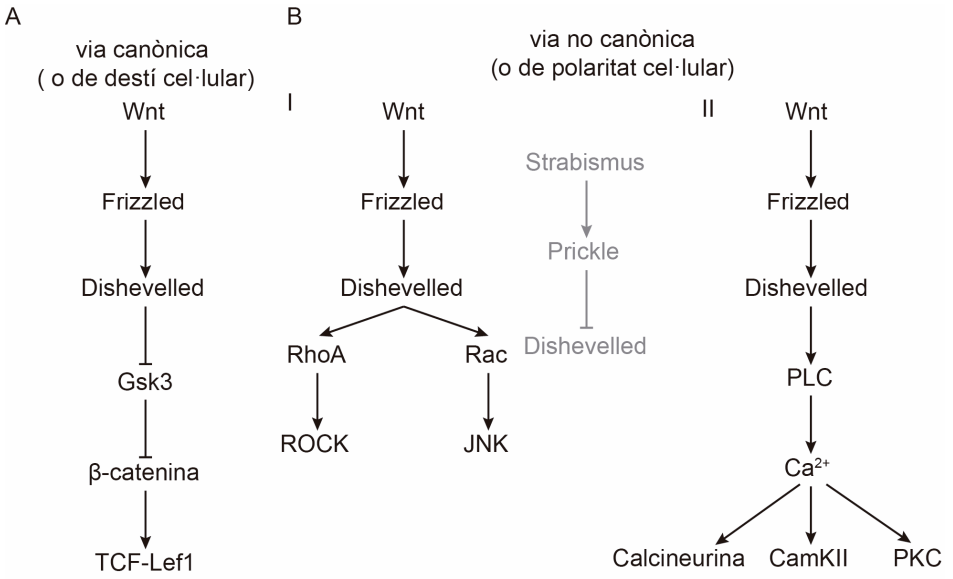


Fig. 14. **Representació esquemàtica de les diferents vies Wnt**. (A) Via canònica o de destí cel·lular. Aquesta via s'activa per una interacció entre el lligant Wnt i el receptor Frizzled, que resulta en l'activació de l'efector citoplasmàtic Dishevelled que al seu torn inhibeix l'antagonista Gsk3, permetent a la β-catenina activar TCF/Lef1. (B) Via no canònica o de polaritat cel·lular. Es mostren dues de les vies Wnt no canòniques més estudiades. (I) En la primera, després de la interacció del lligand Wnt amb un receptor Frizzled i l'activació de la proteïna citoplasmàtica Dishevelled, el senyal es transdueix a través de dues proteïnes G diferents, Rho i Rac, que transmeten polaritat cel·lular a través dels seus respectius efectors ROCK i JNK. Per establir la polaritat, però, el receptor de membrana Strabismus també s'activa, encara que en la cèl·lula veïna (en gris), on a través de Prickle inhibeix la via de polaritat cel·lular segrestant Dishevelled. (II) En la segona via no canònica l'activació de la via Wnt condueix a l'activació del component PLC i un augment en els nivells de calci intracel·lular, que indueixen l'activació de les proteïnes Calcineurina, CamKII i PKC. Modificat de Croce and McClay, 2009.

L'evolució dels Wnt

Tots els organismes tenen múltiples gens Wnt que s'han classificat en 13 subfamílies: des de la subfamília Wnt1 a la Wnt11, a més les Wnt16 i WntA (Albalat and Cañestro, 2016; Croce and Holstein, 2014; Holstein, 2012). Anàlisis filogenètiques i genòmiques a gran escala dels Wnt d'espècies d'animals distribuïdes al llarg de l'escala evolutiva han portat a concloure que les 13 subfamílies són molt antigues ja que l'ancestre comú dels eumetazous (cnidaris + bilaterals) ja tenia membres de totes elles (Kusserow et al., 2005;

Pang et al., 2010; Ukolova, 2012). Tanmateix, els membres de les diverses subfamílies han estat retinguts, perduts o duplicats de forma desigual durant l'evolució dels diferents llinatges. Dins del clade dels protostomats, per exemple, la pèrdua de subfamílies Wnt ha estat freqüent, ja fos de forma ancestral i afectant a tot el clade (p.e. la pèrdua del Wnt3 en tots els fílum), afectant a llinatges concrets (p.e. la pèrdua de Wnt16 en els insectes o la pèrdua de Wnt8 en els mol·luscs) o de forma recurrent en els diferents llinatges (p.e. les almenys 6 pèrdues independents de Wnt9 en artròpodes, anèl·lids, mol·luscs i platihelmints) (Albalat and Cañestro, 2016; Riddiford and Olson, 2011) (fig.15). Per contra, els vertebrats semblen haver estat refractaris a perdre Wnt ja que conserven 12 de les 13 subfamílies, amb l'única excepció de la

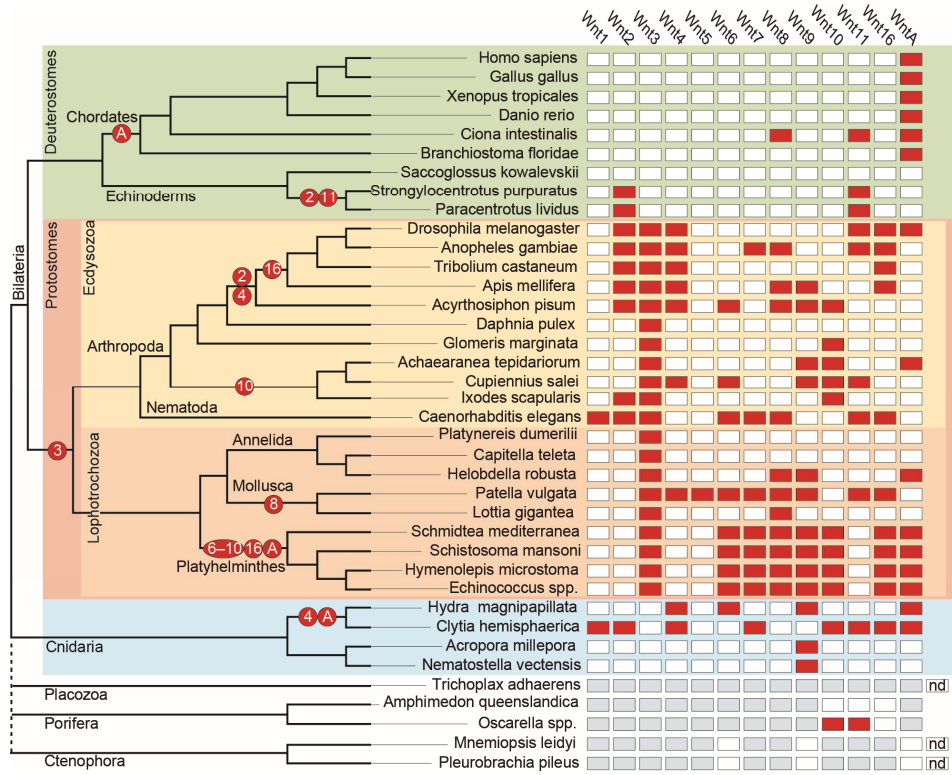


Fig. 15 **Distribució dels genes Wnt en els metazous.** L'extensa pèrdua de gens (quadres vermells) ha afectat a totes les subfamílies de gens Wnt (1 a 11; 16 i A) en tots els taxons dels metazous. Algunes pèrdues gèniques semblen ancestrals (cercles vermells) i, per tant, probablement rellevants per a l'evolució de grups sencers (per exemple, la pèrdua ancestral de Wnt3 en els protostomats de la tija). Altres pèrdues gèniques semblen aparèixer recurrentment en diversos llinatges i mostren una distribució irregular (per exemple, la pèrdua de Wnt11 en alguns cordats, equinoderms, artròpodes, nematodes, mol·luscs i esponges). Les relacions filogenètiques no concloents (branques arborescents) o incerteses en les ortologies de gens (nd) dificulten la capacitat de determinar si l'absència de famílies Wnt en la majoria de metazous basals (caixes grises) es deu a pèrdues gèniques o a guanys genètics. Extret de Albalat and Cañestro, 2016.

subfamília WntA. Si aquesta tendència és específica dels vertebrats o compartida pels altres grups de cordats no se sap ja que el repertori de Wnt en urocordats (tant a ascidis com a *O. dioica*) no està resolt i les dades a cefalocordats són només parcials.

La via de Wnt en la determinació dels patrons axials a l'embrió

La senyalització de Wnt juga a través de la via canònica de la β-catenina un paper crucial en l'establiment de la polaritat axial durant la formació del pla corporal de la majoria dels animals (revisat en Petersen and Reddien, 2009). L'eix primari animal-vegetal (AV) és el primer eix que s'especifica poc després de la fertilització de l'ou i, tot i que els mecanismes de polaritat axial difereixen dràsticament entre els animals degut a les diferents estratègies de desenvolupament (p.e. desenvolupament extern *versus* intern, desenvolupament mosaic *versus* regulat), la via de Wnt contribueix a posicionar les tres capes germinals al llarg d'aquest eix primari. En embrions de diversos metazous, asimetries en la distribució del components materns de la via canònica de Wnt causen l'estabilització de β-catenina en un pol de l'embrió contribuint a determinar la polaritat axial primaria (AV animal-vegetal i DV dorsal-ventral) i la especificació endoderm-mesoderm (Hikasa and Sokol, 2013).

Aquesta distribució asimètrica dels components materns de la via de Wnt/β-catenina va ser investigada als anys 90 en diverses espècies deuterostomats, incloent l'eriçó de mar (Wikramanayake et al., 1998), els ascidis (Sasakura et al., 1998), i els vertebrats *Xenopus* i peix zebra. A *Xenopus*, per exemple, la localització del mRNA matern de Wnt11 en el còrtex vegetal de l'ou (Tao et al., 2005), juntament amb els mRNA materns de Wnt5a (malgrat no trobar-se asimètricament localitzat) i de Dkk1 (un antagonista de Wnt), són els principals determinants de la polaritat axial DV (Cha et al., 2008). A peix zebra, Wnt5 i Wnt11 són també materns (Kilian et al., 2003; Makita et al., 1998) i la pèrdua de Wnt5 impedeix la formació correcte de l'eix DV (Westfall et al., 2003), metre que la pèrdua de la funció de Wnt11 no es essencial pel desenvolupament de l'eix (Heisenberg et al., 2000).

A més de la funció primerenca dels Wnt materns, l'expressió zigòtica dels Wnt també és important per a l'especificació de l'eix AP al llarg de l'embriogènesi. Aquesta funció de Wnt en el patró AP es va revelar per primera vegada a *Xenopus* i a peix zebra, mitjançant estudis de sobre-expressió o d'aboliment de diferents components de la via (revisats en (Hikasa and Sokol, 2013)). Aquests resultats juntament amb estudis amb altres animals de diferents fila (p e. a platihelmints (Petersen and Reddien, 2008), artròpodes (McGregor et al., 2008), nemàtodes (Rocheleau et al., 1997) i equinoderms (Wikramanayake

et al., 2004), van establir les bases del paper clau i evolutivament conservat que te la via de Wnt en el control de la identitat posterior durant la formació del pla corporal de moltes espècies (Hikasa and Sokol, 2013; Petersen and Reddien, 2009).

Objectius

Objectius

Com s'ha explicat en la introducció, la pèrdua gènica és una força evolutiva capaç de generar biodiversitat. En aquesta Tesi Doctoral hem volgut estudiar l'impacte de pèrdua gènica en l'evolució dels mecanismes del desenvolupament dels organismes. Per fer-ho, hem triat l'urocordat *Oikopleura dioica* com a model animal, ja que la pèrdua de gens és una de les seves principals característiques. Hem decidit estudiar les pèrdues gèniques en dues vies de senyalització cabdals pel desenvolupament embrionari: la via de senyalització de l'àcid retinoic, una important via morfogènica per determinar el patró axial en els cordats, i la via de Wnt, un sistema de senyalització crucial per la formació del pla corporal en la majoria dels animals. La interacció de les dues vies en l'establiment de l'eix anteroposterior dels cordats les fa, a més, especialment interessants. Dins d'aquest context, ens hem marcat quatre objectius específics:

- **Objectiu 1**: Establir un sistema de cultiu d'*O. dioica* eficient que permeti disposar del material biològic necessari pels experiments.
- **Objectiu 2**: Caracteritzar l'abast del desballestament de la via de l'àcid retinoic a *O. dioca*.
- **Objectiu 3**: Comparar les vies de Wnt en diferents espècies de cordats, especialment del subfílum dels urocordats, i avaluar les seves històries evolutives.
- **Objectiu 4**: Caracteritzar la via de Wnt a *O. dioica* per visualitzar processos de pèrdues gèniques i entendre l'evolució funcional de la via en aquesta espècie.

Informe dels directors de Tesis

Informe dels directors de tesi

El doctorand Josep Marti-Solans presenta la memòria de la Tesi Doctoral titulada "***Oikopleura dioica* com a model animal per investigar l'impacte de les pèrdues gèniques en l'Evo-Devo: les vies de senyalització de l'àcid retinoic i Wnt com a cas d'estudi**". Els seus co-directors, el **Dr. Ricard Albalat** i el **Dr. Cristian Cañestro**, informem que la tesi doctoral està composta per 6 articles. L'apartat de **Resultats** conté 4 articles, en tres dels quals el doctorand és primer autor, i l'altre és segon autor. D'aquest 4 articles, 2 ja estan publicats, 1 està en procés d'acceptació de la revisió, i l'altre està per ser enviat en les properes dates. L'apartat de l'apèndix conté els altres 2 articles, on el doctorand es segon autor 1 dels quals està publicat i l'altre està en procés de revisió. Fem constar que tots els articles publicats ho han estat en revistes internacionals de prestigi en les seves àrees de coneixement que inclouen la Genètica i Genòmica Evolutiva, la Biologia del Desenvolupament, i la Biologia en general. Totes les revistes on el doctorand ha publicat els seus treballs es troben incloses a Scopus, i en tots els casos es tracta de publicacions que han passat el filtre de avaluadors anònims designats pels editors de les revistes. A continuació, es detalla la contribució científica del doctorand en cada un dels articles, així com el factor d'impacte "*CiteScore*" de les revistes en les que han sigut publicats.

Apartat resultats:

Capítol 1

Oikopleura dioica* culturing made easy: A low-cost facility for an emerging animal model in EvoDevo.** Josep Martí-Solans, Alfonso Ferrández-Roldán, Hector Godoy-Marín, Jordi Badia-Ramentol, Nuria Paz Torres-Águila, Adriana Rodríguez-Marí, Jean Marie Bouquet, Daniel Chourrout, Eric Thompson, Ricard Albalat, Cristian Cañestro. ***Genesis (2015) 53:183-193

- Factor d'Impacte: 2,41
- Posició en l'àrea: 144/311, tercer quartil -Q3- dins de l'àrea *Biochemistry, Genetics and Molecular Biology (Genetics)*

El doctorand Josep Marti-Solans va ser el protagonista principal del disseny i postada en marxa del protocol per establir *O. dioica* al nostre laboratori. Això va incloure mostrejar diversos indrets de la geografia catalana per trobar animals de

l'espècie *O. dioica*. Establir en el laboratori cultius de microalgues per alimentar els animals al laboratori. Determinar tant el règim alimentari com el de renovació d'aigua, determinar la quantitat d'animals a mantenir per tanc i establir un protocol per a mantenir les microalgues. El doctorand, a més, va ser el responsable de dissenyar i realitzar els creuament necessaris per a caracteritzar les característiques biològiques (mida, fecunditat, fertilitat i desenvolupament embrionari) tant població salvatge com de les poblacions amb diferent grau de consanguinitat. Això inclou realitzar tots els creuaments necessaris per determinar la mida de la posta i els percentatges de de fertilitat i de desenvolupament embrionari. Finalment, el doctorand va participar activament en la discussió de resultats i en la redacció del manuscrit.

Capítol 2

Co-elimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. Josep Martí-Solans, Olga V. Belyaeva, Nuria P. Torres-Aguila, Natalia Y. Kedishvili, Ricard Albalat and Cristian Cañestro. ***Mol. Biol. Evol.*** (2016) 33: 2401-2416

- Factor d'Impacte: 7,65
- Posició en l'àrea: 15/311, primer quartil -Q1- dins de l'àrea *Biochemistry, Genetics and Molecular Biology (Genetics)*

El doctorand va ser el principal autor d'aquest treball, sent responsable de la identificació i anàlisi filogenètic dels components de la xarxa gènica del metabolisme de l'RA (RA-MGN), així com llur clonatge i anàlisi d'expressió (tant en condicions normals, com en presencia de toxines). A més, el doctorand va sintetitzar les construccions necessàries per avaluar l'activitat enzimàtica dels enzims Cco i va posar a punt el protocol per a l'expressió heteròloga d'aquestes. El doctorand, també, va recol·lectar el material necessari per determinar la presència de retinoids i carotenoids a *O. dioica*. Finalment, el doctorand va tenir un paper principal en la interpretació dels resultats, la seva discussió , i en la generació de les figures i redacció del manuscrit.

Capítol 3

Wnt evolution and function shuffling in chordate genomes: from conservative amphioxus stasis to liberal ascidian gains and losses. Ildiko M.L. Somorjai, Josep Martí-Solans, Miriam Diaz-Gracia, Hiroki Nishida, Kaoru S.

Imai, Hector Escrivà, Cristian Cañestro and Ricard Albalat. ***Genome Biol.*** (en fase d'acceptació de la segona revisió)

- Factor d'Impacte: 12,66
- Posició en l'àrea: 4/311, primer quartil -Q1- dins de l'àrea *Biochemistry, Genetics and Molecular Biology (Genetics)*

Aquest treball es desenvolupa en col·laboració amb el grup de la Dra. Somorjai (St. Andrews University), qui realitza l'estudi d'expressió a cefalocordats, mentre que el doctorand va col·laborar amb la recol·lecció de totes les seqüències Wnt d'ascidis, la elaboració de l'anàlisi filogenètic, de posicions d'introns i de sintènia. A més, el doctorand va realitzar l'anàlisi d'expressió de WntA a *Halocynthia roretzi*. El doctorand va participar en la finalització de les figures, taules i text del manuscrit.

Capítol 4

***Oikopleura dioica* Wnt signaling: when gene loss, duplication and function shuffling push the limits of chordate EvoDevo**. Josep Martí-Solans, Hector Godoy, Miriam Diaz-Gracia, Takeshi Onuma, Hiroki Nishida, Ricard Albalat and Cristian Cañestro (Manuscrit en preparació).

El doctorand és el principal autor d'aquest treball, responsable de l'anàlisi filogenètic dels lligands, receptors i antagonistes de la via de senyalització per Wnt, així com de l'anàlisi complert d'expressió dels components de la via Wnt a *O. dioica*. A més, el doctorand va dur a terme l'anàlisi funcional de Wnt11a i la descripció del fenotip. Finalment, el doctorand ha estat el principal responsable de redactar el manuscrit, i generar les seves figures.

Apartat Apèndix

Apèndix 1

DNA methylation in amphioxus: from ancestral functions to new role in vertebrates. Ricard Albalat, Josep Martí-Solans, and Cristian Cañestro. Brief. Funct. Genomics. (2012) 11: 142-155

- Factor d'Impacte: 3,17
- Posició en l'àrea: 141/367, segon quartil -Q2- dins de l'àrea *Biochemistry, Genetics and Molecular Biology (Molecular Biology)*

El doctorand va participar en l'anàlisi *in silico* del patró de metilació del DNA en el genoma de l'amfiox.

Apèndix 2

Ecological impact and developmental genetic responses of appendicularians to diatom-bloom derived biotoxins. Nuria Paz Torres-Águila, Josep Martí-Solans, Alfonso Ferrández-Roldán, Alba Almazán-Almazán, Vittoria Roncalli, Salvatore D'Aniello, Giovanna Romano, Anna Palumbo A, Ricard Albalat and Cristian Cañestro. *Commun. Biol.* (en fase de revisió menor)

Factor d'Impacte: No disponible encara

Posició en l'àrea: No disponible encara

El doctorand va participar en l'elaboració dels extractes de microalgues durant una estada a Nàpols, així com va col·laborar en el clonatge i anàlisi d'expressió dels gens del defensoma.

Firmat,

Dr. Ricard Albalat Dr. Cristian Cañestro

Resultats. Capítol 1

***Oikopleura dioica* culturing made easy: A low-cost facility for an emerging animal model in EvoDevo.**

Josep Martí-Solans,[1] Alfonso Ferrández-Roldán,[1] Hector Godoy-Marín,[1] Jordi Badia-Ramentol,[1] Nuria Paz Torres-Águila,[1] Adriana Rodríguez-Marí,[1] Jean Marie Bouquet,[2,3] Daniel Chourrout,[2] Eric Thompson,[2,3] Ricard Albalat,[1] Cristian Cañestro[1]

[1]Departament de Gene_tica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, 08028, Spain

[2]Sars International Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Bergen, Norway

[3]Department of Biology, University of Bergen, Postbox 7803, N-5020 Bergen, Norway

Resum

La seqüenciació del genoma i el desenvolupament de tecnologies de *knockdown* mitjançant RNAi a l'urocordat *Oikopleura dioica* fan d'aquest organisme un model emergent i atractiu en el camp de l'EvoDevo. Per tenir èxit com a nou model animal, però, un organisme ha poder mantenir-se de manera fàcil i accessible al laboratori. Avui en dia, només hi ha dues instal·lacions al món capaces de mantenir indefinidament l'*Oikopleura dioica*, una a l'institut SARS (Bergen, Noruega) i l'altra a la Universitat d'Osaka (Japó). Aquí, es descriu l'establiment d'una nova instal·lació a la Universitat de Barcelona en la qual hem modificat els protocols de cultiu prèviament publicats per optimitzar la producció setmanal de milers d'embrions i centenars d'animals madurs utilitzant la mínima quantitat d'espai, recursos humans i equipament. Aquesta optimització inclou nous protocols per a manteniment a llarg termini, mitjança criopreservació i cultius sòlid, estocs de microalgues -*Chaetoceros calcitrans*, *Isochrysis sp.*, *Rhinomonas reticulate* i *Synechococcus sp.*- necessaris per a l'alimentació *Oikopleura dioica*. El nostre sistema de cultiu manté línies parcialment consanguínies amb característiques similars als animals salvatges i és fàcilment ampliable per satisfer les necessitats de qualsevol laboratori que desitgi utilitzar *Oikopleura dioica* com a organisme model.

TECHNOLOGY REPORT

Oikopleura dioica Culturing Made Easy: A Low-Cost Facility for an Emerging Animal Model in EvoDevo

Josep Martí-Solans,[1] Alfonso Ferrández-Roldán,[1] Hector Godoy-Marín,[1] Jordi Badia-Ramentol,[1] Nuria P. Torres-Aguila,[1] Adriana Rodríguez-Marí,[1] Jean Marie Bouquet,[2,3] Daniel Chourrout,[2] Eric M. Thompson,[2,3] Ricard Albalat,[1]* and Cristian Cañestro[1]*

[1]Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, 08028, Spain

[2]Sars International Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Bergen, Norway

[3]Department of Biology, University of Bergen, Postbox 7803, N-5020 Bergen, Norway

Received 3 June 2014; Revised 4 July 2014; Accepted 7 July 2014

Summary: The genome sequencing and the development of RNAi knockdown technologies in the urochordate *Oikopleura dioica* are making this organism an attractive emergent model in the field of EvoDevo. To succeed as a new animal model, however, an organism needs to be easily and affordably cultured in the laboratory. Nowadays, there are only two facilities in the world capable to indefinitely maintain *Oikopleura dioica*, one in the SARS institute (Bergen, Norway) and the other in the Osaka University (Japan). Here, we describe the setup of a new facility in the University of Barcelona (Spain) in which we have modified previously published husbandry protocols to optimize the weekly production of thousands of embryos and hundreds of mature animals using the minimum amount of space, human resources, and technical equipment. This optimization includes novel protocols of cryopreservation and solid cultures for long-term maintenance of microalgal stocks—*Chaetoceros calcitrans*, *Isochrysis sp.*, *Rhinomonas reticulate*, and *Synechococcus sp.*—needed for *Oikopleura dioica* feeding. Our culture system maintains partially inbred lines healthy with similar characteristics to wild animals, and it is easily expandable to satisfy on demand the needs of any laboratory that may wish to use *Oikopleura dioica* as a model organism. genesis 53:183–193, 2015. © 2014 Wiley Periodicals, Inc.

Key words: tunicate; Appendicularian (larvacean) husbandry; microalga cryopreservation; fecundity and fertilization success; chordate developmental biology

In the last decades, the study of well-established animal models has been crucial for the extraordinary progress of knowledge in the fields of developmental biology and biomedicine, but the scientific challenges of the XXI century, especially those to decipher the evolutionary basis of life diversity, demand the development of new animal models carefully chosen according to their phylogenetic position in the tree of life (Maher, 2009). The larvacean (or appendicularian) *Oikopleura dioica*, for instance, is a planktonic marine organism that

Additional Supporting Information may be found in the online version of this article.

Alfonso Ferrández-Roldán, Hector Godoy-Marín, Jordi Badia-Ramentol, and Nuria P. Torres-Águila contributed equally to this work.

Current Address for Adriana Rodríguez-Marí: Institut de Bioenginyeria de Catalunya, IBEC, Barcelona, 08028, Spain.

*Correspondence to: Ricard Albalat, Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, 08028, Spain. E-mail: ralbalat@ub.edu (or) Cristian Cañestro, Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, 08028, Spain. E-mail: canestro@ub.edu

Contract grant sponsor: Ministerio de Ciencia e Innovación; Contract grant number: BFU2010-14875; Contract grant sponsor: Ministerio de Economía y Competitividad (Spain); Contract grant number: BFU2013-48507-P; Contract grant sponsor: Generalitat de Catalunya; Contract grant number: 2009SGR336, SGR2014-290; Contract grant sponsor: ASSEMBLE, Contract grant number: 1553-SZN7; Contract grant sponsor: NSF, Contract grant number: IBN-0345203; Contract grant sponsor: Heiwa-Nakajima Foundation research/travel grant.

Published online 12 July 2014 in
Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/dvg.22800

possesses many characteristics that make it attractive as animal model: (*i*) *O. dioica* occupies a key phylogenetic position within the chordate phylum as a basally divergent member of the urochordate subphylum (Stach and Turbeville, 2002; Swalla *et al*., 2000; Wada, 1998), which is the sister group of vertebrates (Delsuc *et al*., 2006; Delsuc *et al*., 2008). *O. dioica*, therefore, is a useful model to infer the ancestral condition from which vertebrate and other urochordate species evolved. (*ii*) The genome of *O. dioica* is very small (~70 Mb) and appears to have suffered an extreme process of compaction leading to very small introns (peak at 47 bp) and intergenic regions (53% < 1 kb), and absence of most pan-animal transposable elements (Cañestro and Albalat, 2012; Chavali *et al*., 2011; Denoeud *et al*., 2010). This drastic genome compaction has been accompanied by an extreme shattering of exon–intron organization and physical gene reordering, and massive gene losses accompanied by extensive lineage specific gene duplications. Despite this deep reorganization of the genome, *O. dioica* conserves a typical chordate body plan, which makes this organism an interesting model of extreme genome plasticity and body plan conservation, *a.k.a*. the "inverse paradox of Evo-Devo" (Cañestro *et al*., 2007; Denoeud *et al*., 2010). (*iii*) *O. dioica* has a short generation time of just 5 days at 20°C (Fenaux, 1998) (comparable to *C. elegans*) and a high fecundity with hundreds of eggs per female that can be externally fertilized in vitro. *O. dioica* provides an excellent animal model for studies in developmental biology, as fertilized eggs develop into transparent embryos easy to follow under a stereomicroscope without any further manipulation. Many developmental studies have already described its early embryogenesis (Delsman, 1910; Nishida, 2008), its stereotyped cell fate map (Fujii *et al*., 2008; Stach *et al*., 2008), as well as the development of several body structures such as the notochord (Bassham and Postlethwait, 2000; Nishino *et al*., 2001), the central nervous system (Cañestro *et al*., 2005; Cañestro and Postlethwait, 2007; Soviknes *et al*., 2007), placodal sensory organs (Bassham and Postlethwait, 2005), the digestive system (Burighel and Brena, 2001; Cañestro *et al*., 2010), gonad development (Ganot *et al*., 2007; Ganot *et al*., 2008), or the homolog of the thyroid (Bassham *et al*., 2008; Cañestro *et al*., 2008). *iv*). Finally, recent technical advances in RNAi gene silencing by injecting dsRNA into the female gonad has definitively opened the possibility to make this animal a model for functional studies gaining new insights into genes of our interest (Omotezako *et al*., 2013).

Despite *O. dioica* is cosmopolitan and widely available around the globe, the design of an easy and affordable system to culture *O. dioica* in any laboratory interested in studying this organism remains a challenge. Breeding of *O. dioica* in the laboratory was already successfully achieved 40 years ago (Paffenhöfer, 1973) and since then, several laboratories have tried protocols for its husbandry during a limited number of generations by mixing mature male and female specimens and let them spontaneously spawn in beakers in which animals were maintained in suspension by different devices designed *ad hoc* (Bassham and Postlethwait, 2000; Chioda *et al*., 2002; Fenaux and Gorsky, 1979, 1985; Lopez-Urrutia and Acuña, 1999; Nishino *et al*., 2000; Sato *et al*., 1999; Troedsson et al., 2013). Currently, however, to our knowledge there are only two active laboratories with animal facilities capable to indefinitely maintain *O. dioica* cultures all year around, one in the SARS Institute in Bergen (Norway) (Bouquet *et al*., 2009) and another in the Osaka University (Japan) (Nishida, 2008). These animal facilities have the capability to reliably produce large amounts of adult animals and embryos every week, but they require also large amounts of laboratory space, seawater, microalgal cultures, equipment, and human resources, which make their setups unaffordable for most standard research laboratories. With the aim of simplifying and scaling down the *O. dioica* husbandry system, we have modified and optimized the conditions described for the existing facilities. Despite the animal production is lower in our system, our husbandry protocol is flexible enough to provide the biological material to satisfy the requirements for most standard research groups using the minimum amounts of space, technical equipment, laboratory expenses, and human resources.

METHODS AND RESULTS

Animal Collection, Facility, and Seawater

Adult animals were collected in three different locations of the Catalonian coast near Barcelona (Fig. 1a; see Supporting Information Methods for details) and raised in our animal facility, which was inspirited on the *Oikopleura* culture facilities described in Nishida 2008 and Bouquet *et al*. 2009, but modified to satisfy our experimental requirements. Our animal facility was set in a small room (about 5 m^2) maintained at 19 ± 1°C using a standard air conditioner device. Animals were maintained in seawater filtered through 50 μm and 20 μm polypropylene filters (MBHP050-10 and LTPPBB020-10 AscoFiltri Filtering cartridges, respectively) (Fig. 1b; see Supporting Information Methods for details) and feed with microalgae. For in vitro fertilization experiments and for microalgal cultures, filtered seawater (fSW) was sterilized (sterilized seawater, sSW) using 0.22 μm filters (VacuCap PF Filters 4622, Pall Corporation) (Albalat et al., 2003; Dalfó *et al*., 2001).

Microalgal Production for *O. dioca* Feeding

The diet for *O. dioica* consisted of four microalga species, three true eukaryotic species of microalga, *Chaetoceros calcitrans* (CCAP 1010/11), *Isochrysis sp*. (CCAP

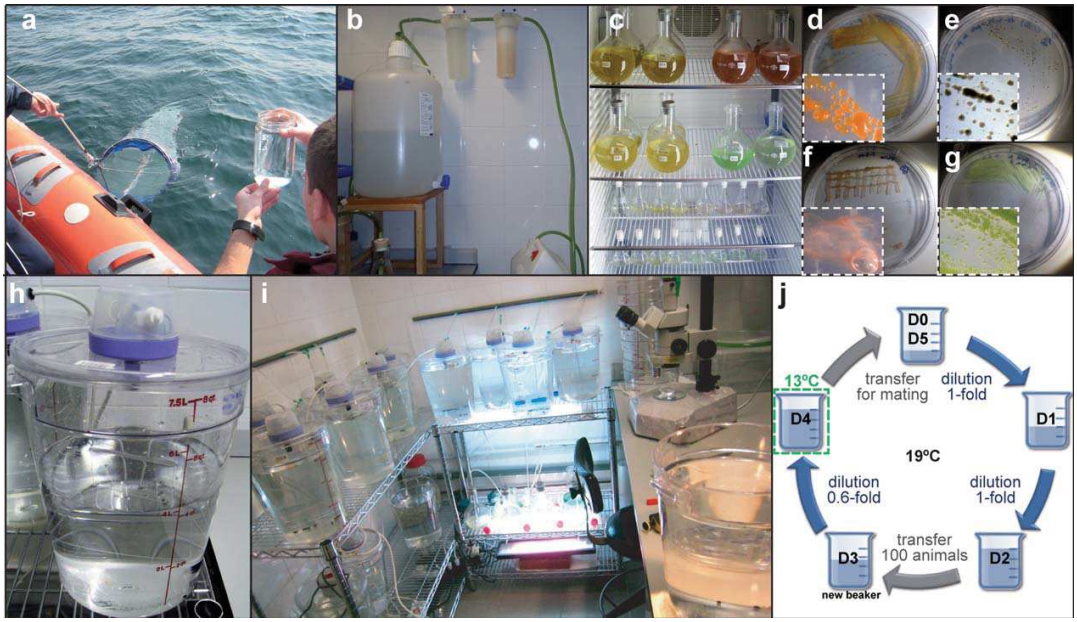


FIG. 1. *O. dioica* facility in the University of Barcelona (Catalonia, Spain). (**a**) Animals were collected in the coast of Catalonia near Barcelona using a plankton net or directly with a bucket. (**b**) Seawater is filtrated at 50–20 μm (fSW) in the facility to remove excess of sand particles that could affect *O. dioica* buoyancy. (**c**) The production of the four microalgae for *O. dioica* feeding (Bouquet *et al*., 2009) was scale down in an adaptable fashion to the weekly needs of the facility (round-bottom glass flasks in upper shelves). Long-term stocks (100 mL Erlenmeyers in lower shelves) were renewed just once per month. (**d-g**) The use of agar plates provides an alternative method to maintain long-term microalgal stocks: (**d**) *Isochrysis sp*., (**e**) *Chaetoceros calcitrans*, (**f**) *Rhinomonas reticulata* and (**g**) *Synechococcus sp*. Solid cultures also allow recovering isolated colonies (insets **d-g**), which can be useful in case of contamination. *R. reticulata* grow better within the agar matrix rather on the surface (notice the bubble in inset **f**), and despite colonies are not easily formed, individual single cells are easily observed. (**h**) *O. dioica* animals were maintained in suspension by the rotation of a paddle driven by a motor mounted on the lid of polycarbonate beakers. (**i**) Animal lines were maintained in a small room (5 m^2) in four shelves (1.5 m^2) at 19°C using a standard air conditioner device. (**j**) The protocol of husbandry has been designed to make a life cycle of five days at 19°C in which each generation start with a mating of 10–15 males and 20–30 females close to spawn (D0). The next two days (D1 and D2) cultures are just diluted onefold (1/2 dilution) with fresh fSW, and at the third day (D3), 100 animals are transferred to a new beaker with fresh fSW. At day 4 (D4), cultures are 0.6-fold diluted (2/3 dilution) with fresh fSW and moved to a 13°C incubator. In the morning of day 5 (D5), most animals synchronously become mature, and are ready to set a new mating for next generation. This protocol of husbandry has been optimized to maximize the reduction of the amount of seawater, number of beakers, volume of microalgae feeding cultures, and tasks of manipulation. This protocol of husbandry is easily expandable on demand (i.e., by increasing the number of matings on D0 or the number of transfers on D3), which allows us to easily generate hundreds of mature animals and tens of thousands of embryos every week.

927/14), *Rhinomonas reticulata* (CCAP 995/2) and one prokaryotic species of cyanobacteria (a.k.a. blue-green algae) *Synechococcus sp*. (K0408), grown in modified Conway medium (Bouquet *et al*., 2009) until their optical density (OD) at 600 nm ranged from 0.20 to 0.25 (8×10^6 to 1×10^7 cells/ml) for *C. calcitrans*, from 0.13 to 0.18 (2×10^6 to 3×10^6 cells/ml) for *Isochrysis sp*., from 0.14 to 0.20 (1×10^6 to 1.5×10^6 cells/ml) for *R. reticulata* and from 0.15 to 0.20 (4.5×10^7 to 6×10^7 cells/ml) for *Synechococcus sp*. (Fig. 1c; see Supporting Information Methods for details).

Long-Term Maintenance of Microalgal Strain Stocks

The maintenance of healthy stocks of the four microalgal strains is crucial for the culturing system of *O. dioica*. This maintenance, however, is a time-consuming task that requires weekly renewal and is sensitive to contamination. To optimize this task, we designed three strategies that minimized the manipulation of the stocks and reduced the risk of contamination: (*i*) long-term liquid cultures; (*ii*) long-term solid cultures; and (*iii*) cryopreserved microalgae stocks.

Long-term liquid cultures. Long-term cultures were generated by finding the inoculum with minimum number of cells that guaranteed the recovery of the culture but delayed the reach of the declining phase as much as possible. We empirically determined that inocula containing OD-estimates of $\approx 10^2$ cells for *R. reticulata*, $\approx 10^3$ cells for *C. calcitrans* and *Isochrysis sp*., and $\approx 10^4$ cells for *Synechococcus sp*. in 50 mL of modified Conway medium were enough to guarantee the recovery

of the cultures and reduced the renewal to just once per month, when their OD reached 0.35–0.60 (Fig. 1c and for details see Supporting Information Fig. 1).

Long-term solid cultures. The second strategy consisted in culturing the microalgae in agar plates that could be maintained in good conditions for long periods, at the same time that allowed the isolation of individual colonies in case of accidental contamination. Agar (Pronadisa #1800.00) was dissolved in modified Conway medium by autoclaving, and vitamins were added to cooled media before agar solidified. The streaks were made from the pellet from 5 mL of a microalgal liquid culture (OD > 0.25) centrifuged for 5 min at 756 g (2,500 rpm). Plates were incubated upside-down at 13°C with 12 h photoperiod until microalgal colonies became apparent after 1–2 weeks (Fig. 1d–g). Agar concentration appeared to be a critically variable for the growth of each microalga, and we found that *Isochrysis sp*. grew well in concentrations from 1.5% to 0.5%, *C. calcitrans* from 1% to 0.5%, whereas *Synechococcus sp* and *R. reticulata* grew better at 0.5% agar and 0.3%, respectively. We have used microalgal colonies plated for at least 3 months to successfully inoculate new plates or liquid medium.

Microalgae cryopreservation. With the exception of *C. calcitrans* (Salas-Leiva, 2011), we did not found specific protocols designed for the cryopreservation of the microalgae that we use for *O. dioica* feeding, and therefore, we modified the conditions of some general protocols that used dimethyl sulfoxide (DMSO) as cryoprotectant (Day, 2007). To cryopreserve our microalgae, 1 mL of densely growth culture (i.e., OD > 0.25 for *Isochrysis sp*. and *R. reticulata;* OD > 0.35 for *C. calcitrans*, and OD > 0.4 for *Synechococcus sp*.) was gently mixed with 1 mL of a DMSO (D8148, Sigma) stock prepared at different concentrations in sSW for each microalga–30%v/v in sSW for *R. reticulata* and *Isochrysis sp*., 20% for *C. calcitrans*, and 10% for *Synechococcus sp*.–in a cryotube (479–6847 Cryo Tube, Nunc Thermo Scientific, VWR), incubated at 19°C during 45 min for *Isochrysis sp*. and *R. reticulata*, 15 min in darkness for *C. calcitrans*, and 5 min for *Synechococcus sp*., transferred into Mr Frosty™ device (5100-001, Nalgene) and incubated in a −80°C ultrafreezer during at least 3 h, and finally stored into liquid nitrogen until used. For culture recovering, cryopreserved microalgae were thawed in a 40°C water-bath during 1–2 min, until ice crystals melted, transferred into 50 mL of modified Conway medium in a 100 mL Erlenmeyer and maintained at 13°C with a 12 h photoperiod. Typically, after 2–3 weeks, microalgal growth was obvious by eye and spectrophotometrically with an OD > 0.1 (Supporting Information Fig. 2). We have shown so far that microalgal cryopreserved stocks were viable at least after 3 months of storage in liquid nitrogen, although additional experiments are needed to validate the quality of the cryopreserved stocks after very long-term storage (i.e., years).

***O. dioica* culturing and mating strategy for line maintenance.** Animals captured from the coast of Barcelona were transferred to fSW in 8-liter polycarbonate beakers (Camwear Round RFSCW8, Cambro) using the large opening of 10 mL plastic pipettes (900036, Deltalab). We added 2–3 pearls of 1-Hexadecanol (258741-500G, Sigma) to each beaker to reduce the surface tension, and 10 g of prerinsed activated charcoal pellets (22631.293 charcoal 0.85–1.7 mm gradient, VWR) to each beaker to preserve water quality, with the exception of the mating beakers in order to not affect embryo development. Animals were maintained in suspension by the rotation of a methacrylate paddle (10 cm × 23 cm) driven at 5–6 rpm by an electric motor (Synchronous motor, Kelvin) mounted on the beaker lid (Fig. 1h,i).

To start an *O. dioica* line (day 0: D0), 20–30 mature females and 10–15 mature males were transferred into a mating beaker with 1.5 liters of fSW at 19°C (Fig. 1j) in which 10 mL of *Isochrysis sp*., 1 mL of *C. calcitrans*, 5 mL of *R. reticulate*, and 5 mL of *Synechococcus sp*. from the 800 mL microalgal cultures had been added. In the next few hours, animals spawned, generating a progeny of thousands of individuals. Next day morning (day 1: D1), one and a half liter of fresh fSW was gently added up to a final volume of 3 liters (1/2 dilution), and animals were fed with 10 mL of *Isochrysis sp*., 2 mL of *C. calcitrans* and 10 mL of *Synechococcus sp*. In the morning of day 2 (D2), 3 liters of fresh fSW were gently added, up to a final volume of 6 liters (1/2 dilution), and animals were fed with 20 mL of *Isochrysis sp*., 4 mL of *C. calcitrans*, 10 mL of *R. reticulate*, and 10 mL of *Synechococcus sp*. At day 3 (D3), from the hundreds of animals available, 100 individual of approximately the same size (to promote that they will mature at the same time) were manually transferred to a new beaker containing 4 liters of fresh fSW, and fed with 20 mL of *Isochrysis sp*., 8 mL of *C. calcitrans*, and 20 mL of *R. reticulata*. At this point, the setting of additional 100-animal beakers allows to scale up on demand the *O. dioica* culture. The beaker with the remaining animals can be maintained as a "backup" population by daily feeding it with 20 mL of *Isochrysis sp*., 8 mL of *C. calcitrans*, 20 mL of *R. reticulate*, and 10 mL of *Synechococcus sp*. At day 4 (D4), two liters of fSW were added gently to each 100-animals culture up to a final volume of 6 liters (2/3 dilution), fed with 30 mL of *Isochrysis sp*., 8 mL of *C. calcitrans*, and 30 mL of *R. reticulata* and incubated at 13°C to slow down the maturation of the gonads. In the morning of day 5 (D5, typically from 9:00 AM to 12:00 PM), most animals were sexually mature and one (or more) new cultures (D0) were started by mating again 20–30 females and 10–15 males.

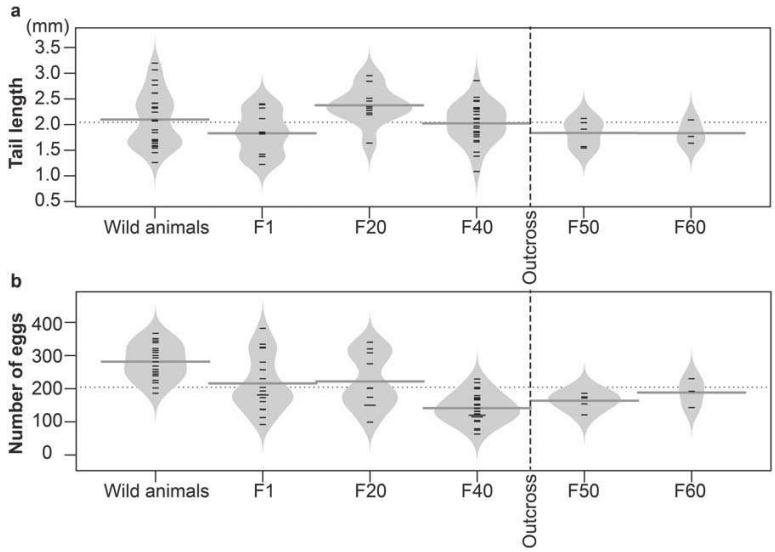


FIG. 2. Analysis of tail lengths (**a**) and number of eggs (**b**) to test the quality and fecundity of inbred populations in the *O. dioica* facility. Tail lengths of mature females and number of eggs of individual spawns are represented as short horizontal lines in one-dimension scatter beanplot (some lines appear bigger because of the coincidence of values in different measurements). Grey areas represent the density distribution of the values, and the thick grey lines indicate the average of tail lengths or number of eggs within each generation. The horizontal dotted line shows the overall average of tail length or number of eggs. The tail length and number of eggs for inbred animals for 1 generation (F1, number of analyzed females $n = 12$), more than 20 (F20, $n = 10$), 40 (F40, $n = 27$), 50 (F50, $n = 5$) or 60 (F60, $n = 5$) generations were compared with those of wild animals ($n = 27$). No significant differences were observed in the length of the tails between inbred animals and wild animals (ANOVA: P-value = 0.061). Comparison of the clutch size was not significantly different between wild animals and inbred animals during the first 20 generations (see Table S1 for statistical significance). At F40, however, the average of the clutch size in some *O. dioica* inbred lines decreased significantly. We then generated an hybrid population by outcrossing two lines independently maintained in the facility (dashed line), and we observed that although the clutch size after the outcross did not significantly improve, the outcross resulted in similar clutch sizes during at least the next 20 generations. The data suggest, therefore, that the inbreeding generated during our husbandry strategy did not affect the average size of animals, and that outcrossing lines can help to limit the reduction of the fecundity observed after many generations of inbreeding. In any case, the overall egg production of our animal facility produced clutch sizes big enough (170 ± 60 eggs in average) to fulfill the experimental needs.

Quality and stability of the *O. dioica* culture through generations. To test the quality and the stability of our *O. dioica* cultures over time, we compared size, fecundity, and fertility parameters between wild animals and inbred animals in our animal facility for 1 (F1) or more than 10 (F10), 20 (F20), 40 (F40), 50 (F50) or 60 (F60) generations. Specifically, the parameters analyzed were the tail length of mature females, the number of eggs per spawn, and the success of fertilization and embryo development up to the hatch. To discard that differences between wild and inbred animals were because of environmental factors (e.g., temperature or food differences in the sea *vs*. our facility) rather than inbreeding or artificial selection derived from our mating strategy, we generated a first generation from wild collected animals maintained in different beakers and raised under our husbandry system. Genetically unrelated animals coming from different beakers were crossed, and the size, fecundity, and fertility values of the born animals were taken as representative measures of those parameters for genetically nonrelated "wild" animals raised under normalized environmental laboratory conditions.

For size comparison, we measured the tails of mature females randomly selected from each group of animals (i.e., wild animals and animals from F1 to F60 generations of inbreeding; $n \geq 10$). Animals were photographed after natural spawning using a Nikon Coolpix E995 camera in a Leica Wild M3Z Kombistereo microscope, and the sizes of their tails were measured with the ImageJ software (http://imagej.nih.gov/ij/). The average and standard deviation of the tail lengths of mature females through different generations F1, F20, F40, F50, and F60 were 1.8 mm ± 0.4 mm, 2.4 mm ± 0.4 mm, 2.0 mm ± 0.4 mm, 1.8 mm ± 0.3 mm, and 1.8 mm ± 0.2 mm, respectively, which were not significantly different among them, neither to the lengths of nonrelated wild animals, 2.1 mm ± 0.5 mm (ANOVA: P-value = 0.061, Fig. 3a). We concluded therefore that neither artificial selection nor inbreeding derived from the crossing strategy of our culture system did affect the size of the animals over time.

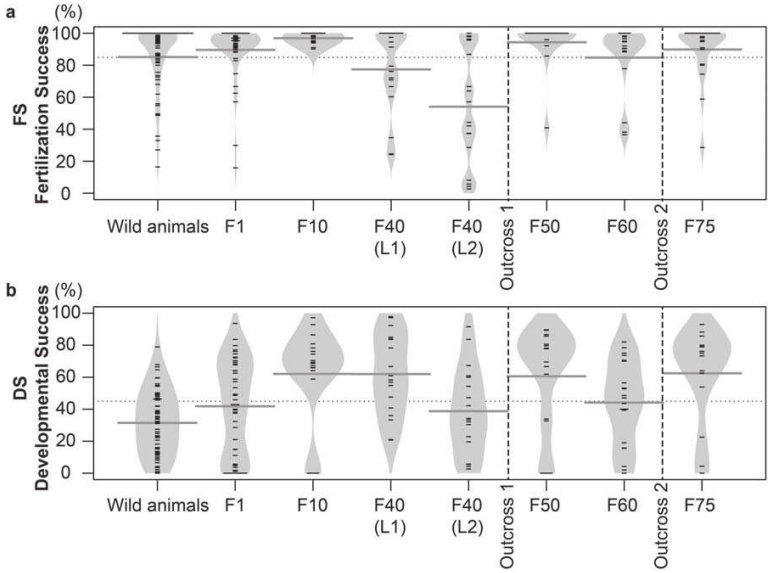


FIG. 3. Analysis of the Fertilization Success (FS) (**a**) and Developmental Success (DS) (**b**) to evaluate the fertility of inbred populations in the *O. dioica* facility. FS and DS values from in vitro fertilization experiments crossing wild animals or inbred animals from the facility (Table 1) are represented as short horizontal lines in one-dimension scatter beanplot (some lines appear bigger because of the coincidence of values in different crosses). Grey areas represent the density distribution of the values, and the thick grey line indicates the average values of FS or DS within each generation. The horizontal dotted line shows the overall FS or DS averages. The FS and DS values of inbred animals for 1 generation (F1), or more than 10 (F10), 40 (F40), 50 (F50), 60 (F60) or 75 (F75) generations were compared with the FS and DS values of wild animals. Interestingly, the low FS and DS averages of wild animals (85% ± 20 and 31% ± 20, respectively) did significantly increase after 10 generations (F10, FS = 97% ± 3 and DS = 62% ± 30; see Table S2 for statistical significance). After maintaining two independent *O. dioica* lines (L1 and L2) for more than 40 generations, however, their FS and DS averages decreased. We then created a new line by crossing animals from L1 with animals from L2 (outcross 1, dashed left line). The FS and DS averages of the new line were improved after the outcross (e.g., F50), but the DS declined again 20 generations later (F60). For this reason, the *O. dioca* population was outcrossed with wild animals at the F74 generation (outcross 2, dashed right line), leading again to a significant improvement of the FS and DS in the next generation F75. The data showed, therefore, that fertility values of the inbred populations in the facility are maintained similar or even better than in wild animals for several generations, but they can decline in the long-term. Periodical outcrosses that increase the genetic variability of the *O. dioica* populations in the facility appear to overcome the decline of fertility because of the inbreeding.

We also counted and photographed the eggs in the spawns of the previously selected females from each group. The averages and standard deviation of clutch sizes from females in F1, F20, and F40 were 283 ± 50; 222 ± 83, and 141 ± 42 eggs, respectively, whereas the average of eggs from nonrelated wild females was 237 ± 81 eggs. The average production of eggs did not show significant differences during at least the first 20 generations (see Table S1 in Supporting Information for *P*-values of the pairwise comparisons). By generation 40, however, the average number was significantly lower, suggesting that the inbreeding or the selection derived from our mating strategy might be diminishing the fecundity of the lines. The outcross of two independent inbred lines stopped, however, this decreasing trend since the hybrid line showed a slight improvement of egg production, which was maintained with no significant differences after 10 or 20 extra generations (164 ± 23 at F50 and 188 ± 44 at F60, respectively). We concluded, therefore, that despite some variability could occur through generations, the overall egg production of our animal facility produced spawns big enough (170 ± 60 eggs in average) for most experimental needs of standard laboratories.

To test if the inbreeding affected the fertility of inbred animals, we measured the percentage of eggs that after fertilization proceeded to the first cleavage (fertilization success, FS) and the percentage of embryos that properly develop at least until hatching (developmental success, DS) through many generations of inbreeding (from F1 to F75). The experimental design for this analysis consisted in making crossing-grids in which a subset of eggs from individual females (e.g., females 1, 2, and 3) were in vitro fertilized with sperm of individual males (e.g., males 1, 2, and 3). The in vitro fertilization protocol followed the next steps. Mature females were individually transferred to glass dishes and rinsed twice with 5 mL of sSW to remove any potential contaminant sperm from the culture. Each female was allowed to naturally release the eggs, and each clutch was subdivided in batches of at least 20 eggs, which were maintained at 13°C for no more than

Table 1
Fertilization Success and Development Success of Wild Animals and Partially Inbred Populations in the O. dioica Facility

	WA	F1	F10	F40 (L1)	F40 (L2)	F50	F60	F75
FS (%±SD)	85 ± 20	90 ± 17	97 ± 3	77 ± 26	54 ± 36	85 ± 22	90 ± 17	94 ± 15
DS (%±SD)	31 ± 20	42 ± 30	62 ± 30	62 ± 25	39 ± 27	61 ± 32	44 ± 26	62 ± 30
Number of eggs	4863	2987	1006	846	848	882	973	613
Number of crosses	65	48	18	18	18	18	21	15

For each generation (from parental wild animals WA to F75 as described in Fig. 3), means (%) and standard deviations (SD) of fertilization success (FS) and development success (DS) were estimated by counting the number of eggs of independent crosses that after in vitro fertilization proceeded to the first cleavage (FS) or properly develop at least until hatch (DS). Wild animals (WA), generations F1-F75, L1 and L2 are different animals lines.

2 hours until being fertilized. Mature males were also rinsed twice in 5 mL of sSW, and placed individually in an eppendorf tube with the smallest possible volume of sSW (e.g., 1–2 μL) to minimize sperm activation after natural spawning but avoiding desiccation. Each male was allowed to naturally release the sperm and just before fertilization, it was diluted in 1 mL of sSW. For the in vitro fertilization, eggs were placed in glass dishes containing 800 μL of sSW and inseminated with 200 μL of the sperm dilution at 19°C. Ten minutes after insemination most polar bodies were visible, eggs were transferred to 3 mL of sSW to avoid polyspermy, and embryo development was allowed to proceed at 19°C. We evaluated the FS and DS values by photographing and counting with ImageJ the number of embryos in each batch that underwent at least one cell division and properly developed until hatching, respectively. Crosses in which one female or one male consistently showed FS below 25% in all its matings were discarded to avoid an underestimation of the FS because of technical problems (i.e., sperm or eggs were not properly manipulated). We performed a total of 156 crosses among inbred animals (Table 1) using 13 females and 11 males from F1, 6 females and 6 males from F10, 16 females and 14 males from F40, 6 females and 6 males from F50, 5 females and 9 males from F60, and 6 females and 6 males from F75. For wild animals, the experiment was done twice performing 65 crosses using 20 females and 19 males from animals collected during the winter-spring seasons of two consecutive years, 2013 and 2014.

The average FS and DS of wild animals were 85% ± 20 and 31% ± 20 (embryos analyzed $n = 4863$; Table 1 and Fig. 3), respectively. In the first generation of inbreeding F1 (embryos analyzed $n = 2987$), the FS was 90% ± 17, and the DS was 42% ± 30, both similar to the FS and DS values of wild animals (Table 1 and Table S2 for *P*-values of pairwise comparisons). Interestingly, the FS and the DS improved after more than 10 generations (F10 embryos $n = 1006$), since FS (97% ± 3) and DS (62% ± 30) were significantly higher than those for wild or F1 animals (*P*-value < 0.05 and < 0.001, respectively; Table 1 and Table S2). After many generations of inbreeding, however, some of the lines appeared to suffer a reduction of their FS and/or DS average values (e.g., see for example FS and DS in the F40 of line 2 or DS in F60 in Fig. 3). To overcome this limitation, we periodically followed an outcrossing strategy either by "internally outcrossing" inbred lines that had been independently maintained (outcross 1 in Fig. 3), or by "externally outcrossing" inbred lines with newly collected wild animals (outcross 2 in Fig. 3) to reduce the inbreeding of the *O. dioica* populations in the facility. These outcrosses resulted in significant improvements of the FS and DS average values as can be observed for instance in the FS average value after outcross 1 (F50) and 2 (F75), or the DS average value after outcross 2 (F75) (*P*-values < 0.05).

Overall, our data indicated that our animal facility is capable to maintain *O. dioica* populations in a healthy reproductive status through many generations, showing FS and DS values similar or even higher than those from wild animals. Despite that this status is not stably maintained since egg production, and the FS and DS parameters might decline after many generations of inbreeding, we have shown that a periodical outcrossing strategy appears as a possible solution to overcome this limitation.

DISCUSSION

Optimization of the *O. dioica* Facility to Develop a Low-Cost System

For many decades, the husbandry of *O. dioica* has been challenging for two main reasons. First, the difficulty to maintain this fragile planktonic animal that lives within a house suspended in a frequently refreshed seawater system. Second, fulfilling the feeding needs for maintaining the physiology and fecundity of *O. dioica* in healthy conditions. Recently, two main facilities—in the SARS Institute and in the University of Osaka—have successfully addressed these challenges and have been able to construct a husbandry system capable to indefinitely maintain *O. dioica* populations throughout the year (Bouquet *et al*., 2009; Nishida, 2008). The SARS facility, for instance, routinely maintains four or five populations, and can generate 6,000 mature animals

per week, which allows the capability to produce tens of thousands of embryos any given day (Bouquet *et al.*, 2009). The production of *O. dioica* at this large scale, however, requires significant amount of space, specialized equipment, large volumes of seawater, and extensive human resources (Bouquet *et al.*, 2009). Our aim was to investigate how to scale-down previously described *O. dioica* culture systems in order to minimize the amount of space, human resources, and technical equipment, but affecting the less as possible to animal production. In comparison to large-scale *O. dioica* production systems, the simplification implemented in our husbandry system implies 95% savings of seawater (from 220 liter to 12 liter per generation), 93% savings of beaker manipulation (from 29 to 2 beakers per generation cycle), 90% savings of beaker maximum occupation (from 10 to 1 beaker in any given day per line), and 87% savings of alga-production (from 2 liter to <0.25 liter per generation). We have also considerably minimized labor tasks such as a 94% reduction in the number of animals to be transferred per line (from more than 2400 to 150), and a reduction to half in the feeding frequency (once rather than twice per day). In our case, the total time dedicated by one person to the maintenance of each animal line is $\sim$3 hours per cycle, ranging from 20 minutes in days D1, D2, and D4, to 1 hour on D0 and D3. In the case of microalgal culturing, we spend 1 hour per week for the setting up of the feeding cultures, and 1 hour per month for long-term stock renewal. Obviously, these savings and reductions were accompanied by a significant cutback of the production capability of our facility (see Discussion below), and the trade-off between production and cost of the facility is a parameter that should be considered in the face of the requirements and resources of each laboratory.

Other important aspects we have also improved are the tasks of maintenance of microalga stocks. The use of long-term liquid or solid cultures of microalgae allowed us to renew the stocks only once every month (Supporting Information Fig. 1), and to isolate single colonies from solid cultures in case of accidental contamination (Fig 1b). We have also developed novel protocols to cryopreserve the four species of microalgae, which allowed us to store microalgae stocks in liquid nitrogen without any additional manipulation, and to recover microalgal cultures in case of accidental loss of a strain or contamination.

Production of a "Low-Cost" *O. dioica* Facility

Our *O. dioica* facility has been able to maintain multiple inbred lines during 2 years, some of them over 75 generations with sporadic outcrosses between genetically nonrelated lines or wild animals. We normally maintain at least two independent inbred lines, and a set of three nonsynchronized "backup" cultures with animals that spontaneously breed. Typically, our *O. dioica* facility generates more than 150 mature animals per week, which allows us to produce more than 10,000 embryos. Despite our production is far below the 6000 mature animals reported for the SARS Institute facility, our system is expandable on demand easily 10-fold by increasing the number of matings at day 0, and by increasing the number of animals transferred at day 3 (Fig. 1j).

The characterization of the animals of our facility after many generations of inbreeding reveals that the average length tail (2.1 mm), which is a measure of animal size that positively correlates with trunk length, gonad length and house size (Lobón *et al.*, 2011), is not significantly different from wild animals collected in the coast of Barcelona. We can conclude, therefore, that the size of the animals did not seem to be affected by the environmental conditions of our husbandry system, neither our mating strategy nor the inbreeding and selection accumulated in our laboratory lines. This average length, however, is lower than the average length reported in animals collected in other locations [e.g., 3.3 mm in animals from the Cantabric sea in the north coast of Spain, (Lobón *et al.*, 2011)], which suggests that length may vary among different populations.

Despite the size of the animals of our facility did not seem to significantly change through generations, we have observed that parameters related to fecundity such as egg production, or fertility such as fertilization and developmental success did appear to be susceptible to the inbreeding accumulated by our mating strategy. For this reason, we suggest to periodically test for egg production, fertilization success, and developmental success, for instance every 20 generations, to evaluate the status of the animal populations of the facility. We also recommend the maintenance of some genetic variability in the facility such as independent inbred lines or independent "backup" beakers, or the sporadic input of new genetic variability from wild animals or cryopreserved sperm (Ouchi *et al.*, 2011) to make out-crosses in order to reduce inbreeding effects when detected. Similar outcross strategies have been also recommended for large-scale facilities (Bouquet *et al.*, 2009), in which values of developmental success around 80–90% are typically observed. The application of out-cross strategies to compensate the negative effects of inbreeding appears therefore fundamental to maintain the cultures healthy, specially in the case of small-scale *O. dioica* facilities like ours that hold a limited amount of genetic variability. The overall average egg production in our facility was 170 ± 60 eggs per female, which is lower than in wild animals collected in the coast of Barcelona (237 eggs per female), or animals from the SARS institute in Norway (303–388, Bouquet *et al.*, 2009; Troedsson *et al.*, 2002), but similar to the reported 165 eggs per female of animals

from the SARS raised in Villefranche-sur-Mer (France) (Lombard *et al*., 2009), or higher than the 122 eggs per female from animals from the Cantabric sea (Lobón *et al*., 2011). It is clear, therefore, that in addition of the differences in egg production that may exist among populations, parameters such as feeding frequency, animal density, and seawater renewal of the animal culture can significantly affect egg production. Overall, these parameters have been optimized in large-scale facilities for producing large quantities of biological material, while in our case we prioritized the reduction of the costs of the facility. Our husbandry protocol opens the possibility that each laboratory can choose the production/cost trade-off that best agrees with its requirements and resources.

We found interesting that the average value of the developmental success was unexpectedly low in wild animals ($31\% \pm 20$), and that this value significantly improved after 10 generations ($62\% \pm 30$). A possible explanation is that our breeding system artificially selects those variants that better adapt to our laboratory conditions, which will be already widespread after 10 generations. An alternative—although not mutually exclusive—explanation is that the low value of the developmental success in wild animals is due to the mutational load resulting from accumulated deleterious mutations in wild populations. The mutational load in *O. dioica* could be high because of the significant population mutation rate reported for this species, $\theta = 4\mathrm{Ne}\mu = 0.0220$ (Denoeud et al., 2010). The combination of the purifying selection and the genetic bottlenecks of our mating strategy may be causing the purge of many deleterious mutations, resulting in improved FS and DS values during the first generations. This purge, however, could be accompanied by a slow widespread of some recessive lethal mutations by genetic drift that started to have a negative effect on FS and DS values due to the inbreeding generated after many generations, for example, F40. This negative effect of the inbreeding on the fertility parameters would be compatible with the observation that fertility can be improved by outcrossing independent inbred lines or by occasionally introducing wild animals in the system, which entail an increase on the genetic variability of the *O. dioica* populations in the facility. Further analyses of *O. dioica* genetic variability related to developmental genes in natural populations will help to find out the mechanisms responsible for the low rate of developmental success in wild animals and to better understand the developmental constraints and plasticity that underlay the evolution of the mechanisms of development, a central question of Evo-Devo.

CONCLUSIONS

In conclusion, we have described how to set up an *O. dioica* "low-cost" facility that is reliable, flexible, easily expandable on demand, and capable to produce thousands of embryos and hundreds of adults every week. Our results suggest that, eventually, inbreeding may affect the fertility of the laboratory populations over many generations, a limitation that can be overcome with outcrossing strategies. Our optimization of the husbandry system of *O. dioica* also includes new protocols for microalgae long-term maintenance and cryopreservation. We hope that the facility described here will facilitate the research with *Oikopleura* and pave the way for spreading *O. dioica* as a model organism.

ACKNOWLEDGMENTS

The authors are thankful to the staff of the *O. dioica* facility in the SARS institute for their expert guidance during the initial set up of our facility and for kindly providing with microalgae cultures. We also thank H. Nishida and T. Omotezako for sharing helpful experiences while visiting their facility and J.H. Postlethwait for helpful advice and support. The authors are grateful to D. Buj, A. Biosca, and B. Colom for their help in the initial setting-up and maintenance of the *O. dioica* facility. The authors thank J.L. Garrido for microalgal culturing material and F. Palau and the staff of the Club Nàutic de Coma-ruga, the Uniland Cementera, S.A., and G. Muñoz and the Badalona City Hall for facilitating the access for animal collections. The authors thank J. Guinea and the staff of the Scientific and Technological Centers (CCiT) at the Facultat de Biologia of the Universitat de Barcelona for providing seawater. They also thank Asociación Española Contra el Cancer for fellowships to AFR and JBR, and AGAUR Generalitat de Catalunya for COLAB fellowships to HGM and JBR, and AAD fellowship to JBR.

LITERATURE CITED

Albalat R, Permanyer J, Cañestro C, Martínez-Mir A, Gonzàlez-Angulo O, Gonzàlez-Duarte R. 2003. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. Cell Mol Life Sci 60:803–809.

Bassham S, Cañestro C, Postlethwait JH. 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol 6:35.

Bassham S, Postlethwait J. 2000. Brachyury (T) expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of T. Dev Biol 220:322–332.

Bassham S, Postlethwait JH. 2005. The evolutionary history of placodes: A molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. Development 132:4259–4272.

Bouquet JM, Spriet E, Troedsson C, Ottera H, Chourrout D, Thompson EM. 2009. Culture optimization for

the emergent zooplanktonic model organism *Oikopleura dioica*. J PLANKTON RES 31:359–370.

Burighel P, Brena C. 2001. Gut ultrastructure of the appendicularian *Oikopleura dioica* (Tunicata). INVERTEBR BIOL 120:278–293.

Cañavate J, Lubirín LM. 1995. Some aspects on the cryopreservation of microalgae used as food for marine species. Aquaculture 136:277–290.

Cañestro C, Albalat R. 2012. Transposon diversity is higher in amphioxus than in vertebrates: Functional and evolutionary inferences. Brief Funct Genomics 11:131–141.

Cañestro C, Albalat R, Postlethwait JH. 2010. *Oikopleura dioica* alcohol dehydrogenase class 3 provides new insights into the evolution of retinoic acid synthesis in chordates. Zool Sci 27:128–133.

Cañestro C, Bassham S, Postlethwait JH. 2005. Development of the central nervous system in the larvacean *Oikopleura dioica* and the evolution of the chordate brain. Dev Biol 285:298–315.

Cañestro C, Bassham S, Postlethwait JH. 2008. Evolution of the thyroid: Anterior-posterior regionalization of the *Oikopleura* endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. Dev Dyn 237: 1490–1499.

Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol 305:522–538.

Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. Nat Rev Genet 8:932–942.

Chavali S, Morais DA, Gough J, Babu MM. 2011. Evolution of eukaryotic genome architecture: Insights from the study of a rapidly evolving metazoan, *Oikopleura dioica*: Non-adaptive forces such as elevated mutation rates may influence the evolution of genome architecture. Bioessays 33:592–601.

Chioda M, Eskeland R, Thompson EM. 2002. Histone gene complement, variant expression, and mRNA processing in a urochordate *Oikopleura dioica* that undergoes extensive polyploidization. Mol Biol Evol 19:2247–2260.

Dalfó D, Cañestro C, Albalat R, Gonzàlez-Duarte R. 2001. Characterization of a microsomal retinol dehydrogenase gene from amphioxus: retinoid metabolism before vertebrates. Chem Biol Interact 130–132:359–370.

Day JG. 2007. Cryopreservation of microalgae and cyanobacteria. Methods Mol Biol 368:141–151.

Delsman HC. 1910. Beiträge zur Entwicklungsgeschichte von *Oikopleura dioica*. Verh. Rijksinst. Onderz. Zee 3:1–24.

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

Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. 2008. Additional molecular support for the new chordate phylogeny. Genesis 46:592–604.

Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Cañestro C, Bouquet JM, Danks G, Poulain J, Campsteijn C, Adamski M, Cross I, Yadetie F, Muffato M, Louis A, Butcher S, Tsagkogeorga G, Konrad A, Singh S, Jensen MF, Cong EH, Eikeseth-Otteraa H, Noel B, Anthouard V, Porcel BM, Kachouri-Lafond R, Nishino A, Ugolini M, Chourrout P, Nishida H, Aasland R, Huzurbazar S, Westhof E, Delsuc F, Lehrach H, Reinhardt R, Weissenbach J, Roy SW, Artiguenave F, Postlethwait JH, Manak JR, Thompson EM, Jaillon O, Du Pasquier L, Boudinot P, Liberles DA, Volff JN, Philippe H, Lenhard B, Roest Crollius H, Wincker P, Chourrout D. 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330:1381–1385.

Fenaux R. 1998. Life history of the Appendicularia. In: Bone Q, editor. The biology of Pelagic tunicates. Oxford: Oxford University Press. pp 151–159.

Fenaux R, Gorsky G. 1979. Techniques d'elevages des Appendiculaires. Ann Inst Oceanogr Paris 55:195–200.

Fenaux R, Gorsky G. 1985. Techniques d'elevages des Appendiculaires. Rapp Comm Int Mer Médit 29: 291–292.

Fujii S, Nishio T, Nishida H. 2008. Cleavage pattern, gastrulation, and neurulation in the appendicularian, *Oikopleura dioica*. Dev Genes Evol 218:69–79.

Ganot P, Bouquet JM, Kallesoe T, Thompson EM. 2007. The *Oikopleura* coenocyst, a unique chordate germ cell permitting rapid, extensive modulation of oocyte production. Dev Biol 302:591–600.

Ganot P, Moosmann-Schulmeister A, Thompson EM. 2008. Oocyte selection is concurrent with meiosis resumption in the coenocystic oogenesis of *Oikopleura*. Dev Biol 324:266–276.

Lobón CM, Acuña JL, López-Álvarez M, Capitanio FL. 2011. Heritability of morphological and life history traits in a pelagic tunicate. Marine Ecol Prog Series 422:145–154.

Lombard F, Renaud F, Sainsbury C, Sciandra A, Gorsky G. 2009. Appendicularian ecophysiology I: Food concentration dependent clearance rate, assimilation efficiency, growth and reproduction of *Oikopleura dioica*. J Marine Sys 78:606–616.

Lopez-Urrutia A, Acuña JL. 1999. Gut throughput dynamics in the appendicularian *Oikopleura dioica*. Mar Ecol Prog Ser 191:195–205.

Maher B. 2009. Evolution: Biology's next top model? Nature 458:695–698.

Nishida H. 2008. Development of the appendicularian *Oikopleura dioica*: Culture, genome, and cell lineages. Dev Growth Differ 50 (Suppl 1):S239–S256.

Nishino A, Satou Y, Morisawa M, Satoh N. 2000. Muscle actin genes and muscle cells in the appendicularian, *Oikopleura longicauda*: phylogenetic relationships among muscle tissues in the urochordates. J Exp Zool 288:135–150.

Nishino A, Satou Y, Morisawa M, Satoh N. 2001. Brachyury (T) gene expression and notochord development in *Oikopleura longicauda* (Appendicularia, Urochordata). Dev. Genes Evol. 211:219–231.

Omotezako T, Nishino A, Onuma TA, Nishida H. 2013. RNA interference in the appendicularian *Oikopleura dioica* reveals the function of the Brachyury gene. Dev Genes Evol 223:261–267.

Ouchi K, Nishino A, Nishida H. Simple procedure for sperm cryopreservation in the larvacean tunicate *Oikopleura dioica*. Zool Sci 2011;28:8–11.

Paffenhöfer G-A. 1973. The cultivation of an appendicularian through numerous generations. Mar Biol 22: 183–185.

Salas-Leiva JS, Enrique D. 2011. Cryopreservation of the microalgae *Chaetoceros calcitrans* (Paulsen): Analysis of the effect of DMSO temperature and light regime during different equilibrium periods. Lat Am J Aquat Res 39:271–279.

Sato R, Yu J, Tanaka Y, Ishimaru T. 1999. New apparatuses for cultivation of appendicularians. Plankton Biol Ecol 46:162–164.

Soviknes AM, Chourrout D, Glover JC. 2007. Development of the caudal nerve cord, motoneurons, and muscle innervation in the appendicularian urochordate *Oikopleura dioica*. J Comp Neurol 503:224–243.

Stach T, Turbeville JM. 2002. Phylogeny of Tunicata inferred from molecular and morphological characters. Mol Phylogenet Evol 25:408–428.

Stach T, Winter J, Bouquet JM, Chourrout D, Schnabel R. 2008. Embryology of a planktonic tunicate reveals traces of sessility. Proc Natl Acad Sci USA 105:7229–7234.

Swalla B, Cameron C, Corley L, Garey J. 2000. Urochordates are monophyletic within the deuterostomes. Syst. Biol. 49:52–64.

Troedsson C, Bouquet J-M, Aksnes DL, Thompson EM. 2002. Resource allocation between somatic growth and reproductive output in the pelagic chordate *Oikopleura dioica* allows opportunistic response to nutritional variation. Marine Ecol Prog Series 243:83–91.

Troedsson C, Bouquet J-M, Lobon CM, Novac A, Nejstgaard JC, Dupont S, Bosak S, Jakobsen HH, Romanova N, Pankoke LM, Isla A, Dutz J, Sazhin AF, Thompson EM, Dupont S, Pörtner H-O. 2013. Effects of ocean acidification, temperature and nutrient regimes on the appendicularian *Oikopleura dioica*: A mesocosm study, Marine Biol 2013;160:2175–2187.

Wada H. 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. Mol Biol Evol 15:1189–1194.

Supporting Information for

***Oikopleura dioica* culturing made easy: A low-cost facility for an emerging animal model in EvoDevo.**

Josep Martí-Solans,[1] Alfonso Ferrández-Roldán,[1] Hector Godoy-Marín,[1] Jordi Badia-Ramentol,[1] Nuria Paz Torres-Águila,[1] Adriana Rodríguez-Marí,[1] Jean Marie Bouquet,[2,3] Daniel Chourrout,[2] Eric Thompson,[2,3] Ricard Albalat,[1] Cristian Cañestro[1]

[1]Departament de Gene_tica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, 08028, Spain

[2]Sars International Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Bergen, Norway

[3]Department of Biology, University of Bergen, Postbox 7803, N-5020 Bergen, Norway

This PDF file includes:

Suplementary text:

Collection of wild animals

Adult animals were collected in three different locations of the Catalonian coast near Barcelona (41°26'24.6"N 2°14'56.2"E, 41°14'22.4"N 1°51'59.6"E, 41°10'25.2"N 1°31'37.9"E) at various distances of the coastline, from 0 to 1 km, and depths ranging from 0 to 30 m (Fig. 1a). Animals were collected either with a plankton net (model 90-50x3 150A, Sea Gear) filtering around 3000 liters of seawater in each pass, or directly from approximately 200 liters of seawater collected from the surface with a bucket. Animal collections were kept cold using portable coolers to avoid undesired increase of the temperature and stress during transportation. Although a systematic sampling has not been performed throughout the year in the coast of Barcelona, we have observed

differences in animal abundance among seasons. During winter, for instance, *O. dioica* was scarce and other larvacean species were present, while a peak of abundance of *O. dioica* occurred during early spring, accompanied by a decrease in the presence of other larvacean species.

Seawater

Seawater was provided by the animal experimentation unit of the Scientific and Technological Centers (CCiT) at the Faculty of Biology of the University of Barcelona. Seawater was collected twice a year by the Institute of Marine Sciences (Barcelona) from an underwater outlet located 300 m from the coast of Barcelona and about 10 m deep, sand filtered, stored in a 35,000 L tank in the dark, and circulated every night for 8 hours for aeration. In our facility, the seawater was filtered through 50 μm and 20 μm polypropylene filters (MBHP050-10 and LTPPBB020-10 AscoFiltri Filtering cartridges, respectively) using the suction force generated by a vacuum system connected with a filling/venting closure (Closure 2162-0830, Nalgene) to a polypropylene 50 L container (Autoclavable Carboys with Spigot 2319-0130, Nalgene) (Fig. 1b). This filtration removes possible sand particles that could affect *O. dioica* buoyancy as well as possible contaminant algae that may grow in the 35,000 L tank. Filtered seawater (fSW) was kept in 50 L polypropylene tanks at 19ºC until use, avoiding storage of filtered water for periods longer than 2 weeks. For *in vitro* fertilization experiments and for microalgal cultures, fSW was sterilized (sterilized seawater, sSW) using 0.22 μm filters (VacuCap PF Filters 4622, Pall Corporation) connected to a vacuum system following manufacture specifications.

Microalgal production for *O. dioca* feeding

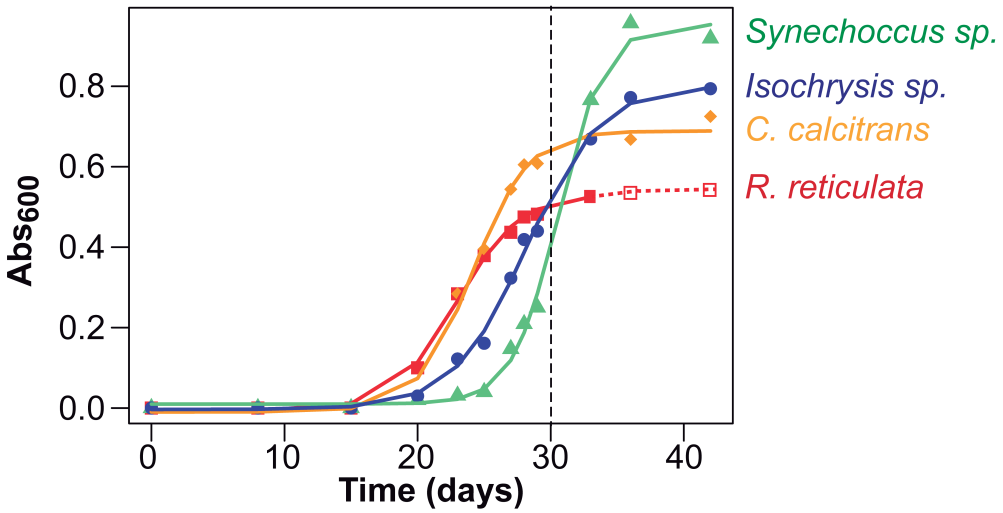
The diet for *O. dioica* was adapted from Bouquet *et al*., (2009) (Bouquet *et al*., 2009) and consisted of four microalga species, three true eukaryotic species of microalga, *Chaetoceros calcitrans* (CCAP 1010/11), *Isochrysis sp.* (CCAP 927/14), *Rhinomonas reticulata* (CCAP 995/2) and one prokaryotic species of cyanobacteria (a.k.a. blue-green algae) *Synechococcus sp.* (K0408), whose starting cultures were

kindly provided by the SARS Institute (Norway). These four microalgal species were grown in modified Conway medium (Bouquet *et al.*, 2009), supplemented with 0,047 mM $Na_2SiO_3 \cdot 5H_2O$ (28092.290 VWR BDH Prolabo) for *C. calcitrans*, at 13ºC with a photoperiod of 12h light/dark (PGA-500 Incubator SCLAB, equipped with three fluorescent tubes 36W color 6500K, Roblan).

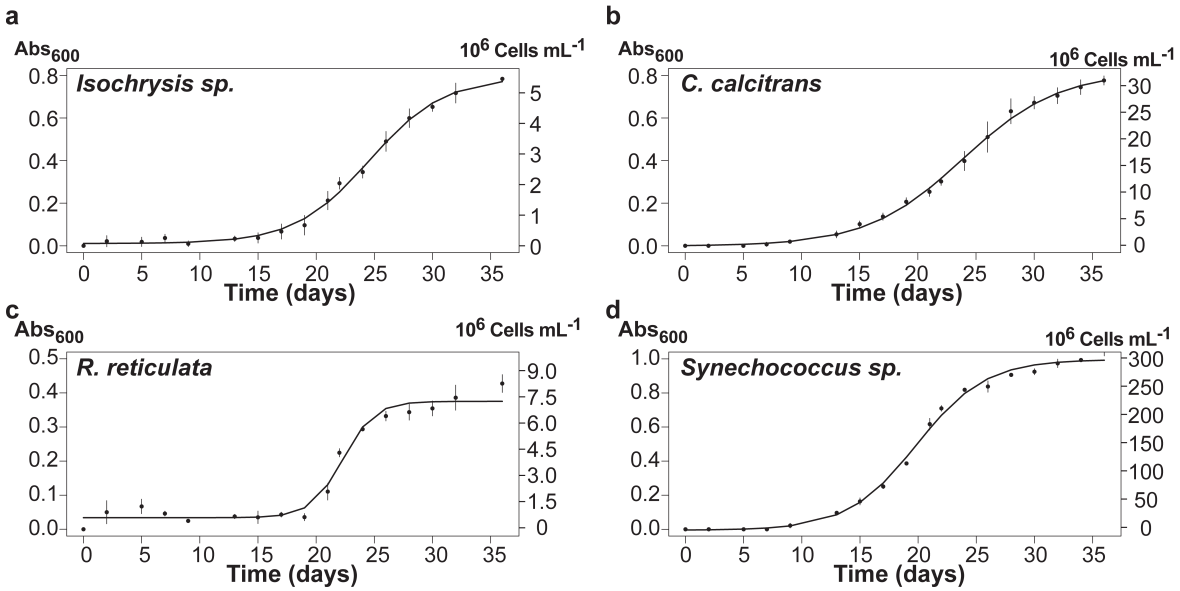
We designed a scalable protocol for the production of the four microalgae in order to minimize the volume of the cultures, but in a flexible way to adapt it on demand to the feeding needs depending on the amount of animals maintained in the laboratory each week. Our protocol aimed to produce in one week the required volume of each microalgal culture within a range of cell density estimated from the optical density (OD) at 600 nm of the culture (Bouquet *et al.*, 2009). OD of the microalgal cultures used for feeding ranged from 0.20 to 0.25 ($8x10^6$ to $1x10^7$ cells/ml) for *C. calcitrans*, from 0.13 to 0.18 ($2 x10^6$ to $3x10^6$ cells/ml) for *Isochrysis sp.*, from 0.14 to 0.20 ($1 x10^6$ to $1.5x10^6$ cells/ml) for *R. reticulata* and from 0.15 to 0.20 ($4.5 x10^7$ to $6x10^7$ cells/ml) for *Synechococcus sp.* Typically, for instance, to produce 800 mL of a microalgae, we inoculated 700 mL of modified Conway medium in a 1L round bottom glass flask with two 50 mL pre-cultures growing in 100 mL Erlenmeyers, which had been, in turn, inoculated one week earlier with a given number of cells ($1 x10^8$ cells for *C. calcitrans*, $4 x10^7$cells for *Isochrysis sp.*, $9 x10^6$ cells for *R. reticulata* and $7 x10^8$ cells for *Synechococcus sp*).

References:

Bouquet JM, Spriet E, Troedsson C, Ottera H, Chourrout D, Thompson EM. 2009. Culture optimization for the emergent zooplanktonic model organism Oikopleura dioica. J PLANKTON RES 31:359-370.



Supplementary Figure 1. Growth curves of monthly renewed long-term liquid cultures of microalgal stocks. Microalgae were grown at 13°C with 12h photoperiod, and growth was inferred from optical densities (Abs600) of each culture at different times (days). We tested inocula ranging from 10 to 5000 cells into 50mL of Conway medium to find the minimal number of cells that guaranteed the recovery of each culture, but delayed the reach of the declining phase as much as possible (data not shown). From this range of cells, we empirically determined that inocula of ≈102 cells for *R. reticulata*, ≈103 cells for *C. calcitrans*, *Isochrysis sp.* and ≈104 cells for *Synechococcus sp.* into 50 mL allowed us to conveniently renew all stocks at the same time just once every 30 days (dotted black line), when all cultures were at the exponential phase or just before entering in the stationary phase. Beyond 40 days, cultures were still viable despite they were in the stationary phase, with the exception for *R. reticulata*, in which after 30 days the condition of the cultured changes as denoted by a color change from the normal vivid red to a dark brown (Abs_{600} measures became unreliable, dotted red line).



Supplementary Figure 2. Growth curves of the four cryopreserved microalgae species at 13°C with 12h photoperiod: (a) *Isochrysis sp.*, (b) *C. calcitrans*, (c) *R. reticulata*, and (d) *Synechococcus sp.*. Cell concentration (Cell mL-1, right axis) at different times (days) was calculated from optical densities (Abs_{600}, left axis) using species-specific equations defined by Bouquet et al. 2009. After thawed, microalgal growth started to be evident by eye and spectrophotometrically with an OD >0.1 approximately at day 20 for *Isochrysis sp*. and *R. reticulata*, at day 15 for C. calcitrans and at day 13 for *Synechococcus sp*..

Table S1. Multiple pairwise comparisons of the mean number of eggs per spawn through generations.

	F1	F20	F40	F50	F60
WA	0.085	0.561	**<0.001*** **	**0.035***	0.387
F1		0.080	**<0.001*** **	**<0.001*** **	**0.017***
F20			**0.007****	0.232	0.554
F40				0.107	0.084
F50					0.381

The number of eggs was not normally distributed (Shapiro-Wilk test; W = 0.955, p-value = 0.009), and therefore, non-parametric Mann-Whitney-Wilcoxon Test was applied. Significant pairwise differences are indicated in bold (*p < 0.05, **p < 0.01, ***p < 0.001). Wild animals (WA), generations F1-F60

Table S2. Multiple pairwise comparisons of the mean of Fertilization Success (FS, up-right values) and Developmental Success (DS, down-left values) through generations.

	WA	F1	F10	F40 (L1)	F40 (L2)	F50	F60	F75
WA		0.057	**0.040***	0.430	**0.001*****	0.776	0.196	**0.006****
F1	0.057		**0.036***	0.370	**0.001*****	0.419	0.254	**0.003****
F10	**<0.001*****	**0.016***		0.064	**<0.001*****	**0.028***	0.653	0.134
F40 (L1)	**<0.001*****	**0.024***	0.681		**0.036***	0.667	0.147	**0.015***
F40 (L2)	0.371	0.671	**0.015***	**0.015***		**0.019***	**0.001*****	**<0.001*****
F50	**<0.001*****	**0.013****	0.680	0.924	**0.041***		0.170	**0.007****
F60	**0.032***	0.691	**0.042***	**0.038***	0.545	**0.048***		0.109
F75	**<0.001*****	**0.008****	0.745	0.789	**0.025***	0.928	**0.019***	

FS and DS values were not normally distributed (Shapiro-Wilk test: W = 0.688, p-value < 0.001; and W = 0.947, p-value < 0.001, respectively), and therefore Mann-Whitney-Wilcoxon test was applied. Significant pairwise differences of FS and DS between generations are indicated in bold (*p < 0.05, **p < 0.01, ***p < 0.001). Wild animals (WA), generations F1-F75, L1 and L2 are different animals lines.

Resultats. Capítol 2

Co-elimination and survival in gene network evolution: dismantling the RA-signaling in a chordate.

Josep Martí-Solans,[1] Olga V. Belyaeva,[2] Nuria P. Torres-Aguila,[1] Natalia Y. Kedishvili,[2] Ricard Albalat[1] i Cristian Cañestro[1]

[1]Departament de Gene`tica, Microbiologia i Estad_ıstica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

[2]Department of Biochemistry and Molecular Genetics, University of Alabama—Birmingham

Resum

La seqüenciació de cada cop més espècies està posant de manifest la pèrdua de gens com una força evolutiva ubiqua que genera diversitat genètica que pot modelar l'evolució de les espècies. Fora de bacteris i llevats, però, la comprensió del procés de pèrdua de gens continua sent esquiva, especialment en l'evolució de les espècies animals. En aquest capítol, utilitzant el desballestament de la xarxa genètica del metabolisme l'àcid retinoic (RA-MGN) a *Oikopleura dioica*, combinem la genòmica comparativa, la filogenètica, la bioquímica i la biologia del desenvolupament per investigar la robustesa mutacional associada a patrons esbiaixats de pèrdua de gens. En aquest capítol, es demostra l'absència de vies alternatives per a la síntesi de àcid retinoic (RA) a *O. dioica*, el que suggereix que la pèrdua de gens de la RA-MGN no és va produir en un escenari mutacional robust, sinó al contrari, es van produir en un escenari d'evolució regressiva. A més, la manca de canvis fenotípics dràstics associats a la pèrdua de senyalització de l'RA proporciona un exemple de la paradoxa inversa d'Evo-Devo. Aquest treball il·lustra com la identificació de patrons de coeliminació gènica -en el nostre cas cinc pèrdues (*Rdh10*, *Rdh16*, *Bco1*, *Aldh1a* i *Cyp26*)- és una estratègia útil per reconèixer mòduls funcionals associats a funcions diferents dins de les xarxes genètiques. El nostre treball també il·lustra com la identificació dels gens supervivents a les pèrdues gèniques d'una via, ajuda a reconeixement d'esdeveniments de neofuncionalització i de les funcions ancestrals. Així, la supervivència i la extensiva duplicació dels enzims *Cco* i *RdhE2* a *O. dioica* es correlaciona amb l'adquisició d'una complexa compartimentació dels dominis d'expressió en el sistema digestiu i un procés de neofuncionalització enzimàtica dels enzims Cco, mentre que la supervivència de l'enzim *Aldh8* podria estar relacionat amb el manteniment d'una funció ancestral detoxificant contra els aldehids.

Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate

Josep Martí-Solans,[1] Olga V. Belyaeva,[2] Nuria P. Torres-Aguila,[1] Natalia Y. Kedishvili,[2] Ricard Albalat,*[,1] and Cristian Cañestro*[,1]

[1]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

[2]Department of Biochemistry and Molecular Genetics, University of Alabama—Birmingham

*Corresponding author: E-mail: ralbalat@ub.edu; canestro@ub.edu.

Associate editor: Yoko Satta

Abstract

The bloom of genomics is revealing gene loss as a pervasive evolutionary force generating genetic diversity that shapes the evolution of species. Outside bacteria and yeast, however, the understanding of the process of gene loss remains elusive, especially in the evolution of animal species. Here, using the dismantling of the retinoic acid metabolic gene network (RA-MGN) in the chordate *Oikopleura dioica* as a case study, we combine approaches of comparative genomics, phylogenetics, biochemistry, and developmental biology to investigate the mutational robustness associated to biased patterns of gene loss. We demonstrate the absence of alternative pathways for RA-synthesis in *O. dioica*, which suggests that gene losses of RA-MGN were not compensated by mutational robustness, but occurred in a scenario of regressive evolution. In addition, the lack of drastic phenotypic changes associated to the loss of RA-signaling provides an example of the inverse paradox of Evo–Devo. This work illustrates how the identification of patterns of gene coelimination—in our case five losses (*Rdh10*, *Rdh16*, *Bco1*, *Aldh1a*, and *Cyp26*)—is a useful strategy to recognize gene network modules associated to distinct functions. Our work also illustrates how the identification of survival genes helps to recognize neofunctionalization events and ancestral functions. Thus, the survival and extensive duplication of *Cco* and *RdhE2* in *O. dioica* correlated with the acquisition of complex compartmentalization of expression domains in the digestive system and a process of enzymatic neofunctionalization of the *Cco*, while the surviving *Aldh8* could be related to its ancestral housekeeping role against toxic aldehydes.

Key words: **gene loss, gene coelimination, regressive evolution, evo–devo, retinoic acid, chordate.**

Introduction

The recent bloom of genomic data is revealing a novel perspective of gene loss as a pervasive source of genetic variation with a great potential to generate phenotypic diversity and to shape the evolution of gene networks (Albalat and Cañestro 2016). Several fundamental questions regarding the evolutionary role of gene loss, however, still remain elusive. How do genes become dispensable and subsequently lost? How do patterns of gene loss appear to be biased rather than stochastic? What is the impact of gene loss on the evolution of the rest of surviving components of gene networks? In a simplified view, gene dispensability and the evolutionary impact of the subsequent loss of dispensable genes have been mainly associated either with the mutational robustness of biological systems (i.e., the capability of maintaining unaltered phenotypes in the face of mutations) or with changes in the functional requirements caused by the relaxation of environmental constraints (Albalat and Cañestro 2016). In the scenario of a robust genetic system, a gene can become dispensable when its loss does not imply the loss of a biological function due either to the presence of "backup genes" (e.g., redundant paralogs with overlapping functions) (Gu et al. 2003) or of "backup pathways" (e.g., alternative pathways by which the biological function can be rerouted) (Wagner 2005). Alternatively, a gene can become dispensable in a scenario of regressive evolution characterized by the loss of useless traits due to relaxation of environmental constraints, in which the loss of trait-related genes becomes neutral—for example, gene losses related to vision and pigmentation in cavefish after colonization of dark environments (Protas et al. 2006)—; or in a scenario of adaptive evolution in which the loss of the function is advantageous, and therefore positively selected—that is, the "less is more" hypothesis, classically exemplified by the advantageous losses of the human genes coding for cell receptors CCR5 and DUFFY that provide resistance to AIDS and vivax malaria, respectively (Olson 1999; Olson and Varki 2003).

Interestingly, patterns of gene loss do not appear to be stochastic, but they show clear biases depending on gene functions or genomic positions of the lost genes (reviewed in Albalat and Cañestro 2016). A special case of functional bias of gene loss is observed in species that suffer relaxation of a given biological or environmental constraint, which leads to the "coelimination" of genes that are functionally linked in

Article

distinct pathways associated with the relaxed constraint (Aravind et al. 2000; Koonin et al. 2004). The identification of patterns of gene coelimination appears therefore as a useful strategy to recognize components of functional modules within gene networks, whereas the identification of "survival genes" after the dismantling of a pathway is useful to recognize "hub" genes that cannot be lost owing their pleiotropic nature acting in multiple pathways (reviewed in Albalat and Cañestro 2016).

Most of the theoretical framework about gene dispensability and biased patterns of gene loss and survival has arisen from laboratory evolution experiments and computational models using bacteria and yeast (reviewed in Albalat and Cañestro 2016). The impact of losses affecting gene networks of multicellular eukaryotes (i.e., animals or plants), however, remains elusive owing their highly elaborated transcriptional regulatory circuits, complex signaling pathways, and frequently convoluted mechanisms of development that are less suitable for experimental or computational studies. Here, we take advantage of the availability of the deeply sequenced genomes of many chordates—including the larvacean urochordate *Oikopleura dioica* characterized by a high propensity to loose genes (Denoeud et al. 2010)—in order to examine in detail the evolution of an animal gene network affected by patterns of gene coelimination and survival. Thus, using the dismantling of the robust retinoic acid metabolic gene network (RA-MGN) in *O. dioica* as a case study, we use a multidisciplinary approach that includes not only evolutionary genomics, and phylogenetic reconstructions but also analytical biochemistry and developmental gene expression analyses, in order to provide experimental evidence supporting a plausible scenario for the gene losses in the RA-MGN and to argue about their impact on the evolution of gene networks and the phenotypic diversification.

The RA-MGN regulates the synthesis and degradation of all-*trans*-retinoic acid (atRA), a vitamin A derived compound that acts as a crucial signaling pathway for embryo development and adult tissue homeostasis in chordates, including humans (fig. 1) (Rhinn and Dolle 2012; Duester 2013; Cunningham and Duester 2015). The canonical pathway that regulates the synthesis of atRA during chordate embryogenesis consists of Rdh10 and Aldh1a (typically Aldh1a1, Aldh1a2, and Aldh1a3; formerly Raldhs) enzymes that catalyze two sequential oxidative reactions, first from all-*trans*-retinol (atROL, vitamin A) to all-*trans*-retinaldehyde (atRAL) and second from atRAL to atRA, respectively (Chen et al. 1995; Niederreither et al. 1999; Dalfó et al. 2002; Duester et al. 2003; Cammas et al. 2007; Belyaeva et al. 2008; Cañestro et al. 2009; Strate et al. 2009). These RA-producing enzymes—Rdh10 and Aldh1a—are coordinated with various RA-degrading enzymes of the cytochrome P450 subfamily 26 (Cyp26; typically Cyp26a, Cyp26b, and Cyp26c) to establish the spatiotemporal levels of atRA (Niederreither et al. 2002; Reijntjes et al. 2005; Schilling et al. 2012).

The RA-MGN has been suggested to be a robust genetic system in vertebrates. In addition to the multiplicity of the enzymes in the canonical pathway, several other enzymes with redundant activities have been suggested to be able to bypass the canonical pathway and produce atRA (fig. 1). Aldh8a1, for instance, may bypass Aldh1a and catalyze the oxidation of atRAL to atRA (Lin and Napoli 2000; Lin et al. 2003; Liang et al. 2008), and RdhE2, Rdh16 and various vertebrate-specific Adhs (e.g., Adh1 and Adh4) may replace Rdh10 and catalyze atROL to atRAL oxidation (Gough et al. 1998; Jurukovski et al. 1999; Cañestro et al. 2000; Molotkov, Deltour, et al. 2002; Molotkov, Fan, et al. 2002; Cañestro et al. 2003, 2010; Lee et al. 2009; Belyaeva et al. 2012). Furthermore, atRAL can be alternatively produced by a "backup pathway" (Lindqvist and Andersson 2004) that consists of the enzymatic cleavage of dietary β-carotene by carotenoid cleavage oxygenases (Cco) enzymes, directly by β,β-carotene-15,15′-monooxygenase (Bco1) enzyme (von Lintig and Vogt 2000; Lampert et al. 2003), or indirectly, via apocarotenoids, by the β,β-carotene-9,10′-dioxygenase (Bco2) enzyme (Wang et al. 1996).

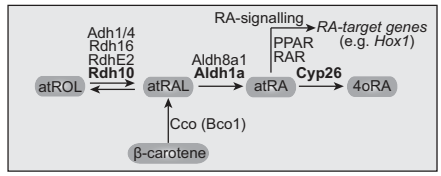


Fig. 1. Schematic representation of the RA-MGN in vertebrates that establishes the levels of atRA, whose signaling regulates developmental genes (e.g., *Hox1*) through nuclear receptors (e.g., RAR and PPAR). The enzymes of the canonical pathway (in bold: Rdh10, Aldh1a, and Cyp26) regulate the synthesis of atRA from atROL (vitamin A) to atRAL precursors. In vertebrates, the RA-MGN is a robust system due to the presence of other redundant enzymes that bypass the canonical pathway (i.e., RdhE2, Rdh16, Adh1/4, Aldh8a1, and Cco).

Previous work had revealed that *O. dioica* has lost two of the components of the canonical RA-MGN (i.e., *Aldh1a* and *Cyp26*) (Cañestro et al. 2006). However, whether *O. dioica* can bypass this loss and is capable of synthesizing atRA using alternative "backup" genes or pathways in a robust genetic system, remained unknown. To evaluate this possibility, in this work, we have first searched for patterns of coelimination affecting the RA-MGN, as well as survival genes of the RA-MGN that could function as redundant genes (e.g., *RdhE2*, *Rdh16*, and *Aldh8a1*) or alternative pathways (e.g., *Cco*). Second, we have investigated whether the dismantling of the RA-MGN was accompanied by the loss of biological function, that is, atRA production. Third, we have analyzed the functions of surviving genes based on their biochemical capabilities and their expression patterns during *O. dioica* development. And finally, we discuss possible evolutionary scenarios that may have facilitated gene losses, and whether the survival genes of the RA-MGN in *O. dioica* were preserved because the retention of ancestral functions or because they underwent neofunctionalization processes.

Results

Genomic Survey of the RA-MGN in *O. dioica*

In order to evaluate the mutational robustness of RA-MGN in *O. dioica*, we first investigated the existence of canonical and

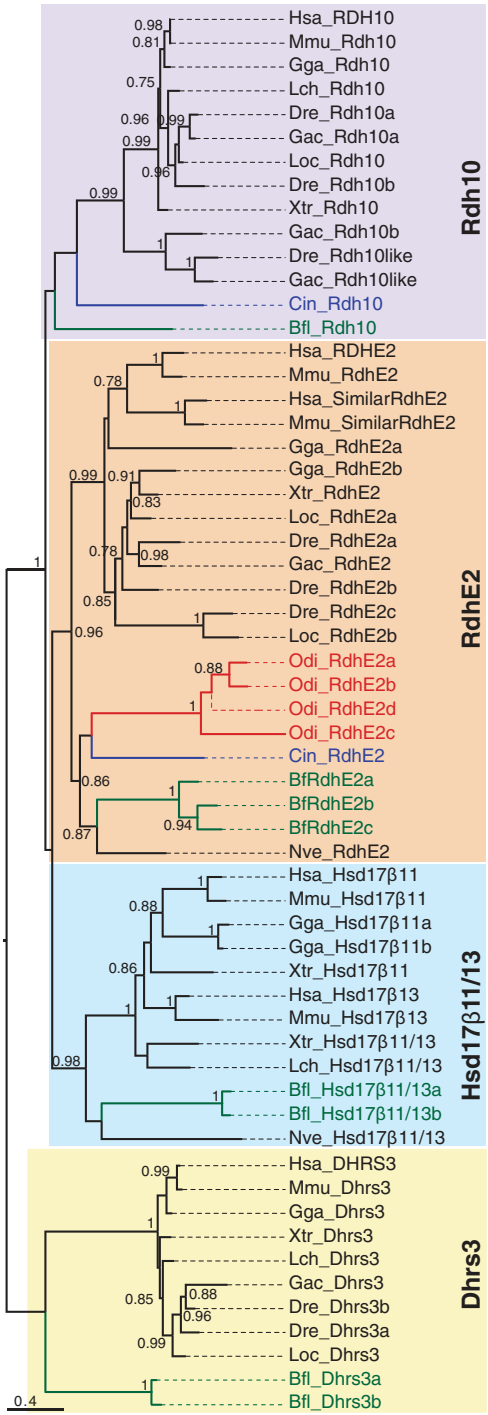


Fig. 2. ML phylogenetic tree of the SDR-16C Rdh10 and RdhE2 subfamilies in chordates revealing the loss of Rdh10, and the surviving and lineage-specific duplication of RdhE2 in *Oikopleura dioica*. The sister SDR-16C subfamily of hydroxysteroid 17-β dehydrogenases 11 and 13 (Hsd17β11/13) and the basal dehydrogenases/reductases member 3 (Dhrs3) as outgroup were included to root the tree. Scale bar indicates amino acid substitutions. Values for the approximate likelihood-ratio test (aLRT) are only shown in nodes with support values greater than 0.7. Vertebrates: *Danio rerio* (Dre), *Gasterosteus aculeatus* (Gac), *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), and *Lepisosteus oculatus* (Loc), *Mus musculus* (Mmu), *Petromizus marinus (Pmr), Xenopus tropicalis* (Xtr); Urochordates: *Ciona intestinalis* (Cin) and *Oikopleura dioica* (Odi); Cephalochordates: *Branchiostoma floridae* (Bfl); Cnidarians: *Nematostella vectensis* (Nve).

backup genes or pathways in this species by surveying its genome databases for all RA-MGN components known in other chordate species (fig. 1).

Rdh10* and *RdhE2: Vertebrate *Rdh10* (for retinol dehydrogenase 10) genes belong to the short-chain dehydrogenase/reductase (SDR)-16C family (Albalat et al. 2011; Belyaeva et al. 2015), and have been suggested to encode for the main enzyme responsible for the synthesis of atRAL from atROL during embryogenesis (Sandell et al. 2007; Belyaeva et al. 2008). *RdhE2* (for epidermal retinal dehydrogenase 2) genes belong to the same SDR-16C family, and they have been shown to have the same biochemical activity (Matsuzaka et al. 2002; Belyaeva et al. 2012). Both Rdh10 and RdhE2 enzymes have the typical domains of SDR enzymes that are essential for their dehydrogenase function: 1) A Gly-rich sequence in the variable N-terminal region critical for accommodation and binding of the cofactor (Lesk 1995), 2) one acidic residue located about 20 residues downstream the Gly-rich sequence necessary for NAD(H) specificity (Wierenga et al. 1986), and 3) the Asn-Ser-Tyr-Lys tetrad that forms the active site necessary for catalysis (Albalat et al. 1992; Filling et al. 2002). BLAST searches of *O. dioica* genome database with *Homo sapiens* and *Ciona intestinalis* Rdh10 or RdhE2 retrieved four *O. dioica* genes (supplementary file S1, Supplementary Material online, CBY07137, CBY07135, CBY07673, and CBY12938; E-values $< e^{-50}$). Their best reciprocal BLAST hits (BRBH) retrieved back RdhE2, but no Rdh10, with E-values ranging from $2e^{-82}$ to $1e^{-89}$. In agreement with the BRBH results, phylogenetic gene trees inferred by maximum likelihood (ML) method showed that the four *O. dioica* genes did not group with the Rdh10 enzymes, but grouped together within the RdhE2 cluster, suggesting that *O. dioica* lacks an ortholog of chordate *Rdh10* genes (fig. 2). The fact that *Rdh10* genes had been identified in the genomes of the cephalochordate *Branchiostoma floridae* and the urochordate *C. intestinalis* (fig. 2) (Albalat et al. 2011; Belyaeva et al. 2015) indicated that the absence of an *O. dioica Rdh10* was due to a gene loss in the RA-MGN of the larvacean lineage. The four *O. dioica* paralogs (named *RdhE2a* to *RdhE2*d) appeared to be orthologous to chordate *RdhE2* genes, and they arose by gene duplications during the evolution of the larvacean lineage (fig. 2). The lineage-specific origin of *O. dioica RdhE2* paralogs was further supported by the presence of several introns in conserved positions among the four *O. dioica RdhE2* genes (red arrowheads in supplementary file S2A, Supplementary Material online), but absent in any other analyzed species. Sequence analysis of the four *O. dioica* RdhE2 enzymes showed amino acid conservation in the most important

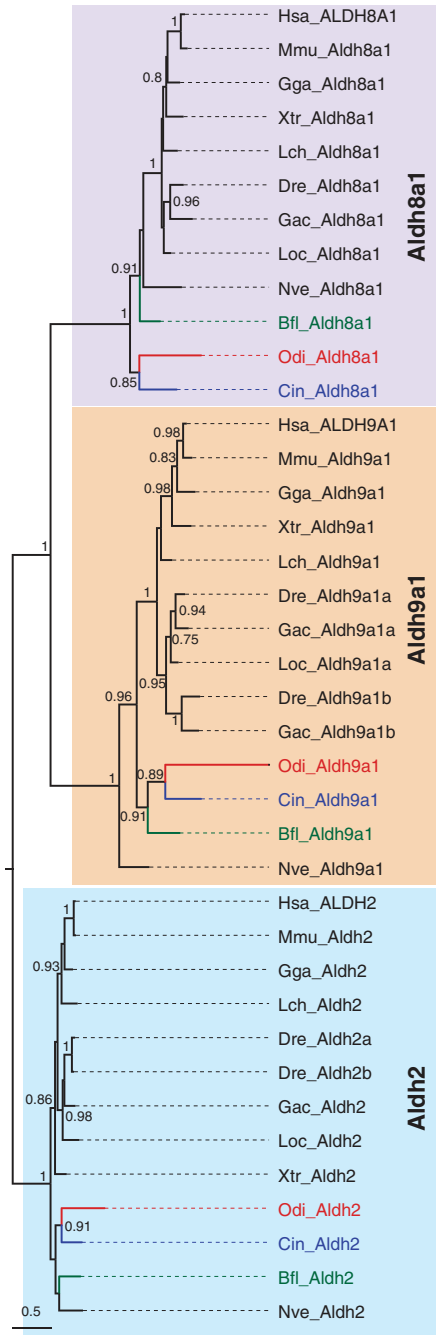


Fig. 3. ML phylogenetic tree of the Aldh8a1 family in chordates revealing the surviving of this gene to the dismantling of the RA-MGN in *Oikopleura dioica*. The sister Aldh9a1 and the Aldh2 outgroup were included to root the tree. Scale bar indicates amino acid substitutions. Values for the approximate likelihood-ratio test (aLRT) are only shown in nodes with support values greater than 0.7. Abbreviations are as in figure 2.

functional SDR domains (supplementary file S1, Supplementary Material online): TGAG$^{S}/_{N}$G$^{I}/_{L}$G at the Gly-rich sequence, D for NAD(H) specificity, and the $NX_{27}SX_{12}YX_3K$ tetrad in the active site, supporting that the *O. dioica* enzymes had the typical NAD(H)-dependent dehydrogenase activity.

Rdh16: Vertebrate *Rdh16* genes (a.k.a. *Rodh4*) belong to SDR-9C family, and have been proposed to encode enzymes that oxidize atROL into atRAL, showing therefore a redundant activity with Rdh10 and RdhE2 (Gough et al. 1998). BLAST searches using human RDH16 as query retrieved only two *O. dioica* genes (CBY10263 and CBY06797) with significant similarity. The reciprocal BLAST with the *O. dioica* sequences did not return *Rdh16* but yielded *Bdh1* (for 3-hydroxybutyrate dehydrogenase type 1) gene as the best hit (E-value = $6e^{-33}$ and $6e^{-20}$ with human *BDH1*), which is another member of the SDR-9C family that catalyzes the interconversion between acetoacetate and (R)-3-hydroxybutyrate. ML phylogenetic trees corroborated the BRBH analysis (supplementary file S3, Supplementary Material online). The presence of *Rdh16* pro-orthologs in ascidians and cephalochordates (Dalfó et al. 2001, 2007; Belyaeva and Kedishvili 2006) indicates that its absence in *O. dioica* was due again to a gene loss in the RA-MGN occurred during the evolution of the larvacean lineage.

Aldh8a1: *Aldh8a1* (for aldehyde dehydrogenase 8, member a1; a.k.a. *ALDH12* in human and *Raldh4* in zebrafish, mouse and rat) gene encodes an enzyme capable of catalyzing the conversion of atRAL into atRA (Lin and Napoli 2000; Sima et al. 2009), being therefore functionally redundant with Aldh1a enzymes. Aldh8a1 has been found in most animal phyla, from cnidarians to mammals (Albalat et al. 2011). Its amino acid sequence is highly conserved, showing five invariant residues critical for the catalytic activity (Gly231, Gly284, Cys287, Glu391, and Phe393; numbers referred to human ALDH8A1), and 11 highly conserved residues critical for the structure and function (Arg69, Gly146, Asn155, Pro157, Gly172, Lys178, Gly255, Pro395, Gly441, Asn446, and Gly459; >95% conservation in 145 Aldh sequence comparison) (Perozich et al. 1999). BLAST searches identified one *O. dioica* sequence (CBY33202), which based on BRBH analysis (E-value = 0.0 with *Xenopus tropicalis Aldh8a1*) and ML phylogenetic tree (fig. 3) appeared to be a clear *O. dioica* ortholog of *Aldh8a1*. The high sequence conservation of *O. dioica* Aldh8a1 (supplementary file S1, Supplementary Material online), including the five invariant catalytic residues and 10 out of the 11 highly conserved structural and functional residues (Gly441 is Ala in *O. dioica*), suggested that *O. dioica* Aldh8a1 might have equivalent aldehyde oxidizing activity than Aldh8a1 enzymes of other chordates.

***Cco* (*Bco1*, *Bco2*, and *Rpe65*)**: Vertebrate β-carotene-cleaving enzymes Bco1 (for beta-carotene monooxygenase 1) and Bco2 (for beta-carotene dioxygenase 2), and retinoid isomerohydrolyzing enzyme Rpe65 (for retinal pigment epithelium-specific 65 KDa protein) belong to the Cco family. The three enzymes are involved in the production of RAL, either from the cleavage of β-carotene by Bco1 or Bco2 (Wyss

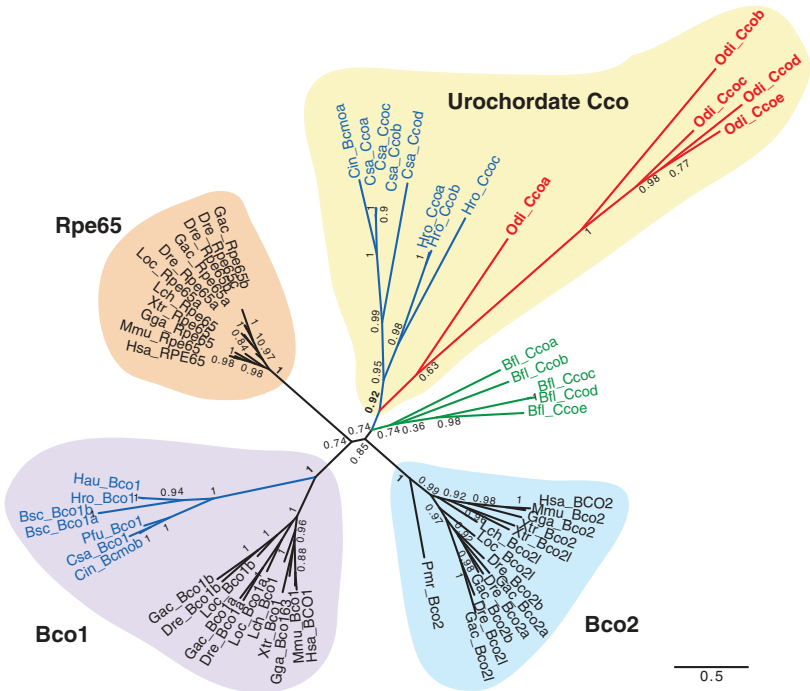


Fig. 4. ML phylogenetic tree of the Cco family (Bco1, Rpe65, and Bco2) in chordates revealing the loss of Bco1, and the surviving and lineage specific duplication of uro-Cco paralogs in *Oikopleura dioica*. Scale bar indicates amino-acid substitutions. The tree is unrooted because the absence of closely related gene family that could render a reliable sequence alignment. Values for the approximate likelihood-ratio test (aLRT) are shown in nodes. Abbreviations are as in figure 2. In addition to the Cco of the ascidian *Ciona intestinalis*, in silico survey of the genomes of five additional ascidian species in the Aniseed database (http://aniseed.cnrs.fr/, last accessed May 2016)—*Botryllus schlosseri* (Bsc), *Ciona savignyi* (Csa), *Halocynthia roretzi* (Hro), *Halocynthia aurantium* (Hau) and *Phallusia fumigata* (Pfu)—allowed us to identify and include 13 new Cco sequences in the phylogenetic analysis in order to increase the robustness of tree and to clarify the position of *O. dioica* Cco within the uro-Cco group, characterized by multiple lineage-specific duplicated paralogs in most analyzed species.

2004) or from the cleavage and isomerization of all-*trans*-retinyl esters by Rpe65 (Mata et al. 2004). Cco enzymes are characterized by four iron-coordinating His (His180, His241, His313, and His527) and three acidic residues that form a second coordination sphere essential for its enzymatic activity (Glu/Asp148, Glu/Asp417, and Glu/Asp469; numbers referred to human RPE65) (Sui et al. 2013). BLAST searches using human Ccos (BCO1, BCO2, and RPE65) and ascidians (*C. intestinalis* Bcmoa and Bcmob) identified five *O. dioica* genes (named *Ccoa* to *Ccoe*, CBY18123, CBY10633, CBY23475, CBY07484, and CBY07038) that were positive in BRBH analyses (E-values ranging from $2e^{-47}$ to $5e^{-97}$). The five *O. dioica* Cco showed strict conservation of the four iron-coordinating His and the three Glu/Asp acidic residues (table 1 and alignment in supplementary file S1, Supplementary Material online), clearly supporting *O. dioica* Ccos as potential carotenoid cleavage oxygenases. None of the *O. dioica* Ccos, however, contained any of the seven conserved residues responsible for fine-tuning/adaptation of the different Bco1/Bco2/Rpe65 enzymes (Poliakov et al. 2012) and none of the 13 residues considered relevant for Rpe65 enzymes (Albalat 2012) (table 1 and alignment in supplementary file S1, Supplementary Material online). The overall lack of conservation in these critical positions and the high sequence divergence besides of the seven conserved structural residues questioned that any of the five *O. dioica Cco* had β-carotene cleaving activity as vertebrate Bco1 or Bco2 enzymes, or an isomerohydrolyzing activity similar to vertebrate Rpe65.

The evolutionary relationship between *Cco* of vertebrate and nonvertebrate chordates was still unclear (Albalat 2009, 2012; Poliakov et al. 2012). In order to improve the robustness of the topology of Cco evolutionary tree of chordates of our analyses, we included in 13 new urochordate Cco sequences that we have identified after surveying the genomes of five ascidian species (sequences are provided in supplementary file S1, Supplementary Material online). ML phylogenetic analyses revealed a well-supported group of ascidian genes within the Bco1 clade, suggesting that duplications that gave rise to Bco1/Bco2/Rpe65 genes may have predated the urochordate-vertebrate split (fig. 4). None of the five *O. dioica* sequences, however, grouped within this ascidian/vertebrate Bco1 group, suggesting the loss of Bco1 in the *O. dioica* lineage. The five *O. dioica* Ccos grouped in a subcluster that we have named "urochordate-Cco group" (uro-Cco), which included eight Cco paralogs from three ascidian species (fig. 4). The internal tree topology of the uro-Cco group suggested

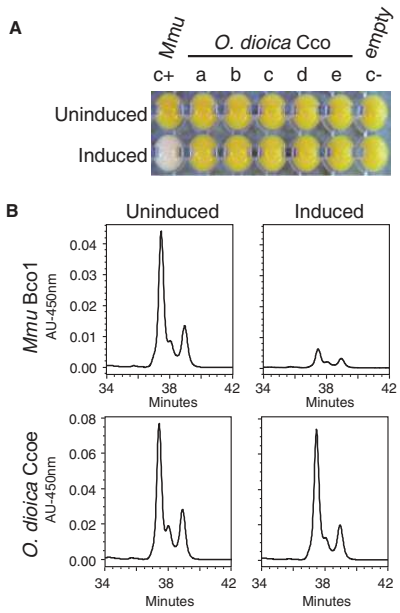


Fig. 5. *Oikopleura dioica Cco* paralogs do not show β-carotene cleaving activity. (*A*) Induction of heterologous expression of *O. dioica* Cco enzymes (Ccoa to Ccoe) in p-orange *Escherichia coli* strain did not result in a color shift from yellow to white. In contrast, the mouse Bco1 that was used as positive control cleaved accumulated β-carotene and rendered pellets with an obvious white color in comparison to the uninduced condition. (*B*) HPLC analysis of β-carotene content of p-orange *E. coli* cultures expressing mouse Bco1 (positive control, top) and *O. dioica* Ccoe (bottom). While a 9-fold reduction in the β-carotene content was observed in the *Mmu* Bco1-expressing cultures, no reduction was observed in any of the *O. dioica* Cco-expressing cultures, as represented by Ccoe-induced cultures as an example. Thus, HPLC analysis supported the observation that none of the *O. dioica* Cco cleaves β-carotene to generate atRAL.

that this group expanded due to recurrent independent duplications of *Cco* in different species within the urochordate clade. The conservation of several conserved intron positions unique to *O. dioica Ccos* was consistent with an origin by lineage-specific duplications (supplementary file S2*B*, Supplementary Material online, red arrowheads). The phylogenetic relationship between the uro-Cco group and Bco2 or Rpe65 was unclear and not well supported in the tree. The tree topology was compatible with the uro-Cco having an ancient origin, at least back to the chordate ancestor, with no surviving homolog in vertebrates, or alternatively, the uro-Cco may represent the homolog of the vertebrate Bco2 or Rpe65 families, but no clustering within those groups because of the long-branches of most urochordates Ccos (fig. 4). In either case, the five *O. dioica* sequences were lineage-specific duplications that represent the closest Cco members to the vertebrate Cco enzymes.

In conclusion, our genomic survey showed that the dismantling of RA-MGN in *O. dioica* involved the coelimination of the main genes of the canonical metabolic pathway, that was the loss of *Aldh1a*, *Cyp26*, *Rdh10*, *Rdh16*, and *Bco1* genes, while *RdhE2*, *Aldh8a1*, and some *Cco* genes survived the dismantling. Based on the biochemical properties of the vertebrate RdhE2, Aldh8a1, and Cco, the finding that these genes survived the RA-MGN dismantling was compatible with a hypothetical backup noncanonical pathway for atRA synthesis still present in *O. dioica*.

Biochemical Characterization of *O. dioica* Cco Paralogs

To test for the existence of such hypothetical backup pathway in *O. dioica*, we first investigated whether any of the *O. dioica* Ccos could catalyze β-carotene cleavage for atRAL production despite their sequence divergence. We used the experimental assay described in von Lintig and Vogt (2000), which has also been successfully used for ascidian Ccos (Poliakov et al. 2012). Each *O. dioica* Cco was heterologously expressed in the p-orange *Escherichia coli* strain that had been modified to synthesize and accumulate β-carotene, which yielded a yellow/orange color to bacterial pellets. In this assay, the yellow color shifted to white when β-carotene was cleaved by the induced expression of a β-carotene oxygenase enzyme (fig. 5*A*, compare p-orange *E. coli* expressing positive control (c+) mouse Bco1 (induced) and uninduced bacterial pellet). None of the *O. dioica* Cco did shift the color of bacterial pellets (fig. 5*A*), suggesting that these enzymes could not cleave β-carotene. To eliminate the possibility that the apparent lack of color shift in *O. dioica* Cco assays was due to a low activity of the expressed enzymes that was not evident by visual examination, we analyzed by high performance liquid chromatography (HPLC) the β-carotene content of p-orange *E. coli* pellets before and after the induction of Cco expression. Although expression of mouse Bco1 yielded a 9-fold reduction in the β-carotene content in comparison with the uninduced sample, no reduction was observed in any of the cultures expressing *O. dioica* Ccos in the assayed conditions (fig. 5*B*). These results suggest that in contrast to vertebrate Ccos none of *O. dioica* Ccos was a β-carotene-cleaving enzyme.

Retinoid and Carotenoid Content in *O. dioica*

Considering that *O. dioica* Cco did not seem to contribute to the synthesis of atRA from β-carotene cleavage, we analyzed the retinoid content of *O. dioica* by HPLC to test if RdhE2 and Aldh8a, or any other unidentified enzyme could still account for a noncanonical "backup" pathway for atRA synthesis in this species. In this analysis, we paid special attention to the presence of atRA, which is the most biologically active retinoid, but also to its intermediate metabolites—that is, atROL and atRAL—and carotenoids that could act as potential precursors.

Analyses of *O. dioica* samples from different stages, including unfertilized eggs, 7-h postfertilization (hpf) embryos, day-4 nonmature adults (at day-4, males and females were indistinguishable), and day-5 mature females and males revealed that no atRA (or 9-*cis*-, 11-*cis*-, or 13-*cis*-isomers) was detected in any of the samples (fig. 6*A*, compare the chromatograms with the "atRA standard" peak). The lower limit of reliable atRA quantification in our method was approximately 1 pmol per

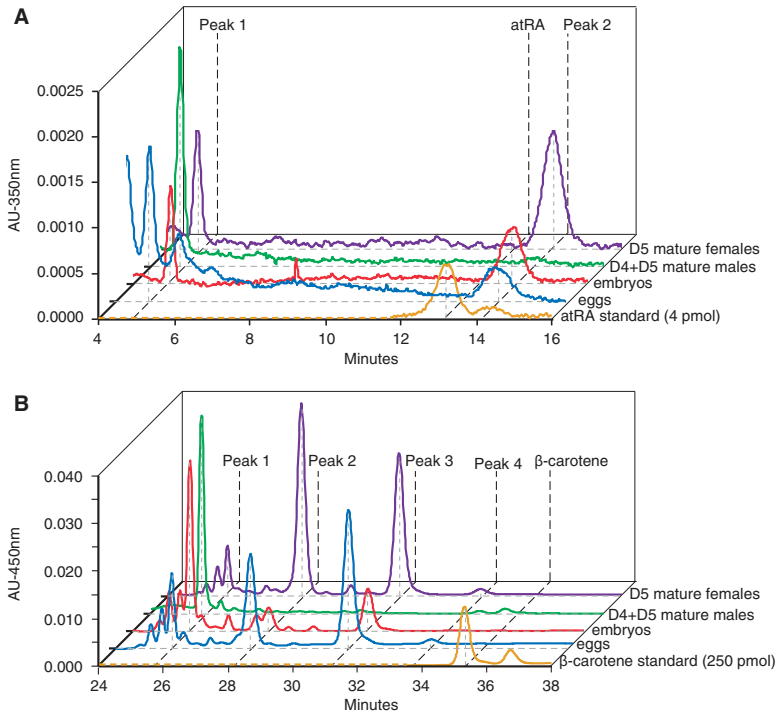


Fig. 6. HPLC analysis of the retinoid (*A*) and carotenoid (*B*) content of *Oikopleura dioica* samples from different stages, including unfertilized eggs, 7-hpf embryos, day-4 nonmature adults, and day-5 mature males and day-5 mature females. (*A*) *Oikopleura dioica* extracts analyzed in normal phase HPLC. Chromatogram extracted at 350 nm. No atRA or its 9-*cis*-, 11-*cis*-, or 13-*cis*-isomers were detected. Unidentified peak 2 eluted at 14.2 min, may represent an endogenous retinoid, while peak 1 eluted at 5.0 min, is an unrelated compound (see text and supplementary file S4, Supplementary Material online, for details). (*B*) Analysis of carotenoid content in *O. dioica* by reverse phase HPLC. Chromatograms extracted at 450 nm showed four peaks eluted at 25.8, 28.3, 31.3, and 33.8 min in most *O. dioica* samples. None of the peaks appeared to be β-carotene standard. Peaks 2,3 and 4 have absorbance spectra typical for carotenoids, while peak 1 represents a different compound. Carotenoids (peaks 2, 3, and 4) likely have a dietary origin (see supplementary file S5, Supplementary Material online, for the carotenoid content of the four microalgae species used in the *O. dioica* diet). The inverse relative abundance of peak 1 and the carotenoid peaks at different stages suggested that peak 1 might be derived from carotenoids and that *O. dioica* might have the ability to actively store carotenoids in eggs, and to metabolize them throughout their life cycle. The presence of β-carotene in the dietary algae (supplementary file S5, Supplementary Material online) suggests that the absence of atRA in *O. dioica* was not due to a dietary deficiency of β-carotene. The absence of a β-carotene in *O. dioica* samples could be explained by its transformation into astaxanthin, which appears to be the one of the major carotenoids found in larvaceans (Mojib et al. 2014).

sample. Considering that the average weight of *O. dioica* sample in this study was 81 mg, we would have been able to reliable quantify the amount of RA corresponding to 12 pmol/g of wet weight, which is comparable to the concentration of atRA reported for midgestation mouse embryos (Billings et al. 2013). Since the limit of detection is much lower than the quantification limit, and we did not detect even trace amount of atRA, we conclude that *O. dioica* did not contain atRA at concentrations that were likely to play any role in developmental or physiological processes. The enzymes encoded by the surviving *RdhE2*, *Cco*, and *Aldh8* genes might be involved, therefore, in other biological functions. Only trace amount of atROL was detected in some of the samples, indicating that vitamin A might not be abundant in *O. dioica*, in contrast to its higher concentrations observed in other nonvertebrate chordates (Dalfó et al. 2002). Retinyl esters, vitamin A storage form in vertebrates, were neither detected in any of *O. dioica* samples. Interestingly, HPLC analysis revealed the consistent presence of two peaks that may represent endogenous retinoids in *O. dioica* (supplementary fig. S4, Supplementary Material online). Peak 1 in the reverse phase chromatogram had elution time and absorbance spectrum similar to those of *cis*-RAL isomers, but it did not coelute precisely with at-, 9-*cis*-, 13-*cis*-, or 11-*cis*-RAL standards, leaving a possibility that it could be a double *cis*-isomer, or a different derivative of RAL. In normal phase chromatogram, which allowed for a better separation of *cis*-RAL isomers, this peak was not observed, though it could have been masked by the presence of much larger unrelated peak (peak 1 eluted at 5 min in fig. 6A). Peak 2 in figure 6A (corresponding to peak 2 in reverse phase chromatogram in supplementary fig. S5, Supplementary Material online) was present in all analyzed samples. Though its absorbance spectrum was also similar to RAL, the elution time was greatly delayed in comparison to

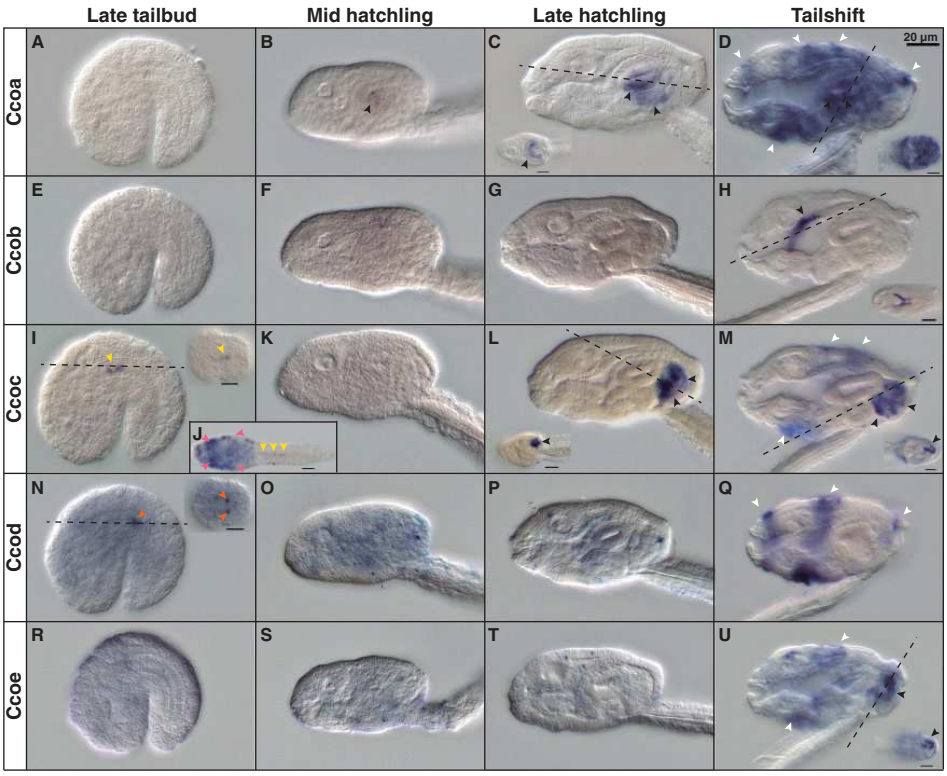


Fig. 7. Developmental expression patterns of *Oikopleura dioica Cco* paralogs. Whole-mount in situ hybridization in *O. dioica* late tailbubs (*A*, *E*, *I*, N, and *R*), midhatchlings (*B*, *F*, *K*, *O*, and *S*), late hatchlings (*C*, *G*, *L*, *P*, and *T*), and tailshift juveniles (*D*, *H*, *M*, *Q*, and *U*). *Ccoa* first expression signal was observed in the stomach primordium by midhatch stage (*B*, arrowhead), and by late-hatch stage it was strong and restricted to the right wall of the left stomach lobe, in the connection between both stomach lobes, and in the ventral part of the right stomach lobe (*C*). *Ccob* expression signal appeared as a bilateral domain in the posterior pharynx, presumptively in the peripharyngeal bands (*H* and dorsal view in inset). *Ccoc* expression signal was first observed as a faint signal in few cells near the anterior tip of the notochord (yellow arrowheads) in tailbud embryos. This notochordal domain was temporarily maintained in early hatchlings (*J*), together with some broad expression signal in the trunk, with special intensity in epidermal cells symmetrically (pink arrowheads) (*J*). *Ccoc* expression appeared to be temporarily downregulated in midhatchlings (*K*), but it became again obvious in the vertical intestine of late-hatchlings and tailshift juveniles (*L* and *M*). In tailshift juveniles, *Ccoc* expression signal appeared in different epithelial oikoplastic fields (*M*; presumptively the posterior part of the field of Fol, the middle ventral surface, the anterior crescent and the posterior rosette, white arrowheads). *Ccod* was the only paralog that did not show any clear expression domain in the digestive system, but it was temporally detected in late tailbuds a bilateral pair of cells adjacent near the seventh notochordal cell (*N* inset), and at later stages in different epithelial oikoplastic fields (*Q*). Finally, *Ccoe* expression signal appeared in the vertical intestine of tailshift juveniles (*U*) also labeled by *Ccoc*, although the onset of the former seems to be later. Large image of each panel correspond to left lateral view oriented anterior toward the left and dorsal toward the top. Inset images are dorsal views of optical cross sections at the levels of the dashed lines. Black arrowheads label expression in the digestive system, yellow in the notochord, orange in a pair of bilateral cells adjacent to the notochord, pink in the epidermis, and white in the oikoplastic epithelium. In the printed version, color codes correspond to their equivalent intensity in the grey scale. Scale bar = 20 µm.

atRAL or *cis*-RAL standards. This compound may also represent a derivative of RAL, such as hydroxylated forms described in invertebrates (Seki et al. 1987).

To test for the presence of potential dietary precursors of these endogenous compounds, we analyzed carotenoid content of *O. dioica* samples and the four algae species used in the diet to culture this organism. HPLC results revealed the presence of at least four peaks in *O. dioica* samples (fig. 6*B*), three of which had absorbance spectra typical for carotenoids (peaks 2–4), while peak 1 represented a different compound. The fact that none of the carotenoids corresponded to β-carotene (fig. 6*B*, compare the chromatograms with the "β-carotene standard" peak) suggested that this typical provitamin A precursor in other chordates was not used for that purpose in *O. dioica*, which was consistent with the loss of *Bco1* and the loss of the RA-signaling in this species. Although we did not identify the nature of the detected carotenoids in *O. dioica*, their presence in the samples of the microalgae used to feed the animals suggested that they had a dietary origin (supplementary file S5, Supplementary Material online). The relative amount of the three carotenoid peaks detected in *O. dioica* varied among the analyzed samples,

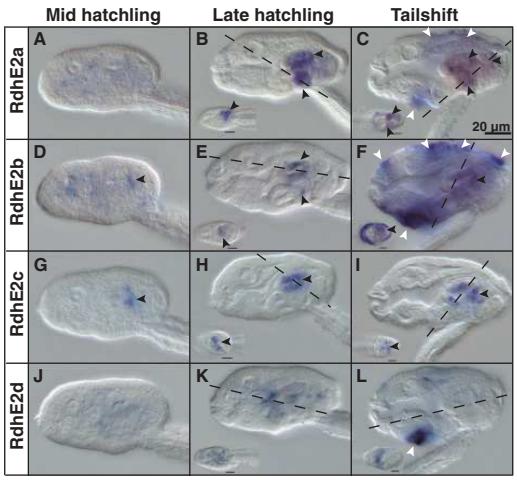


Fig. 8. Developmental expression patterns of *Oikopleura dioica RdhE2* paralogs. Whole-mount in situ hybridization in *O. dioica* mid-hatchlings (*A*, *D*, *G*, and *J*), late hatchlings (*B*, *E*, *H*, and *K*), and tailshifts (*C*, *F*, *I*, and *L*) revealed *RdhE2* expression domains in the digestive system (black arrowheads) and the oikoplastic epithelium (white arrowheads). *RdhE2a* expression signal was strong in the right stomach lobe and midintestine, and weak in the vertical intestine and the left stomach lobe (*B*, *C*). *RdhE2b* expression signal was detected in the right stomach lobe and in the midintestine (*E*, *F*). *RdhE2c* showed the earliest expression onset of all *RdhE2* paralogs in the primordium of the stomach by mid hatchling (*G*), and it was detected in both stomach lobes in later stages (*H*, *I*). *RdhE2d* was the only paralog with no clear expression in the digestive system. Similar to *Cco* paralogs, different fields of the oikoplastic epithelium appears to have also recruited the expression of *RdhE2a*, *RdhE2b*, and *RdhE2d* in different fields (e.g., the field of FoI, the middle ventral surface, the anterior crescent and the posterior rosette). Large image of each panel correspond to left lateral view oriented anterior toward the left and dorsal toward the top. Inset images are dorsal views of optical cross sections at the levels of the dashed lines. Scale bar = 20 µm.

showing their highest level in eggs and mature females full of eggs, and gradually depleting in 7-hpf embryos, day-4 immature adults, and day-5 females (fig. 6*B*). On the other hand, peak 1 showed an inverse abundance relative to the three other peaks, increasing in embryos and day-4 immature animals in comparison to eggs and mature females. *Oikopleura dioica*, therefore, appeared to have the capability to actively store carotenoids in eggs, and to metabolize these compounds throughout their life cycle. The interdependence of the relative abundance between the different peaks suggested that they could belong to the same metabolic pathway.

Expression Analysis of *O. dioica RdhE2*, *Aldh8a1*, and *Cco* Genes

Having established that the surviving *O. dioica Cco*, *RdhE2*, and *Aldh8a1* were not involved in the synthesis of atRA, we analyzed the expression of these genes in order to explore their possible biological functions during embryogenesis.

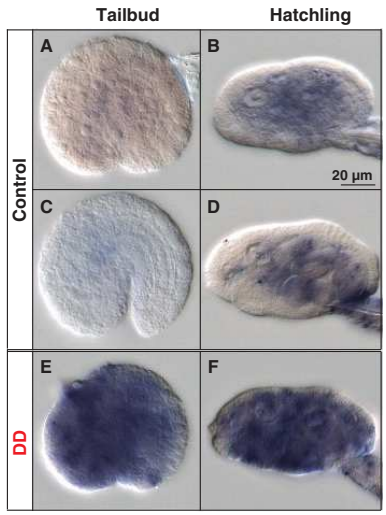


Fig. 9. Developmental expression patterns of *Oikopleura dioica Aldh8a1*. Whole-mount in situ hybridization in *O. dioica* early tailbud stage (*A* and *E*), late tailbud stage (*C*) and midhatchling (*B*) and late-hatchling (*D* and *F*). *Aldh8a1* expression signal did not show obvious tissue-specificity, but it appeared to be faintly and broadly distributed throughout the entire embryo in all analyzed developmental stages (*A*–*D*). Embryos treated with 0.25 µm/ml *trans*, *trans*-2,4-decadienal (DD), a model aldehyde for diatom-derived polyunsaturated aldehydes (PUAs), showed an obvious up-regulation of the signal throughout the embryo (*E*, *F*), suggesting a housekeeping role of the Aldh8a1 in aldehyde detoxification. Panels show left lateral views oriented anterior toward the left and dorsal toward the top. Scale bar = 20 µm.

- ***Cco.*** Most of the expression domains of the five *O. dioica Cco* genes were mainly, but not exclusively, observed in different compartments of the digestive system at late developmental stages (fig. 7*A*–*U*). *Ccoa* was the first paralog to show incipient expression in the stomach primordium by midhatchling stage (fig. 7*B*, black arrowhead). The signal of *Ccoa* became intense in the ventral part of the right stomach lobe and in the right wall of the left stomach lobe, in the region connecting both lobes by late hatchling stage (fig. 7*C*). By that stage, *Ccoc* expression signal was evident in the vertical intestine (fig. 7*L*). In tailshift juveniles, *Ccoa* expression signal became strong throughout the trunk, especially in both stomach lobes (fig. 7*D*), *Ccob* signal appeared in the posterior part of the pharynx but not in the esophagus (fig. 7*H* and inset), and *Ccoc* and *Ccoe* were expressed in the vertical intestine (fig. 7*M* and *U*, black arrowheads). Remarkably, by tailshift stage, all *Cco* genes, with the exception of *Ccob*, appeared to be also expressed in different fields of the oikoplastic epithelium (fig. 7*D*, *M*, *Q*, and *U*, white arrowheads). In summary, most of the expression domains of *Cco* genes became restricted to specific compartments of the digestive system during different developmental stages of *O. dioica*, suggesting that each *Cco* gene might have evolved distinct enzymatic activities related with a

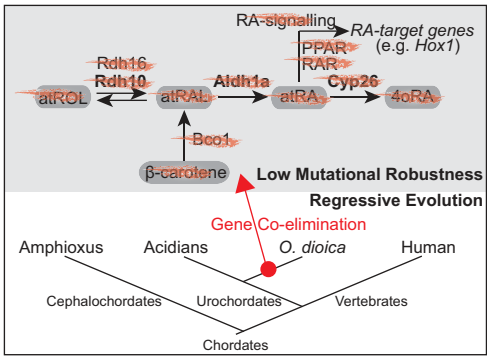


Fig. 10. Functional biased pattern of gene loss by coelimination of genes of the retinoic acid metabolic gene network (RA-MGN). The losses of the genes *Rdh10*, *Rdh16*, *Bco1*, *Aldh1a*, and *Cyp26* occurred in a context of low mutational robustness with no compensatory rerouting for the synthesis of atRA by any alternative pathway, and likely in an scenario of regressive evolution in which the loss of RA-signaling and its nuclear receptor genes (e.g., *RAR*, *PPAR*) did not imply major phenotypic changes related to RA-target genes, such as *Hox1* that show the same expression pattern in *Oikopleura dioica* and ascidians.

complex physiological compartmentalization of the digestive system of *O. dioica*.

- ***RdhE2.*** Similarly to *Ccos*, most *O. dioica RdhE2s* were expressed in the digestive system at late developmental stages (fig. 8*A*–*L*). While no obvious expression was detected before hatch, by midhatchling stage, the primordium of the stomach showed the first faint expression of *RdhE2b* and *RdhE2c* (fig. 8*D* and *G*, black arrowheads). In late-hatchlings and tail-shift juveniles, the right stomach lobe expressed three out of the four *RdhE2* genes (*RdhE2a*, *RdhE2b*, and *RdhE2c*), the midintestine expressed two of them (*RdhE2a* and *RdhE2b*), the left gastric lobe expressed *Rdh2Ec* and faintly *RdhE2a*, the vertical intestine expressed *RdhE2a*, and the dorsal part of the rectum expressed *RdhE2c*. In contrast to *Cco*, no expression was observed in the pharynx at any developmental stage. The signal of *RdhE2d* was weak and hardly distinguishable from background during all developmental stages (fig. 8*J*–*L*), with the exception of an obvious ventral domain in the oikoplastic epithelium, which also expressed *RdhE2a* and *RdhE2b*. In summary, similarly to Cco, the expression of most *O. dioica RdhE2* is restricted to distinct regions of the digestive system during different developmental stages, suggesting a functional specialization for each RdhE2 enzyme, and reinforcing the idea that the digestive system of *O. dioica* is complex and functionally highly compartmentalized.
- ***Aldh8a1.*** *Aldh8a1* expression signal, in contrast to *Ccos* and *RdhE2s*, was not tissue specific but it appeared to be faintly and broadly distributed throughout the entire embryo in all analyzed developmental stages (fig. 9*A*–*D*). This ubiquitous pattern was compatible with a housekeeping role for aldehyde detoxification, which is, indeed, the role proposed for many members of superfamily of Aldh enzymes (Vasiliou and Nebert 2005). To experimentally analyze this possibility, we checked if *O. dioica Aldh8a1* expression was affected by exposition to aldehydes. We treated *O. dioica* embryos with 0.25 µg/ml *trans,trans*-2,4-decadienal (DD), a model for polyunsaturated aldehydes produced by diatoms (Caldwell et al. 2002; Romano et al. 2010), and to which *O. dioica* might be exposed in natural environments. *Aldh8a1* expression analyzed by whole-mount in situ hybridization in DD-treated *O. dioica* embryos before and after the hatch showed an obvious *Aldh8a1* upregulation throughout the entire body of treated embryos (fig. 9*E* and *F*). These results pointed to *O. dioica Aldh8a1* acting as an enzyme member of the "chemical defensome" that responds to environmental stressors with detoxification functions against biogenic or xenobiotic aldehydes.

Discussion

Gene Coelimination Pattern of the RA-MGN Associated to the Loss of Function of RA Signaling in *O. dioica*

Our results from HPLC analyses of retinoid content combined with those from the genome survey reveal that the loss of the ability of *O. dioca* to synthesize atRA was associated with a pattern of gene coelimination affecting five genes of the RA-MGN (i.e., *Rdh10*, *Rdh16*, *Bco1*, *Aldh1a*, and *Cyp26*) (fig. 10). This coelimination pattern suggests that the enzymes encoded by these five genes function together in a distinct, and probably specialized, pathway for atRA metabolism, and therefore, they can be considered a functional module in which the loss of one gene will likely be accompanied by the loss of the others. Interestingly, this pattern of gene coelimination does not appear to be limited to the RA-MGN, but it seems to affect other components of the regulatory gene network dependent on RA-signaling, such as the loss of nuclear receptors *RAR* and *PPAR*, which mediate RA signaling by regulating the expression of RA-target genes, or the loss of *Isx* (intestine-specific homeobox), a transcription factor activated by RAR that represses the expression of Bco1 (Cañestro et al. 2006 and data not shown).

Our results showing the absence of atRA (and other intermediate retinoids) in *O. dioica* also indicate that genes surviving the dismantling of the RA-MGN (i.e., *RdhE2*, *Aldh8a1*, and *Ccos*) do not form a backup pathway for atRA synthesis, at least in *O. dioica*. The absence of a backup pathway for atRA synthesis suggests that the loss of the main RA-MGN genes in *O. dioica* did not occur in a scenario of high mutational robustness able to compensate the loss, but it was likely favored by a scenario of regressive evolution in which RA-signaling became dispensable. Several observations are consistent with the notion that RA-signaling became dispensable during the evolution of the *O. dioica* lineage, in which the loss of RA-MGN did not have a negative effect on its fitness, nor a drastic phenotypic impact. First, for instance, the fact that the expression pattern of *O. dioica* homologs of RA-target genes in other chordates (e.g., *Hox1*) has been preserved unaltered

despite the loss of RA-signaling (Cañestro and Postlethwait 2007). Second, the disintegration of the *Hox*-cluster in *O. dioica* (Seo et al. 2004) likely relaxed the constraints to maintain the RA-signaling since this signaling requires an intact cluster to be able to regulate the temporal collinear expression of its genes (Cañestro and Postlethwait 2007; Cañestro et al. 2007; Garstang and Ferrier 2013). And third, the loss of RA-signaling did not account for a drastic remodeling of the archetypal chordate body plan in *O. dioica*, as it would be expected owing its crucial in anteposterior axial patterning or organogenesis in all other chordates (Nagatomo and Fujiwara 2003; Fujiwara 2006; Cañestro and Postlethwait 2007; Koop et al. 2010; Cunningham and Duester 2015). The loss of RA-signaling of *O. dioica*, therefore, stands as a good example to illustrate gene dispensability and patterns of gene coelimination in gene networks of multicellular eukaryotes, and provides a paradigmatic case of the "inverse paradox" of Evo–Devo, which argues that organisms might develop fundamentally similar morphologies (i.e., phenotypic unity) despite having important differences in their genetic toolkits (i.e., genetic diversity), in this case due to extensive gene losses (Cañestro et al. 2007; Albalat and Cañestro 2016). We anticipate that the study of the impact of the loss of RA-signaling on the evolution of other developmental pathways such as FGF and WNT, which are known to counteract RA-signaling in other chordates (Davidson et al. 2006; Pasini et al. 2012; Wagner and Levine 2012), will be intriguing since these developmental pathways might have been remodeled upon the loss of RA-signaling. This study will likely provide new clues about the evolution of the crosstalks between these fundamental signaling pathways.

Functions of the Genes Surviving the RA-MGN Dismantling

The absence of atRA in *O. dioica* indicates that the survival of *Cco*, *RdhE2*, and *Aldh8a1* genes to the dismantling of the RA-MGN might be linked to other functions not related to RA synthesis.

Cco and *RdhE2* Gene Family Expansions Correlate with a Complex Functional Compartmentalization of the Digestive System in *O. dioica*

The distinct expression of several *O. dioica Cco* and *RdhE2* paralogs in different compartments of the digestive system (i.e., pharynx and stomach lobes) (figs. 7 and 8) suggests that the recurrent duplications affecting these gene families have allowed the evolution of a complex functional specialization of the different segments of the digestive system of *O. dioica*. Moreover, the differences in the temporal onset of the expression of *Cco* and *RdhE2* paralogs likely reflect the process of differentiation of digestive cells occurring as soon as the delineation of the first organ primordia by midhatchling stage.

The endodermal expression domains of *O. dioica Cco* paralogs resemble the expression patterns of the ascidian *C. intestinalis Bcmoa* in the gills and intestine (Takimoto et al. 2006) and *Bcmob* in the meso-endoderm of the trunk of tailbud and larval stages (Satou et al. 2001). *Bco1* and *Bco2* in zebrafish (Lampert et al. 2003; Lobo et al. 2012), rodents (Kiefer et al. 2001; Raghuvanshi et al. 2015), and humans (Lindqvist and Andersson 2004) are also expressed in the digestive system, including the pharynx, liver, and gut. Bco1 shows narrow substrate specificity for cleaving β-carotene to produce RAL (Lindqvist and Andersson 2002). Bco2, in contrast, displays broad substrate specificity and can also cleave non-provitamin A carotenoids such as xanthophylls (e.g., zeaxanthin and lutein) that can serve as blue light filters and antioxidants in lipophilic environments such as in the macula of the retina improving visual acuity and protection against light damage (Krinsky and Johnson 2005; von Lintig 2010). The apparent absence of β-carotene cleaving activity in any of the *O. dioica* Cco set (fig. 5) and their high sequence divergence, which is reflected by long phylogenetic tree branches in comparison with other ascidian Cco (fig. 4) and by high amino acid variability at important functional positions for Bco1/Bco2/Rpe65 (table 1 in supplementary file S1, Supplementary Material online), are compatible with a process of neofunctionalization of their enzymatic activities. The expansion of the Cco family in *O. dioica*, as well as in some ascidians, may have facilitated its capability of metabolizing diverse carotenoids, which are especially abundant in marine environments and display great structural diversity (Liaaen-Jensen 1991; Maoka 2011). This hypothesis is supported by the detection of several unknown carotenoids, likely from a dietary origin (supplementary file S5, Supplementary Material online), already stored in eggs, which change their relative contents throughout different stages of their life cycle (fig. 6). The physiological functions of carotenoids and their derivative products, especially in marine animals, are varied and essential, including photoprotection against UV light, immunity enhancement, defense to predation by camouflage, signaling as breeding color, and normal growth, survival, and reproduction (Torrissen and Christiansen 1995; Kawakamia et al. 1998; Chew and Park 2004; Krinsky and Johnson 2005; von Lintig 2010; Maoka 2011).

The expression patterns of the *RdhE2* paralogs in the digestive system of *O. dioica* is comparable with the ascidian expression pattern of *RdhE2* in the mesoendoderm at tailbud stage (Satou et al. 2001), and with the *RdhE2* expression of frogs, zebrafish, and human in the liver, stomach, and gut (Matsuzaka et al. 2002; Cheng et al. 2006; Belyaeva et al. 2012). The gastrointestinal expression and the sequence conservation of *O. dioica* RdhE2 enzymes suggests they might be NAD^+-dependent dehydrogenases active against substrates of a dietary origin, including marine steroids and sterols that are important chemical constituents of microalgae and a major nutritional component in the diet of marine organisms (Cardozo et al. 2007). The presence of RdhE2 orthologs in cnidarians and placozoans (Albalat et al. 2011; Belyaeva et al. 2015), which are basally divergent metazoans that appear to lack RA-signaling (Cañestro et al. 2006), and the shared endodermal expression found in *O. dioica* and all other RdhE2 analyzed so far suggest that the digestive function might be the ancestral role of the RdhE2, and its role in RA-metabolism

is an evolutionary innovation that occurred during chordate evolution (Albalat et al. 2011; Belyaeva et al. 2015).

Interestingly, none of the *O. dioica* Cco and RdhE2 appeared to be expressed in the CNS, notochord, or sensory organs, which contrasts with the strong neural expression of *C. intestinalis Bcmoa* in the photoreceptor organ of the neural complex, and *Bcmob* in the brain vesicle, including ocellus photoreceptor cells (Takimoto et al. 2007), or with the vertebrate *RdhE2* expression in the CNS and notochord (Belyaeva et al. 2012, 2015). The absence of any photoreceptor cell or pigment in the brain of *O. dioica*, in contrast to ascidians, and the lack of any atRA-metabolic role for these enzymes appear as plausible explanations accounting for the loss of neural and notochordal expression domains of *O. dioica Cco* and *RdhE2* genes in a context of regressive evolution. The survival of these genes to the RA-MGN might be therefore due to their pleiotropic condition acting in digestive functions not related to atRA metabolism, as discussed above. Future knockdown experiments might provide further clues about the physiological roles and the evolutionary process of these duplicated enzymes.

Detoxification Role of Aldh8a

The broad expression pattern of *O. dioica Aldh8a* is similar to its ascidian ortholog *Aldh8a*, but differ from its *Aldh8a* ortholog in vertebrates, which is also highly expressed in the liver, intestine, and kidney (Lin et al. 2003; Marlier and Gilbert 2004; Liang et al. 2008). The biological function of Aldh8a1 is unclear, and despite a significant activity against 9-*cis*-RAL (but low against atRAL) has been shown in vitro (Lin and Napoli 2000; Lin et al. 2003), Aldh8a1 is most active against benzaldehyde and aliphatic aldehydes, and other metabolic aldehydes (Lin and Napoli 2000). Such substrate promiscuity and the recent finding that *Aldh8a1* was one of the genes of copepods that was upregulated upon stress due to the presence of polyunsaturated aldehydes produced by blooms of diatoms (e.g., DD) (Lauritano et al. 2012) inspired us to test and discover that *O. dioica Aldh81a* also responds in the same way. This finding suggests, therefore, that *O. dioica* Aldh8a1 enzyme might play a detoxification role at least against certain polyunsaturated aldehydes similarly to other marine organisms. The presence of *Aldh8a1* orthologs already in cnidarians (Albalat et al. 2011), together with the promiscuous biochemical activity of this enzyme suggest that the detoxification role observed in *O. dioica* might represent the ancestral function, and its contribution to the RA metabolism would be a secondary evolutionary innovation.

Conclusions

This works shows how gene loss is an important evolutionary force generating differences in the developmental genetic toolkits of different organisms. The recognition of biased patterns of gene loss due to the coelimination of set of genes is a useful strategy to identify gene network modules acting in distinct pathways related to specific biological functions. Our work demonstrates the absence of mutational robustness regarding atRA synthesis in *O. dioica*, suggesting, therefore that the loss of RA-signaling in *O. dioica* likely occurred in a scenario of regressive evolution. According to such scenario, the loss of RA-signaling did not seem to have accounted for drastic morphological or functional changes in comparison to related organisms such as ascidians in which RA-signaling still is maintained, providing an example of the inverse paradox in Evo–Devo. Our work also illustrates how the identification of genes surviving the dismantling of pathways can be a useful to recognize ancestral functions and neofunctionalizations, and to reveal the pleiotropic nature of some genes by studying their roles in a simplified context in which some functions (e.g., RA-signaling) have been lost.

Materials and Methods

Biological Material

Oikopleura dioica specimens were obtained from the Mediterranean coast of Barcelona (Catalonia, Spain). Culturing of *O. dioica* and embryo collections have been performed as previously described (Marti-Solans et al. 2015).

Genome Database Searches, Phylogenetic Analyses, and Intron Comparisons

Protein sequences of the RA-MGN from vertebrate *H. sapiens* and urochordate *C. intestinalis* were used as queries in BLASTp and tBLASTn searches in *O. dioica* genome databases (https://www.genoscope.cns.fr and http://oikoarrays.biology.uiowa.edu/Oiko, last accessed May 2016). Homologies of the *O. dioica* sequences were initially assessed by BRBH strategy (Wall et al. 2003). Homologies were then corroborated by phylogenetic tree analyses based on ML inferences calculated with PhyML v3.0 and automatic mode of selection of substitution model (Guindon et al. 2010) using protein alignments generated with MUSCLE and reviewed by hand (Edgar 2004). *Ciona intestinalis* Cco sequences were also used as queries in tBlastn searches in the genome database http://www.aniseed.cnrs.fr/aniseed (last accessed May 2016) to find Cco orthologs in five additional ascidian species (i.e., *Botryllus schlosseri*, *Ciona savignyi*, *Halocynthia roretzi*, *Halocynthia aurantium*, and *Phallusia fumigata*) to provide wide species representation and to increase the robustness to the phylogenetic analysis of the Cco family. Accession numbers and protein alignments for phylogenetic tree reconstructions are provided in a supplementary file S1, Supplementary Material online.

The CIWOG program (Wilkerson et al. 2009) from GECA package (Fawal et al. 2012) was used to compare the intron/exon organization of the different orthologous genes based on position and sequence conservation in the corresponding protein alignments, with no restrictions on the proportion of identical amino acids required to define a common intron. The predicted intron/exon structure of each *O. dioica* gene was confirmed by polymerase chain reaction (PCR) amplification, cloning, and sequencing of the corresponding cDNA or by available expressed sequence tag sequences. Briefly, total RNA was obtained from *O. dioica* embryos homogenized with TRI Reagent RT (MCR, RT111) using Direct-zol™ (Zymo Research, R2050). First strand cDNA was synthetized using an oligo-d(T)-anchor primer with a T7 promoter sequence at

its 5′-end (table 2 in supplementary file S1, Supplementary Material online) and Superscript reverse transcriptase III (Life Technologies), and second strand was synthesized with *E. coli* Polymerase I (New England Biolabs). Each cDNA was amplified with gene-specific primers (table 2 in supplementary file S1, Supplementary Material online), cloned in the pCR4-TOPO vector (Invitrogen), transformed in TOP10 *E. coli* competent cells (Invitrogen), and sequenced using vector flanking primers. Cloned *O. dioica* cDNAs sequences were deposited in the Genbank under following accession numbers: Aldh8a1, KX118710; *Ccoa*, KX118711; *Ccob*, KX118712; *Ccoc*, KX118713; *Ccod*, KX118714, and *Ccoe*, KX118715.

Heterologous Expression and Biochemical Activity of *O. dioica* Ccos

To heterologously express *O. dioica Cco* genes, their cDNAs were amplified by two consecutive PCRs, first with a specific forward primer (table 2 in supplementary file S1, Supplementary Material online) and a reverse primer complementary to the T7 promoter sequence of the oligo-d(T)-anchor primer, and second with a specific pair of primers carrying restriction sites for *AscI* and *NotI* enzymes (table 2 in supplementary file S1, Supplementary Material online). PCR products were digested with *AscI*/*NotI* enzymes, cloned into pBAD-TOPO vector (Invitrogen) previously modified for containing three extra restriction sites (*AscI*, *SpeI*, and *NotI*), and transformed in a β-carotene-producing *E. coli* strain kindly provided by Dr Johannes von Lintig (von Lintig and Vogt 2000). The β-carotene-producing *E. coli* strain transformed with each *O. dioica Cco* cDNAs was inoculated in 25 ml of LB medium in 100-ml flasks, and grown at 28 °C in the dark until reached a 600 nm-absorbance of 0.8. Expression of each *Cco* was then induced by adding 0.1% w/v of L-arabinose for 20 h. The β-carotene cleaving activity of the recombinant proteins in the *E. coli* cultures was first evaluated by visual examination of the shift of the color of the bacterial pellets harvested by centrifugation—mouse Bco1 was used as positive control β-carotene cleavage *Cco* enzyme for comparison (von Lintig and Vogt 2004)—, and later confirmed by HPLC analysis of the β-carotene content in the bacterial pellets.

Analysis of Carotenoid and Retinoid Content by HPLC

Oikopleura dioica unfertilized eggs (96.1 mg, ∼35,000 eggs), 7-hpf embryos (52.3 mg, ∼31,000 embryos), day-4 nonmature adults (97.6 mg, ∼2,000 individuals), day-5 mature males (70.3 mg, ∼200 individuals), day-5 mature females (90.5 mg, ∼280 individuals), and pellets of the four algae species used in *O. dioica* diet –*Isochrysis sp* (11.1 mg), *Chaetoceros calcitrans* (22.9 mg), *Rhinomonas reticulate* (24.2 mg) and *Synechococcus* sp. (101.9 mg)—grown as described in (Marti-Solans et al. 2015)—were frozen at −80 and stored until analyses. For the analysis of *O. dioica* and algal samples, the extraction procedure described in Napoli and Horst (1998) and Belyaeva et al. (2009) was modified as follows to allow for the extraction of different classes of retinoids from the same sample: samples were homogenized on ice in 0.5 ml of PBS, mixed with 0.35 ml of methanol and extracted with 2.5 ml of hexane:dichloromethane (4:1). The upper organic phase was collected in a siliconized glass tube and dried. Ethanol (0.65 ml) containing 0.025 N potassium hydroxide was added to the aqueous phase, and the extraction was repeated with 3 ml of hexane. The upper hexane phase was pooled with the first extraction. The aqueous phase was acidified by the addition of 40 µl of 4 N hydrochloric acid, and extracted again with 3 ml of hexane. The hexane phase was collected and pooled with the previous extractions, and dried. The residue was reconstituted in 120 µl of the mobile phase and 100 µl were analyzed by normal phase HPLC. Fractions of elute containing peaks of interest were collected, pooled, dried, reconstituted in 50 µl of mobile phase, and analyzed by reverse phase HPLC. Normal and reverse phase HPLC analysis was performed essentially as described in Adams et al. (2014).

For the analysis of β-carotene content in bacterial culture, heterologous expression of *O. dioica* Cco enzymes was performed as described in the section Heterologous Expression and Biochemical Activity of *O. dioica* Ccos. Bacterial pellets were obtained from 12.5 ml of the cultures, rinsed with PBS, homogenized in 0.5 ml of PBS, mixed with 1 ml of ethanol, and extracted twice with 4 ml of hexane. Hexane extractions were pooled, dried, and analyzed by reverse phase HPLC as described in Adams et al. (2014).

Gene Expression Analysis by Whole-Mount In Situ Hybridization

Fragments of *O. dioica* genes were PCR amplified and cloned to synthesize gene-specific riboprobes (table 2 in supplementary file S1, Supplementary Material online). Whole-mount in situ hybridization on fixed embryos was performed as previously described (Bassham and Postlethwait 2000; Cañestro and Postlethwait 2007) with minor modifications. To promote permeabilization, embryos were incubated with 1% DMSO and 0,2% Tween-20 in PBS for 30 min before prehybridization, and washed with a PBT solution with 0.2% Tween-20 in the posthybridization.

Pharmacological Treatments

Avoiding light exposure, 5 µl of *trans,trans*-2,4-decadienal (DD; Sigma-Aldrich W313505) were diluted in 200 µl of DMSO (Sigma-Aldrich D8418) to prepare a stock solution of 15 mg/ml inferred from optical density at 282 nm following the Beer-Lambert law, and it was stored at 4 °C until used. Working solution of DD (0.03 mg/ml) was made by diluting the stock 1:500 in 0.2 µM-filtered sterilized seawater (sSW). Eggs were preincubated for 10 min before fertilization in 0.25 µg/ml of DD in a volume 900 µl of sSW (50 µl of DD-working solution in 6 ml of sSW). Ten minutes after fertilization, excess of sperm was washed by transferring the zygotes into a new 5.1 ml of 0.25 µg/ml DD solution in sSW in 50 mm-diameter glass dishes. Control embryos were incubated in 0.0017% DMSO solution. Embryos were maintained at 19 °C for 3 h and 20 min postfertilization (until late tailbud stage) or for 9 h 45 min postfertilization (until late hatchling stage) and fixed for whole-mount in situ hybridization analysis.

Supplementary Material

Supplementary files S1–S5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors are grateful to Johannes von Lintig for kindly sharing the β-carotene-producing *E. coli* strain and mouse Bco1 construct, and to D. Buj and Professor D. Chourrout for its technical assistance and scientific advise at the beginning of the project. The authors also thank the team members of the lab for thoughtful discussions on gene loss. They also thank to F. Palau and the staff of the Club Nàutic de Coma-ruga, and to G. Muñoz and the Badalona City Hall for facilitating us the access for animal collections. To J. Guinea and the staff of the Scientific and Technological Centers (CCiT) at the Facultat de Biologia of the Universitat de Barcelona for providing seawater. This work was supported by National Institutes of Health (grant AA12153), Ministerio de Ciencia e Innovación (grant number BFU2010-14875), and by Generalitat de Catalunya (grant number SGR2014-290).

References

Adams MK, Belyaeva OV, Wu L, Kedishvili NY. 2014. The retinaldehyde reductase activity of DHRS3 is reciprocally activated by retinol dehydrogenase 10 to control retinoid homeostasis. *J Biol Chem*. 289:14868–14880.

Albalat R. 2009. The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations. *Mol Cell Endocrinol*. 313:23–35.

Albalat R. 2012. Evolution of the genetic machinery of the visual cycle: a novelty of the vertebrate eye? *Mol Biol Evol*. 29:1461–1469.

Albalat R, Brunet F, Laudet V, Schubert M. 2011. Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective. *Genome Biol Evol*. 3:985–1005.

Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat Rev Genet*. 17:379–391.

Albalat R, Gonzàlez D, Atrian S. 1992. Protein engineering of *Drosophila* alcohol dehydrogenase. The hydroxyl group of Tyr152 is involved in the active site of the enzyme. *FEBS Lett*. 308:235–239.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A*. 97:11319–11324.

Bassham S, Postlethwait J. 2000. Brachyury (T) expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of T. *Dev Biol*. 220:322–332.

Belyaeva OV, Chang C, Berlett MC, Kedishvili NY. 2015. Evolutionary origins of retinoid active short-chain dehydrogenases/reductases of SDR16C family. *Chem Biol Interact*. 234:135–143.

Belyaeva OV, Johnson MP, Kedishvili NY. 2008. Kinetic analysis of human enzyme RDH10 defines the characteristics of a physiologically relevant retinol dehydrogenase. *J Biol Chem*. 283:20299–20308.

Belyaeva OV, Kedishvili NY. 2006. Comparative genomic and phylogenetic analysis of short-chain dehydrogenases/reductases with dual retinol/sterol substrate specificity. *Genomics* 88:820–830.

Belyaeva OV, Lee SA, Adams MK, Chang C, Kedishvili NY. 2012. Short chain dehydrogenase/reductase rdhe2 is a novel retinol dehydrogenase essential for frog embryonic development. *J Biol Chem*. 287:9061–9071.

Belyaeva OV, Lee SA, Kolupaev OV, Kedishvili NY. 2009. Identification and characterization of retinoid-active short-chain dehydrogenases/reductases in *Drosophila melanogaster*. *Biochim Biophys Acta*. 1790:1266–1273.

Billings SE, Pierzchalski K, Butler Tjaden NE, Pang XY, Trainor PA, Kane MA, Moise AR. 2013. The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. *FASEB J*. 27:4877–4889.

Caldwell GS, Olive PJ, Bentley MG. 2002. Inhibition of embryonic development and fertilization in broadcast spawning marine invertebrates by water soluble diatom extracts and the diatom toxin 2-trans,4-trans decadienal. *Aquat Toxicol*. 60:123–137.

Cammas L, Romand R, Fraulob V, Mura C, Dolle P. 2007. Expression of the murine retinol dehydrogenase 10 (Rdh10) gene correlates with many sites of retinoid signalling during embryogenesis and organ differentiation. *Dev Dyn*. 236:2899–2908.

Cañestro C, Albalat R, Postlethwait JH. 2010. *Oikopleura dioica* alcohol dehydrogenase class 3 provides new insights into the evolution of retinoic acid synthesis in chordates. *Zool Sci*. 27:128–133.

Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH. 2009. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet*. 5:e1000496.

Cañestro C, Godoy L, Gonzàlez-Duarte R Albalat R.. 2003 Comparative expression analysis of Adh3 during arthropod, urochordate, cephalochordate and vertebrate development challenges its predicted housekeeping role. *Evol Dev*. 5:157–162.

Cañestro C, Hjelmqvist L, Albalat R, Garcia-Fernàndez J, Gonzàlez-Duarte R, Jörnvall H. 2000. Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. *Eur J Biochem*. 267:6511–6518.

Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol*. 305:522–538.

Cañestro C, Postlethwait JH, Gonzàlez-Duarte R, Albalat R. 2006 Is retinoic acid genetic machinery a chordate innovation? *Evol Dev*. 8:394–406.

Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. *Nat Rev Genet*. 8:932–942.

Cardozo KH, Guaratini T, Barros MP, Falcao VR, Tonon AP, Lopes NP, Campos S, Torres MA, Souza AO, Colepicolo P, et al. 2007. Metabolites from algae with economical impact. *Comp Biochem Physiol C Toxicol Pharmacol*. 146:60–78.

Chen H, Namkung MJ, Juchau MR. 1995. Biotransformation of all-trans-retinol and all-trans-retinal to all-trans-retinoic acid in rat conceptal homogenates. *Biochem Pharmacol*. 50:1257–1264.

Cheng W, Guo L, Zhang Z, Soo HM, Wen C, Wu W, Peng J. 2006. HNF factors form a network to regulate liver-enriched genes in zebrafish. *Dev Biol*. 294:482–496.

Chew BP, Park JS. 2004. Carotenoid action on the immune response. *J Nutr*. 134:257S–261S.

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

Dalfó D, Albalat R, Molotkov A, Duester G, Gonzàlez-Duarte R. 2002. Retinoic acid synthesis in the prevertebrate amphioxus involves retinol oxidation. *Dev Genes Evol*. 212:388–393.

Dalfó D, Cañestro C, Albalat R, Gonzàlez-Duarte R. 2001. Characterization of a microsomal retinol dehydrogenase gene from amphioxus: retinoid metabolism before vertebrates. *Chem Biol Interact*. 130-132:359–370.

Dalfó D, Marqués N, Albalat R. 2007. Analysis of the NADH-dependent retinaldehyde reductase activity of amphioxus retinol dehydrogenase enzymes enhances our understanding of the evolution of the retinol dehydrogenase family. *FEBS J*. 274:3739–3752.

Davidson B, Shi W, Beh J, Christiaen L, Levine M. 2006. FGF signaling delineates the cardiac progenitor field in the simple chordate, *Ciona intestinalis*. *Genes Dev*. 20:2728–2738.

Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Cañestro C, et al. 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330:1381–1385.

Duester G. 2013. Retinoid signaling in control of progenitor cell differentiation during mouse development. *Semin Cell Dev Biol*. 24:694–700.

Duester G, Mic FA, Molotkov A. 2003. Cytosolic retinoid dehydrogenases govern ubiquitous metabolism of retinol to retinaldehyde followed by tissue-specific metabolism to retinoic acid. *Chem Biol Interact*. 143-144:201–210.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Fawal N, Savelli B, Dunand C, Mathe C. 2012. GECA: a fast tool for gene evolution and conservation analysis in eukaryotic protein families. *Bioinformatics* 28:1398–1399.

Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U. 2002. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem*. 277:25677–25684.

Fujiwara S. 2006. Retinoids and nonvertebrate chordate development. *J Neurobiol*. 66:645–652.

Garstang M, Ferrier DE. 2013. Time is of the essence for ParaHox homeobox gene clustering. *BMC Biol*. 11:72.

Gough WH, VanOoteghem S, Sint T, Kedishvili NY. 1998. cDNA cloning and characterization of a new human microsomal NAD+-dependent dehydrogenase that oxidizes all-trans-retinol and 3alpha-hydroxysteroids. *J Biol Chem*. 273:19778–19785.

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421:63–66.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.

Jurukovski V, Markova NG, Karaman-Jurukovska N, Randolph RK, Su J, Napoli JL, Simon M. 1999. Cloning and characterization of retinol dehydrogenase transcripts expressed in human epidermal keratinocytes. *Mol Genet Metab*. 67:62–73.

Kawakamia T, Tsushimab M, Katabamia Y, Mine M, Ishida A, Matsuno M. 1998. Effect of b,b-carotene, b-echinenone, astaxanthin, fucoxanthin, vitamin A and vitamin E on the biological defense of the sea urchin *Pseudocentrotus depressus*. *J Exp Mar Biol Ecol*. 226:165–174.

Kiefer C, Hessel S, Lampert JM, Vogt K, Lederer MO, Breithaupt DE, von Lintig J. 2001. Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A. *J Biol Chem*. 276:14110–14116.

Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol*. 5:R7.

Koop D, Holland ND, Semon M, Alvarez S, de Lera AR, Laudet V, Holland LZ, Schubert M. 2010. Retinoic acid signaling targets Hox genes during the amphioxus gastrula stage: insights into early anterior-posterior patterning of the chordate body plan. *Dev Biol*. 338:98–106.

Krinsky NI, Johnson EJ. 2005. Carotenoid actions and their relation to health and disease. *Mol Aspects Med*. 26:459–516.

Lampert JM, Holzschuh J, Hessel S, Driever W, Vogt K, von Lintig J. 2003. Provitamin A conversion to retinal via the beta,beta-carotene-15,15'-oxygenase (bcox) is essential for pattern formation and differentiation during zebrafish embryogenesis. *Development* 130:2173–2186.

Lauritano C, Carotenuto Y, Miralto A, Procaccini G, Ianora A. 2012. Copepod population-specific response to a toxic diatom diet. *PLoS One* 7:e47262.

Lee SA, Belyaeva OV, Kedishvili NY. 2009. Biochemical characterization of human epidermal retinol dehydrogenase 2. *Chem Biol Interact*. 178:182–187.

Lesk AM. 1995. NAD-binding domains of dehydrogenases. *Curr Opin Struct Biol*. 5:775–783.

Liaaen-Jensen S. 1991. Marine carotenoids: recent progress. *Pure Appl Chem*. 63:1–12.

Liang D, Zhang M, Bao J, Zhang L, Xu X, Gao X, Zhao Q. 2008. Expressions of Raldh3 and Raldh4 during zebrafish early development. *Gene Expr Patterns*. 8:248–253.

Lin M, Napoli JL. 2000. cDNA cloning and expression of a human aldehyde dehydrogenase (ALDH) active with 9-cis-retinal and identification of a rat ortholog, ALDH12. *J Biol Chem*. 275: 40106–40112.

Lin M, Zhang M, Abraham M, Smith SM, Napoli JL. 2003. Mouse retinal dehydrogenase 4 (RALDH4), molecular cloning, cellular expression, and activity in 9-cis-retinoic acid biosynthesis in intact cells. *J Biol Chem*. 278:9856–9861.

Lindqvist A, Andersson S. 2002. Biochemical properties of purified recombinant human beta-carotene 15,15'-monooxygenase. *J Biol Chem*. 277:23942–23948.

Lindqvist A, Andersson S. 2004. Cell type-specific expression of beta-carotene 15,15'-mono-oxygenase in human tissues. *J Histochem Cytochem*. 52:491–499.

Lobo GP, Isken A, Hoff S, Babino D, von Lintig J. 2012. BCDO2 acts as a carotenoid scavenger and gatekeeper for the mitochondrial apoptotic pathway. *Development* 139:2966–2977.

Maoka T. 2011. Carotenoids in marine animals. *Mar Drugs*. 9:278–293.

Marlier A, Gilbert T. 2004. Expression of retinoic acid-synthesizing and -metabolizing enzymes during nephrogenesis in the rat. *Gene Expr Patterns*. 5:179–185.

Martí-Solans J, Ferrández-Roldán A, Godoy-Marín H, Badia-Ramentol J, Torres-Águila NP, Rodríguez-Marí A, Bouquet JM, Chourrout D, Thompson EM, Albalat R, et al. 2015. *Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* 53:183–193.

Mata NL, Moghrabi WN, Lee JS, Bui TV, Radu RA, Horwitz J, Travis GH. 2004. Rpe65 is a retinyl ester binding protein that presents insoluble substrate to the isomerase in retinal pigment epithelial cells. *J Biol Chem*. 279:635–643.

Matsuzaka Y, Okamoto K, Tsuji H, Mabuchi T, Ozawa A, Tamiya G, Inoko H. 2002. Identification of the hRDH-E2 gene, a novel member of the SDR family, and its increased expression in psoriatic lesion. *Biochem Biophys Res Commun*. 297:1171–1180.

Mojib N, Amad M, Thimma M, Aldanondo N, Kumaran M, Irigoien X. 2014. Carotenoid metabolic profiling and transcriptome-genome mining reveal functional equivalence among blue-pigmented copepods and appendicularia. *Mol Ecol*. 23:2740–2756.

Molotkov A, Deltour L, Foglio MH, Cuenca AE, Duester G. 2002. Distinct retinoid metabolic functions for alcohol dehydrogenase genes Adh1 and Adh4 in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. *J Biol Chem*. 277:13804–13811.

Molotkov A, Fan X, Deltour L, Foglio MH, Martras S, Farres J, Pares X, Duester G. 2002. Stimulation of retinoic acid production and growth by ubiquitously expressed alcohol dehydrogenase Adh3. *Proc Natl Acad Sci U S A*. 99:5337–5342.

Nagatomo K, Fujiwara S. 2003. Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated *Ciona intestinalis* embryos. *Gene Expr Patterns*. 3:273–277.

Napoli JL, Horst RL. 1998. Quantitative analyses of naturally occurring retinoids. In: Redfern, editor. Methods Mol Biol. Totowa, NJ: Humana Press. 29–40.

Niederreither K, Abu-Abed S, Schuhbaur B, Petkovich M, Chambon P, Dolle P. 2002. Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development. *Nat Genet*. 31:84–88.

Niederreither K, Subbarayan V, Dolle P, Chambon P. 1999. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet*. 21:444–448.

Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet*. 64:18–23.

Olson MV, Varki A. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat Rev Genet*. 4:20–28.

Pasini A, Manenti R, Rothbacher U, Lemaire P. 2012. Antagonizing retinoic acid and FGF/MAPK pathways control posterior body patterning in the invertebrate chordate *Ciona intestinalis*. *PLoS One* 7:e46193.

Perozich J, Nicholas H, Wang BC, Lindahl R, Hempel J. 1999. Relationships within the aldehyde dehydrogenase extended family. *Protein Sci*. 8:137–146.

Poliakov E, Gubin AN, Stearn O, Li Y, Campos MM, Gentleman S, Rogozin IB, Redmond TM. 2012. Origin and evolution of retinoid isomerization machinery in vertebrate visual cycle: hint from jawless vertebrates. *PLoS One* 7:e49975.

Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*. 38:107–111.

Raghuvanshi S, Reed V, Blaner WS, Harrison EH. 2015. Cellular localization of beta-carotene 15,15′ oxygenase-1 (BCO1) and beta-carotene 9′,10′ oxygenase-2 (BCO2) in rat liver and intestine. *Arch Biochem Biophys*. 572:19–27.

Reijntjes S, Blentic A, Gale E, Maden M. 2005. The control of morphogen signalling: regulation of the synthesis and catabolism of retinoic acid in the developing embryo. *Dev Biol*. 285:224–237.

Rhinn M, Dolle P. 2012. Retinoic acid signalling during development. *Development* 139:843–858.

Romano G, Miralto A, Ianora A. 2010. Teratogenic effects of diatom metabolites on sea urchin *Paracentrotus lividus* embryos. *Mar Drugs*. 8:950–967.

Sandell LL, Sanderson BW, Moiseyev G, Johnson T, Mushegian A, Young K, Rey JP, Ma JX, Staehling-Hampton K, Trainor PA. 2007. RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. *Genes Dev* 21:1113–1124.

Satou Y, Takatori N, Yamada L, Mochizuki Y, Hamaguchi M, Ishikawa H, Chiba S, Imai K, Kano S, Murakami SD, et al. 2001. Gene expression profiles in *Ciona intestinalis* tailbud embryos. *Development* 128:2893–2904.

Schilling TF, Nie Q, Lander AD. 2012. Dynamics and precision in retinoic acid morphogen gradients. *Curr Opin Genet Dev*. 22:562–569.

Seki T, Fujishita S, Ito M, Matsuoka N, Tsukida K. 1987. Retinoid composition in the compound eyes of insects. *Exp Biol*. 47:95–103.

Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, et al. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* 431:67–71.

Sima A, Parisotto M, Mader S, Bhat PV. 2009. Kinetic characterization of recombinant mouse retinal dehydrogenase types 3 and 4 for retinal substrates. *Biochim Biophys Acta*. 1790:1660–1664.

Strate I, Min TH, Iliev D, Pera EM. 2009. Retinol dehydrogenase 10 is a feedback regulator of retinoic acid signalling during axis formation and patterning of the central nervous system. *Development* 136:461–472.

Sui X, Kiser PD, Lintig J, Palczewski K. 2013. Structural basis of carotenoid cleavage: from bacteria to mammals. *Arch Biochem Biophys*. 539:203–213.

Takimoto N, Kusakabe T, Horie T, Miyamoto Y, Tsuda M. 2006. Origin of the Vertebrate Visual Cycle: III. Distinct Distribution of RPE65 and b-carotene 15,15′-Monooxygenase Homologues in *Ciona intestinalis*. *Photochem Photobiol*. 82:1468–1474.

Takimoto N, Kusakabe T, Tsuda M.. 2007 Origin of the vertebrate visual cycle. *Photochem Photobiol*. 83:242–247.

Torrissen O, Christiansen R. 1995. Requirements for carotenoids in fish diets. *J Appl Ichthyol*. 11:225–230.

Vasiliou V, Nebert DW. 2005. Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. *Hum Genomics*. 2:138–143.

von Lintig J. 2010. Colors with functions: elucidating the biochemical and molecular basis of carotenoid metabolism. *Annu Rev Nutr*. 30:35–56.

von Lintig J, Vogt K. 2000. Filling the gap in vitamin A research. Molecular identification of an enzyme cleaving b-carotene to retinal. *J Biol Chem*. 275:11915–11920.

von Lintig J, Vogt K. 2004. Vitamin A formation in animals: molecular identification and functional characterization of carotene cleaving enzymes. *J Nutr*. 134:251S–256S.

Wagner A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27:176–188.

Wagner E, Levine M. 2012. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* 139:2351–2359.

Wall DP, Fraser HB, Hirsh AE. 2003. Detecting putative orthologs. *Bioinformatics* 19:1710–1711.

Wang XD, Russell RM, Liu C, Stickel F, Smith DE, Krinsky NI. 1996. Beta-oxidation in rabbit liver in vitro and in the perfused ferret liver contributes to retinoic acid biosynthesis from beta-apocarotenoic acids. *J Biol Chem*. 271:26490–26498.

Wierenga RK, Terpstra P, Hol WG. 1986. Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol*. 187:101–107.

Wilkerson MD, Ru Y, Brendel VP. 2009. Common introns within orthologous genes: software and application to plants. *Brief Bioinform*. 10:631–644.

Wyss A. 2004. Carotene oxygenases: a new family of double bond cleavage enzymes. *J Nutr*. 134:246S–250S.

Table 1. Residues at functional positions conserved in Cco enzymes[a] (in bold), at positions that diverged between Bco and Rpe65 enzymes[b] (in italic), and at positions considered relevant for Rpe65 enzymes[c] (in bold italic).

Position[d]	**180**	**241**	**313**	**527**	**148**	**417**	**469**	*49*	*64*	*92*	*332*	*415*	*437*	*451*	***91***	***95***	***127***	***132***	***231***	***288***	***321***	***331***	***341***	***368***	***434***	***450***	***457***
Hsa_RPE65	H	H	H	H	E	E	E	L	Q	A	K	A	L	N	R	E	E	A	C	W	N	W	L	Y	A	L	T
Dre_Rpe65a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	S	.	S	.	.	.	.	.	.
Dre_Rpe65b	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	S	.	Q	.	.	.	.	.	.
Dre_Rpe65c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	S	.	Q	.	.	.	.	.	.
Hsa_BCO1	.	.	.	.	.	.	.	M	L	N	E	G	T	D	T	A	D	C	S	Y	D	Y	F	F	V	Y	S
Dre_Bco1a	.	.	.	.	.	.	.	.	M	N	.	G	T	D	K	A	D	N	I	V	D	Y	F	F	F	V	H
Dre_Bco1b	.	.	.	.	.	.	.	M	M	N	D	G	C	D	S	A	D	C	.	L	D	Y	F	F	S	F	Q
Hsa_BCO2	.	.	.	.	.	.	.	K	M	N	D	G	C	D	A	K	A	T	S	R	Q	Q	Y	F	F	V	L
Dre_Bco2a	.	.	.	.	.	.	.	K	M	N	D	G	C	D	E	H	K	.	S	V	.	S	Y	F	F	M	I
Dre_Bco2b	.	.	.	.	.	.	.	K	M	C	D	G	C	D	K	R	K	.	P	I	D	S	F	F	Y	M	L
Dre_Bco2l	.	.	.	.	.	.	.	K	M	N	D	G	C	D	Q	K	K	.	A	I	.	G	F	.	F	M	L
Cin_Bcmoa	.	.	.	.	D	.	.	E	H	N	.	G	I	.	T	N	R	.	S	S	K	S	G	I	I	I	I
Cin_Bcmob	.	.	.	.	.	.	Q	V	M	N	.	G	P	D	K	A	.	C	S	R	G	Y	F	I	F	I	R
Odi_Ccoa	.	.	.	.	.	.	.	Q	H	N	N	G	N	.	K	F	Q	C	A	S	E	P	.	F	L	V	M
Odi_Ccob	.	.	.	.	D	D	.	K	S	N	N	G	A	.	T	A	S	T	S	R	S	A	.	F	Y	V	R
Odi_Ccoc	.	.	.	.	.	D	.	K	S	N	G	G	V	Q	A	.	S	T	N	Y	E	T	F	.	F	S	R
Odi_Ccod	.	.	.	.	.	D	.	K	S	N	-	G	.	K	L	.	F	T	N	F	-	-	.	.	F	S	R
Odi_Ccoe	.	.	.	.	.	D	.	Q	S	N	P	G	V	Q	A	K	F	T	N	Y	E	T	Y	.	F	S	R

[a]From (Poliakov, et al. 2005) and (Sui, et al. 2013).

[b]From (Poliakov, et al. 2012). From (Albalat 2012).

[c]

[d]Numbering refers to human RPE65. Dots indicate identity to human RPE65 and dashes denote gaps. Hsa, *H. sapiens*; Dre, *D. rerio*; Cin, *C. intestinalis*; Odi. *O. dioica*

Table 2. Information about the cloning of fragments for riboprobes and heterologous expression.

Gene	Forward	Reverse	Use: source; length; vector; RNA-pol/digestion enzyme
Aldh8a1	5'..GATTTAAACAAAAAATGGAGCCGATTG..3'	5'..TTTATTGCTTGCTAATTGTATTAGCTTGAAG..3'	In situ probe: cDNA exon 4 to 3'UTR; 746 bp; pCR4 – TOPO; T3/XhoI
RdhE2a	5'..GCGCCAGATGAGAGAAAGTGC..3'	5'..TAACTCCAGCCCAAGAAAGGAAAGTC..3'	In situ probe: cDNA 5'UTR to exon 6; 890 bp; pCRII-TOPO; Sp6/EcoRV
RdhE2b	5'..GAATGGCGAAATCGTCTGCATAAC..3'	5'..CGTGATTCTTTTCAACCATCTTCGG..3'	In situ probe: cDNA exon3 to exon 4; 391 bp; pCR4-TOPO; T7/PstI
RdhE2d	5'..GATTCAAAGAAAAATGGATTTTTCGG..3'	5'..ACCAAACCAAGACAAGAACGTCATC..3'	In situ probe: cDNA 5'UTR to exon 3; 873 bp; pCRII-TOPO; Sp6/EcoRV
RdhE2c	5'..TTGCTGAAAGGAGAAGTTGCAGTTG..3'	5'..GCCCCATAAGCAAAGATGTAAGnAAC..3'	In situ probe: cDNA exon1 to exon 5; 765 bp; pCRII-TOPO; Sp6/EcoRV
OdCcoa [a]	5'..GAAAATGAGCTCGCACGGTTCG..3'	5'..TCAGAAATCAGAGGGCAAGAAGG..3'	In situ probe: genomic from exon 8 to exon 9; 612 bp; pCR4 – TOPO; T3/NotI
	5'..GGGCGCGCCGAGTAACAGGGTTACACAGAG..3'	5'..GTATGTATGCGGCCGCTCAGAAATCAGAGGGCAAGAAGG..3'	Heterologous expression: full length cDNA 1740 bp, pBAD-ASN
OdCcob	5'..CGACGGGATTCTCTACCGAAATG..3'	5'..AGCGATCGAGTGAAAATATGGCAT..3'	In situ probe: genomic exon 2; 733 bp; pCR4 – TOPO; T3/NotI
	5'..GGGCGCGCCAAATTTGAGATTTTTTGTTCTCTTCC..3'	5'..GTATGTATGCGGCCGCTCAATTAAGTTCTTCTGTTTTGTAAAA..3'	Heterologous expression: full length cDNA 1791 bp, pBAD-ASN
OdCcoc	5'..CTGTCTACGTGGTGATAAAATTCA..3'	5'..CCGCATTTCCAGTTTGTAACGT..3'	In situ probe: genomic exon 3; 432 bp; pCR4 – TOPO; T3/NotI
	5'..GGGCGCGCCATTTCTCTTTTTTATCGCTGCAATTTC..3'	5'..GTATGTATGCGGCCGCTTATAAATGTGGAATAAAAACACCGTG	Heterologous expression: full length cDNA 1701 bp, pBAD-ASN
OdCcod	5'..GCTGCAAACATCTTCTCAATTCCTT..3'	5'..ATCATAAAATCCTTGAATCGGATGC..3'	In situ probe: genomic exon 2; 656 bp; pCR4 – TOPO; T7/PstI
	5'..GGGCGCGCCTTTTCTTCTTTTGATTTTCGCCGGC..3'	5'..GTATGTATGCGGCCGCGCAGATTCTCGAAATAGAAATTGTCAA..3'	Heterologous expression: full length cDNA 1796 bp [c], pBAD-ASN
OdCcoe	5'..AAAGTTTTCAATTTTTTCTTTTCTCTTGCG..3'	5'..CATCGATAACTCCTCGTCTTCC..3'	In situ probe: cDNA exon1 to exon 2; 399 bp; pCR4 – TOPO; T3/NotI
	5'..GGGCGCGCCAGTTTTCAATTTTTTCTTTTCTCTTGCG..3'	5'..GTATGTATGCGGCCGCTTATCAGTGTTTTACGAAAGTTCCATG..3'	Heterologous expression: full length cDNA 1689 bp, pBAD-ASN
oligo-d(T)-anchor primer [b]	5'..AAGCAGTGGTATCAACGCAGAGT**ACGGCCAGTGAATTGTAATACGACTCACTATAGGG**AGGCGGTTTTTTTTTTTTTTTTTTTTTTTVN..3'		1st strand cDNA synthesis

[a] Underlined nucleotides correspond to the restriction sites used to clone the sequences.

[b] Bold nucleotides correspond to T7 promotor sequence.

[b] Despite the annotation of *Ccod* in the assembly of the Oikobase consists of 5 exons hypothetically coding for a full length cds, all our cDNA clones (n=3) as well as ESTs showed the retention of intron 2, resulting in a premature stop codon that generates a truncated protein at the C-terminal part (Genebank: KX118714).

Fig S1. Protein sequence alignments of the species used for the phylogenetic inferences of this work can be found at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4989114/

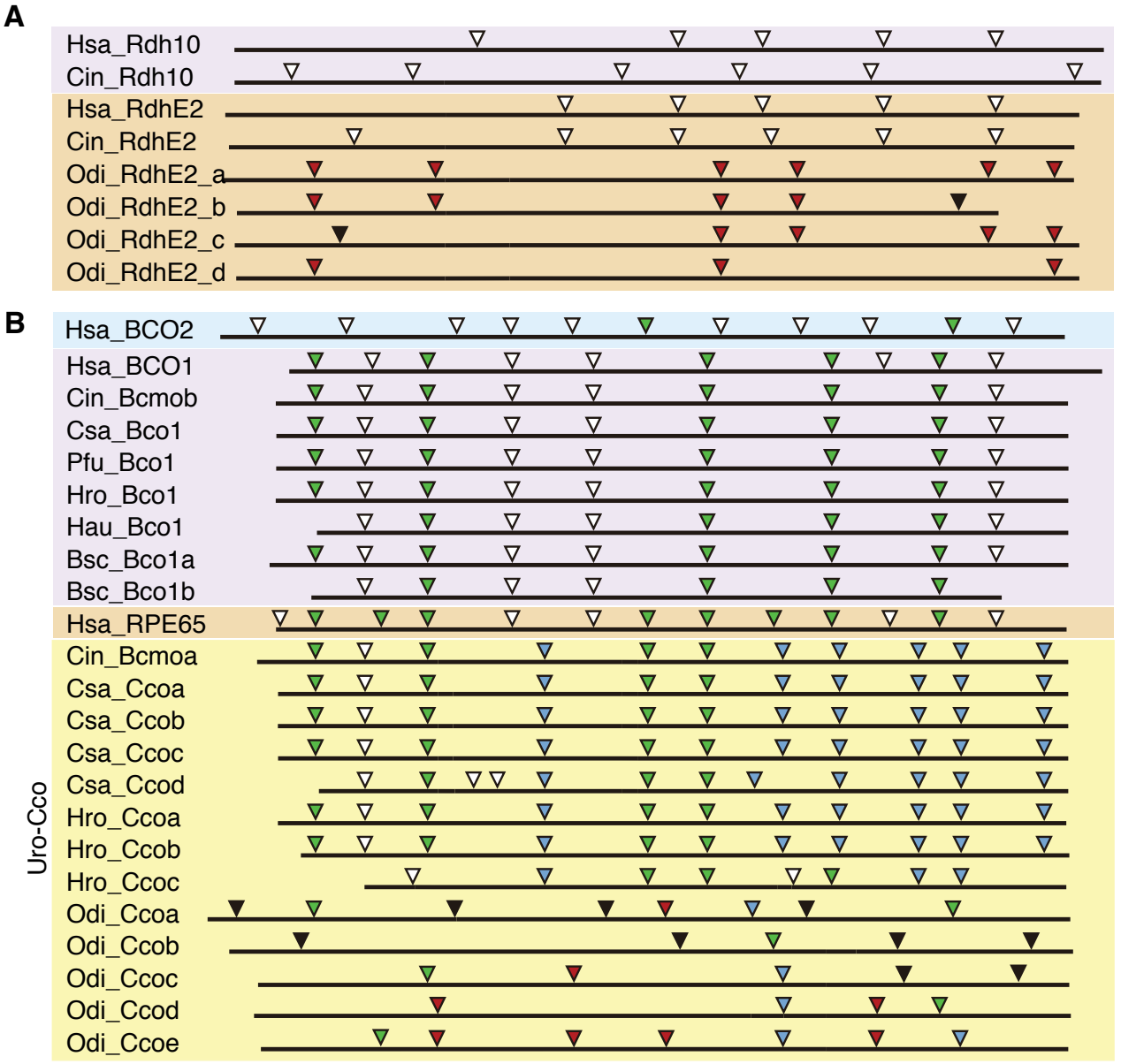

A
Hsa_Rdh10
Cin_Rdh10
Hsa_RdhE2
Cin_RdhE2
Odi_RdhE2_a
Odi_RdhE2_b
Odi_RdhE2_c
Odi_RdhE2_d
B
Hsa_BCO2
Hsa_BCO1
Cin_Bcmob
Csa_Bco1
Pfu_Bco1
Hro_Bco1
Hau_Bco1
Bsc_Bco1a
Bsc_Bco1b
Hsa_RPE65
Cin_Bcmoa
Csa_Ccoa
Csa_Ccob
Csa_Ccoc
Csa_Ccod
Hro_Ccoa
Hro_Ccob
Hro_Ccoc
Odi_Ccoa
Odi_Ccob
Odi_Ccoc
Odi_Ccod
Odi_Ccoe
Uro-Cco

Fig. S2

Fig. S2. Schematic comparison of intron positions between human, ascidian and *O. dioica* genes. Black horizontal lines represent aligned sequences and arrowheads represent intron positions. Red arrowheads denote species-specific introns shared by more than one *O. dioica* gene supporting *O. dioica* specific duplications. Black arrowheads indicate gene-specific *O. dioica* introns. (A) Six out of the eight *O. dioica* introns are shared between at least two *O. dioica RdhE2*, supporting a lineage-specific origin of *O. dioca* paralogs. Consistent with the drastic intron reorganization suffered by most *O. dioica* genes (Edvardsen, et al. 2004), none of the *O. dioica RdhE2* genes share any position with human and *C. intestinalis RdhE2* and *Rdh10* genes. (B) Comparison of *O. dioica Cco* genes with human *BCO1*, *BCO2* and *RPE65*, and ascidian *Bco1* and *Cco*. The presence of intron positions exclusively shared within the uro-Cco group (blue arrowheads), including three between *O. dioica* and ascidian *Cco*, supports a monophyletic origin of the uro-Cco group. Four (red arrowheads) out of the fourteen (black arrowheads) *O. dioica*-specific intron positions are shared by at least two *O. dioica Cco*, supporting a lineage-specific origin by gene duplication of *O. dioca* paralogs. From the positions shared by at least one gene of the uro-Cco group and any of the human *BCO1*, *BCO2* or *RPE65* genes (green arrowheads), the 8 positions shared with the *RPE65* suggest a closer relation between Uro-Cco and *RPE65* than with *BCO2*, with which only 2 intron positions are shared.

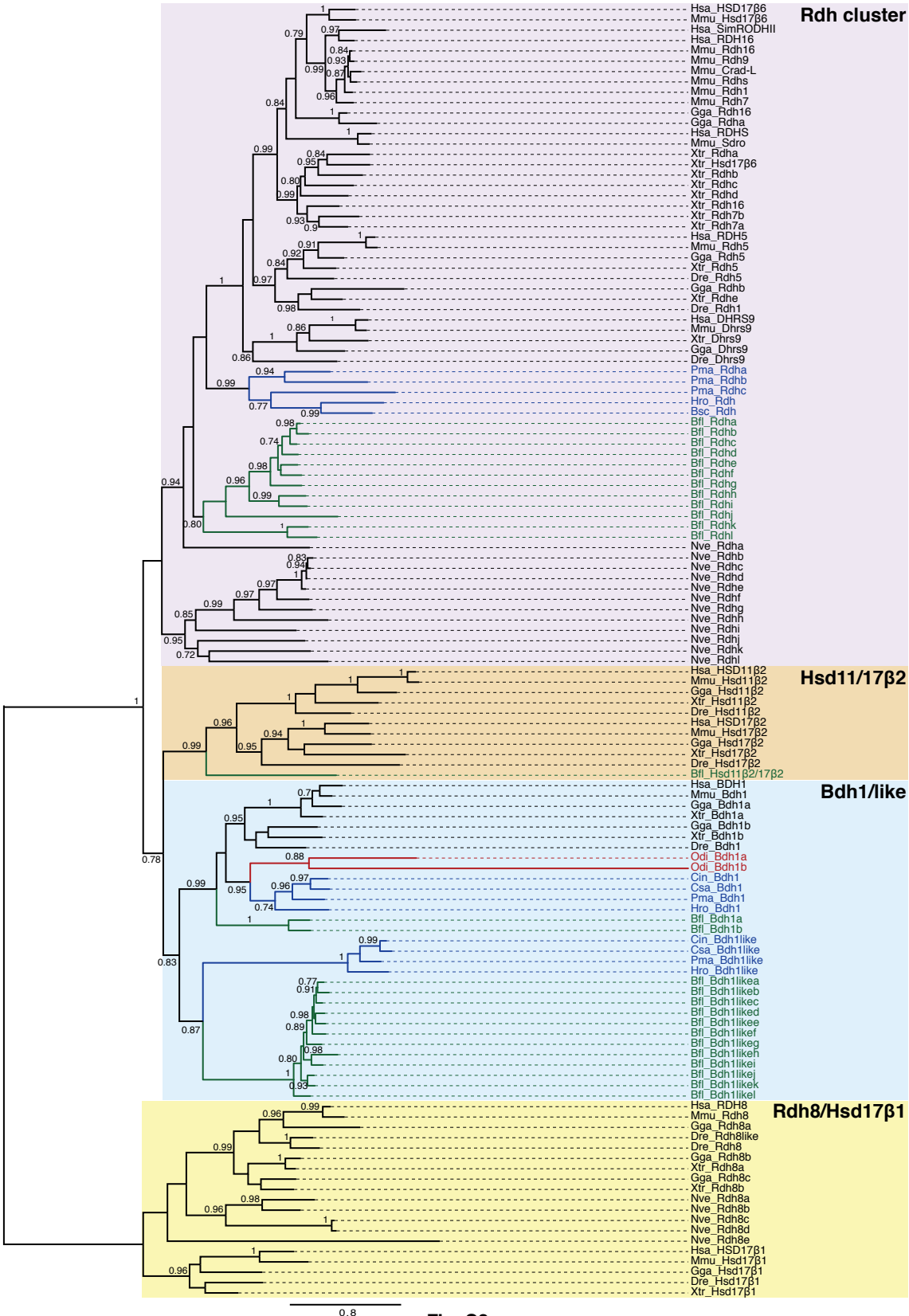

Rdh cluster
Hsa_HSD17β6
Mmu_Hsd17β6
Hsa_SimRODHII
Hsa_RDH16
Mmu_Rdh16
Mmu_Rdh9
Mmu_Crad-L
Mmu_Rdhs
Mmu_Rdh1
Mmu_Rdh7
Gga_Rdh16
Gga_Rdha
Hsa_RDHS
Mmu_Sdro
Xtr_Rdha
Xtr_Hsd17β6
Xtr_Rdhb
Xtr_Rdhc
Xtr_Rdhd
Xtr_Rdh16
Xtr_Rdh7b
Xtr_Rdh7a
Hsa_RDH5
Mmu_Rdh5
Gga_Rdh5
Xtr_Rdh5
Dre_Rdh5
Gga_Rdhb
Xtr_Rdhe
Dre_Rdh1
Hsa_DHRS9
Mmu_Dhrs9
Xtr_Dhrs9
Gga_Dhrs9
Dre_Dhrs9
Pma_Rdha
Pma_Rdhb
Pma_Rdhc
Hro_Rdh
Bsc_Rdh
Bfl_Rdha
Bfl_Rdhb
Bfl_Rdhc
Bfl_Rdhd
Bfl_Rdhe
Bfl_Rdhf
Bfl_Rdhg
Bfl_Rdhh
Bfl_Rdhi
Bfl_Rdhj
Bfl_Rdhk
Bfl_Rdhl
Nve_Rdha
Nve_Rdhb
Nve_Rdhc
Nve_Rdhd
Nve_Rdhe
Nve_Rdhf
Nve_Rdhg
Nve_Rdhh
Nve_Rdhi
Nve_Rdhj
Nve_Rdhk
Nve_Rdhl
Hsd11/17β2
Hsa_HSD11β2
Mmu_Hsd11β2
Gga_Hsd11β2
Xtr_Hsd11β2
Dre_Hsd11β2
Hsa_HSD17β2
Mmu_Hsd17β2
Gga_Hsd17β2
Xtr_Hsd17β2
Dre_Hsd17β2
Bfl_Hsd11β2/17β2
Bdh1/like
Hsa_BDH1
Mmu_Bdh1
Gga_Bdh1a
Xtr_Bdh1a
Gga_Bdh1b
Xtr_Bdh1b
Dre_Bdh1
Odi_Bdh1a
Odi_Bdh1b
Cin_Bdh1
Csa_Bdh1
Pma_Bdh1
Hro_Bdh1
Bfl_Bdh1a
Bfl_Bdh1b
Cin_Bdh1like
Csa_Bdh1like
Pma_Bdh1like
Hro_Bdh1like
Bfl_Bdh1likea
Bfl_Bdh1likeb
Bfl_Bdh1likec
Bfl_Bdh1liked
Bfl_Bdh1likee
Bfl_Bdh1likef
Bfl_Bdh1likeg
Bfl_Bdh1likeh
Bfl_Bdh1likei
Bfl_Bdh1likej
Bfl_Bdh1likek
Bfl_Bdh1likel
Rdh8/Hsd17β1
Hsa_RDH8
Mmu_Rdh8
Gga_Rdh8a
Dre_Rdh8like
Dre_Rdh8
Gga_Rdh8b
Xtr_Rdh8a
Gga_Rdh8c
Xtr_Rdh8b
Nve_Rdh8a
Nve_Rdh8b
Nve_Rdh8c
Nve_Rdh8d
Nve_Rdh8e
Hsa_HSD17β1
Mmu_Hsd17β1
Gga_Hsd17β1
Dre_Hsd17β1
Xtr_Hsd17β1
0.8

Fig. S3

Fig. S3. Maximum likelihood phylogenetic tree of SDR-9C family showing the absence of any *O. dioica* pro-ortholog to the vertebrate *Rdh16* within the Rdh cluster. The closest *O. dioica* homologs identified by BRBH clearly belonged to the sister subfamily Bdh1 in the tree. The SDR-28C family (Rdh8 + Hsd17β1) was used as outgroup to root the tree. Values for the approximate likelihood-ratio test (aLRT) supporting each node are shown. Species abbreviations: Vertebrates: Hsa, *Homo sapiens*; Mmu, *Mus musculus*; Gga, *Gallus gallus*; Xtr, *Xenopus tropicalis*; Dre, *Danio rerio*. Urochordate-ascidians (blue): Cin, *Ciona intestinalis*, Csa, *Ciona savignyi*; Hro, *Halocynthia roretzi*; Pma, *Phallusia mammillata*; Bsc, *Botryllus schlosseri*; Urochordate-larvaceans (red): Odi, *Oikopleura dioica*; Cephalochordates (green): Bfl, *Branchiostoma floridae*.

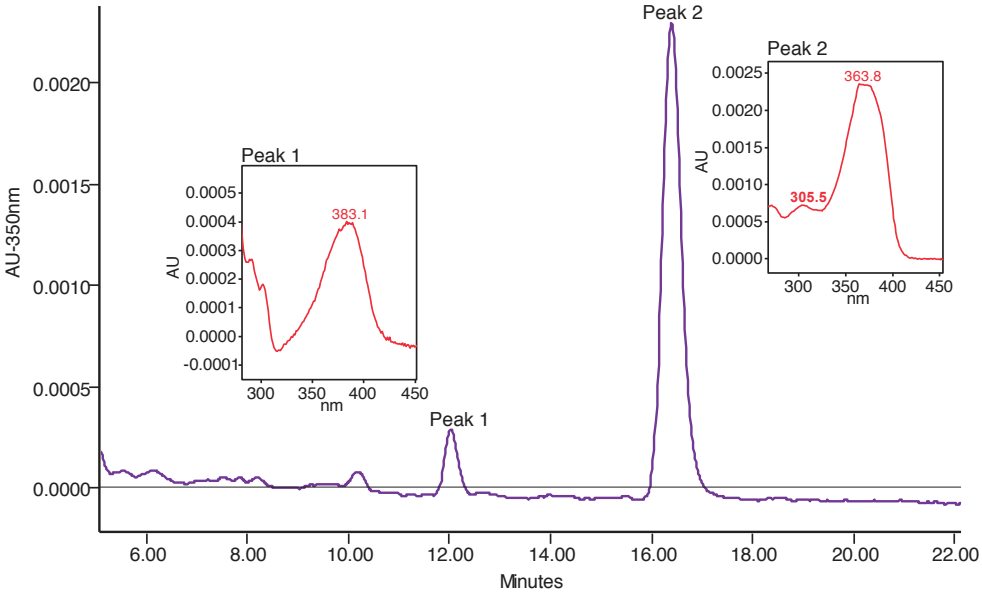

Peak 2
Peak 2
363.8
305.5
AU
nm
Peak 1
383.1
AU
nm
Peak 1
AU-350nm
Minutes
0.0020
0.0015
0.0010
0.0005
0.0000
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00

Fig. S4

Fig. S4. Reverse phase HPLC analysis of the retinoid content in *O. dioica* samples. Chromatogram is extracted at 350 nm. Peaks 1 and 2 may represent endogenous retinoids in *O. dioica*. Peak 1 is characterized by the absorbance spectrum and elution time similar to those of *cis*-RAL isomers. Peak 2 has an absorbance spectrum similar to RAL, but different elution time, and may represent different RAL derivative. Day-5 mature females sample is shown as a representative example.

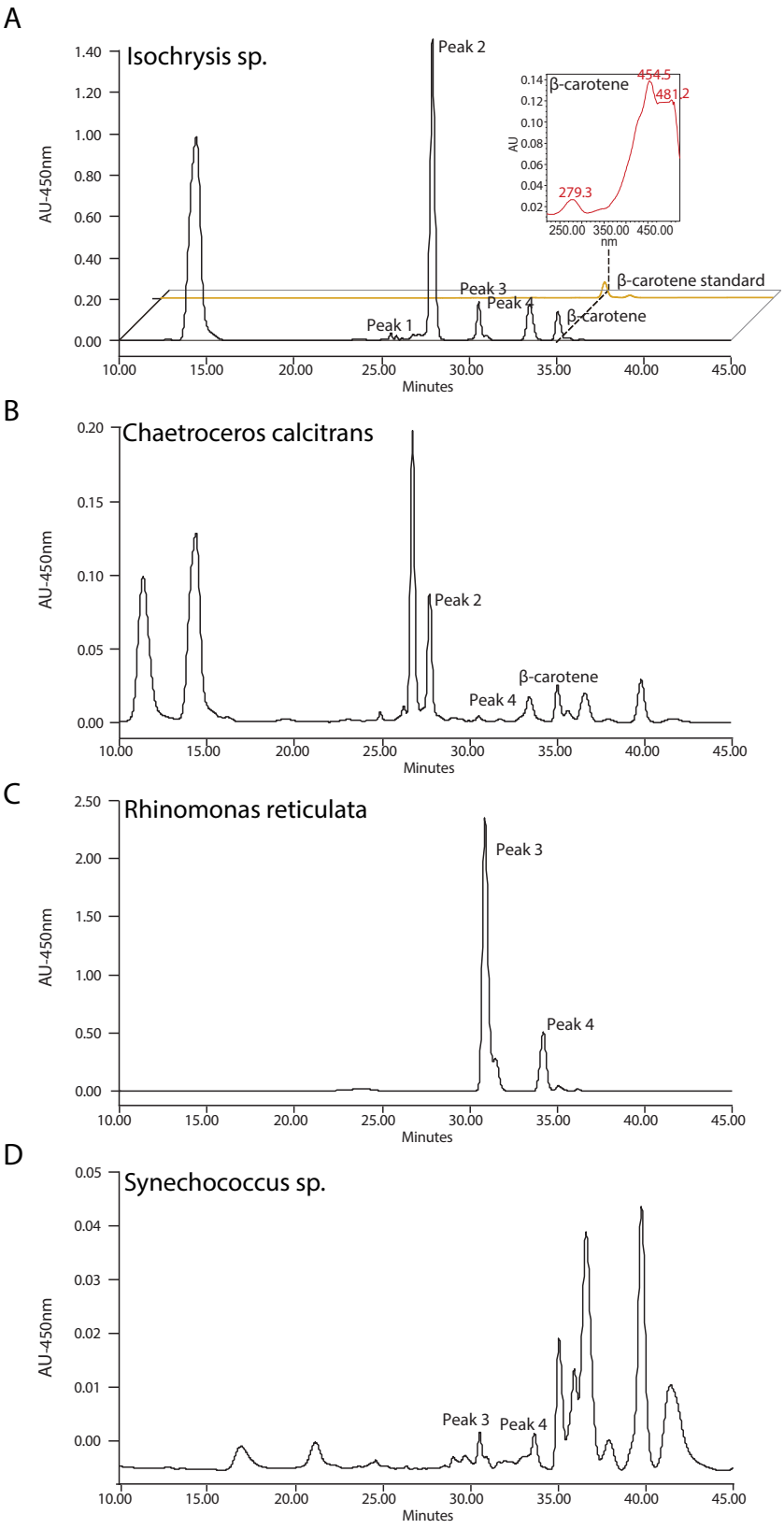

A
Isochrysis sp.
Peak 2
β-carotene
454.5
481.2
279.3
AU
nm
β-carotene standard
Peak 3
Peak 4
Peak 1
β-carotene
AU-450nm
Minutes
B
Chaetroceros calcitrans
Peak 2
β-carotene
Peak 4
AU-450nm
Minutes
C
Rhinomonas reticulata
Peak 3
Peak 4
AU-450nm
Minutes
D
Synechococcus sp.
Peak 3
Peak 4
AU-450nm
Minutes

Fig. S5

Fig. S5. HPLC profile of the carotenoid content of extracts of the four microalgae species used in the *O. dioica* diet: *Isochrysis sp.* (A), *Chaetroceros calcitrans* (B), *Rhinomonas reticulata* (C) and *Synechococcus sp* (D). Chromatogram extracted at 450 nm reveals four peaks, with peaks 2, 3 and 4 corresponding to carotenoids detected in *O.dioica* samples (Figure 5B), suggesting that *O. dioica* has the ability to absorb, store and metabolize carotenoids from the diet. The presence of β-carotene in at least two algae (*Isochrysis sp*. and *Chaetroceros calcitrans*) suggests that the absence of atRA in *O. dioica* was not due to the lack of β-carotene in the diet. Inset in A shows the typical β-carotene spectrum of the peak eluted at 35 min.

Resultats. Capítol 3

Wnt evolution and function shuffling in chordate genomes: from conservative amphioxus stasis to liberal ascidian gains and losses.

Ildiko M.L. Somorjai,[1,2] Josep Martí-Solans,[3] Miriam Diaz-Gracia,[3] Hiroki Nishida,[4] Kaoru S. Imai,[4] Hector Escrivà,[5] Cristian Cañestro[3] i Ricard Albalat[3]

[1]Biomedical Sciences Research Complex, School of Biology, University of St Andrews, North Haugh, St Andrews, KY16 9ST, Scotland, United Kingdom

[2]Scottish Oceans Institute, School of Biology, University of St Andrews, East Sands, St Andrews, KY16 8LB, Scotland, United Kingdom

[3]Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

[4]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan.

[5]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650, Banyuls/Mer, France.

Resum

L'impacte que ha tingut la pèrdua de gens sobre l'evolució dels mecanismes moleculars, i com la reassignació de funcions (*function shuffling*) ha afectat als gens retinguts que governen processos biològics essencials, són preguntes que es mantenen obertes. Per investigar aquests problemes, hem analitzat l'evolució del repertori de lligands Wnt en el fílum dels cordats com a cas pràctic. En aquest treball, hem realitzat una cerca exhaustiva en bases de dades genòmiques per trobar els gens *Wnt* en 13 espècies de cordats no vertebrats, identificant un total de 156 gens *Wnt*. Aquest resultats, representen el catàleg de gens *Wnt* més complet del fílum dels cordats, i ens han permès resoldre ambigüitats anteriors sobre l'ortologia de molts gens Wnt, inclosa la identificació de WntA per primera vegada en cordats. A més, hem creat el primer atles d'expressió completa per a la família Wnt durant el desenvolupament d'amfiox, oferint un recurs útil per investigar l'evolució de l'expressió Wnt a través de la radiació dels cordats. Les nostres dades subratllen l'estasi genòmica extraordinària en els cefalocordats, que contrasta amb els patrons evolutius liberals i dinàmics de pèrdua i duplicació de gens en genomes de cordats. La nostra anàlisi ens ha permès deduir funcions Wnt ancestrals compartides entre tots els cordats, diversos casos de function shuffling entre paràlegs de Wnt, així com dominis d'expressió únics per als gens Wnt que probablement reflecteixen innovacions funcionals particulars de cada llinatge. Finalment, proposem una relació potencial entre l'evolució de WntA i l'evolució de la boca en cordats.

Wnt evolution and function shuffling in chordate genomes: from conservative amphioxus stasis to liberal ascidian gains and losses.

Ildiko M.L. Somorjai[1,2] ✉, Josep Martí-Solans[3], Miriam Diaz-Gracia[3], Hiroki Nishida[4], Kaoru S. Imai[4], Hector Escrivà[5], Cristian Cañestro[3]✉ and Ricard Albalat[3]✉

[1]Biomedical Sciences Research Complex, School of Biology, University of St Andrews, North Haugh, St Andrews, KY16 9ST, Scotland, United Kingdom

[2]Scottish Oceans Institute, School of Biology, University of St Andrews, East Sands, St Andrews, KY16 8LB, Scotland, United Kingdom

[3]Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

[4]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan.

[5]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650, Banyuls/Mer, France.

✉ Correspondence to imls@st-andrews.ac.uk, canestro@ub.edu and ralbalat@ub.edu

Abstract

Background. What impact gene loss has had on the evolution of molecular mechanisms, and how function shuffling has affected retained genes driving essential biological processes, remain open questions in the fields of genome evolution and EvoDevo. To investigate these problems, we have analyzed the evolution of the Wnt ligand repertoire in the chordate phylum as a case study.

Results. We have conducted an exhaustive survey of *Wnt* genes in genomic databases, identifying 156 *Wnt* genes in 13 non-vertebrate chordates. This represents the most complete *Wnt* gene catalogue of the chordate subphyla, and has allowed us to resolve previous ambiguities about the orthology of many *Wnt* genes, including the identification of *WntA* for the first time in chordates. Moreover, we have created the first complete expression atlas for the Wnt family during amphioxus development, providing a useful resource to investigate the evolution of *Wnt* expression throughout the radiation of chordates.

Conclusions. Our data underscore extraordinary genomic stasis in cephalochordates, which contrasts with the liberal and dynamic evolutionary patterns of gene loss and duplication in urochordate genomes. Our analysis has allowed us to infer ancestral Wnt functions shared among all chordates, several cases of function shuffling among *Wnt* paralogues, as well as unique expression domains for *Wnt* genes that likely reflect functional innovations in each chordate lineage. Finally, we propose a potential relationship between the evolution of *WntA* and the evolution of the mouth in chordates.

Keywords: Wnt evolution, genome stasis, gene loss, function shuffling, chordate WntA, ascidians, vertebrates, amphioxus

1.Background

The era of comparative genomics is providing a new perspective on the evolution of living diversity by revealing an unexpected and significant amount of genetic complexity that already existed in ancestral organisms. This new perspective implies that evolutionary simplification –and not only complexification resulting from the acquisition of gene novelties [1] or from the co-option of pre-existing or duplicated genes for novel functions [2-5]– has been a prominent trend across the Tree of Life (reviewed in [6]). At the genetic level, simplification has often been accompanied by pervasive gene loss, providing an important source of genetic variation, in many cases even eliciting major evolutionary adaptive responses –'the less is more' principle [7]– (reviewed in [8]). However, understanding the significance of gene loss [8] and function shuffling among duplicated genes [9] on the generation of biodiversity, and especially their impact on the evolution of the genetic mechanisms of development of complex multicellular animals, is still a fundamental problem in Evolutionary and Developmental Biology. To explore this problem, we focus here on the evolution of the Wingless/Wnt family in chordates as paradigm; the Wnt family is among the best characterized of all metazoan gene families (reviewed in [8, 10, 11]), and plays conserved roles in fundamental developmental processes in all animals, including determination of the primary body axis, spatial cell patterning, cell fate specification, and cell proliferation and migration (reviewed in [11-13]).

The Wnt family encodes a set of secreted glycoprotein ligands that trigger a variety of signal transduction pathways to regulate gene transcription in target cells (e.g. [14]). The transduction of Wnt signaling can occur via two main pathways, the "canonical" and the "non-canonical", although they are not mutually exclusive [11]. The "canonical" Wnt/β-catenin pathway (a.k.a. cell-fate pathway) is mediated by the stabilization and transport of β-catenin into the nucleus, where it binds to transcription factors that regulate the expression of Wnt-target genes, and thus specify cellular fates. The various Wnt signaling pathways that act independently of β-catenin have been described as 'non-canonical", and despite their

diverse functions, they can be broadly grouped into the so-called “Wnt-cell polarity” pathway [11]. The ascription of each Wnt family member to a particular pathway is not straightforward, and it largely depends on the ability of each ligand to modulate β-catenin availability, or alternatively, to mediate cell behaviours. From cnidarians to vertebrates, for instance, Wnt1 and Wnt3 have been generally considered to signal through the “canonical” pathway, while Wnt5 and Wnt11 have been typically assigned to the Wnt-cell polarity pathway [15-17]. The fact, however, that certain Wnt ligands can be promiscuous and activate more than one pathway (e.g. [18]) makes it difficult to assign them to any specific group.

Wnt genes are a metazoan novelty [1] found from sponges to humans, and which duplicated and diversified into 13 subfamilies –Wnt1 to Wnt11, Wnt16 and WntA–, before the bilaterian and cnidarian split [19, 20]. Large-scale phylogenetic and genomic analyses have revealed that several *Wnt* genes have been lost and retained during animal evolution ([8, 10, 11, 21-23]). For instance, the gastropod *Patella vulgata* is the protostome that has suffered the greatest number of losses (9 out of 13, only preserving Wnt1, Wnt2, Wnt10 and WntA subfamilies), which is in stark contrast with the only 2 losses (Wnt3 and Wnt8) seen in another gastropod, *Lottia gigantea* [24, 25]. Other species such as *Drosophila melanogaster* and *Caenorhabditis elegans,* which have lost 6 (Wnt2 to 4, 11, 16 and A) and 8 (Wnt1 to 3, 6 to 8, 11 and 16) subfamilies, respectively, make evident that each animal lineage has shaped its own repertoire of *Wnt* genes. They also reveal that while many gene losses are recurrent and occurred independently in many lineages, others are ancestral and possibly important for the evolution of specific clades. For instance, the patchy pattern of absence/presence of *Wnt9* suggests that it has been lost at least 8 times during the evolution of arthropods, annelids, platyhelminthes, and cnidarians, while the absence of *Wnt3* in all protostomes suggests an early loss of *Wnt3* in the stem protostome ancestor, likely affecting the evolution of the entire group (reviewed in [8, 10, 11]).

In contrast to protostomes, vertebrates appear refractory to the loss of entire Wnt subfamilies. They have preserved at least one member in 12 out of the 13 subfamilies, with WntA the only subfamily that has not been found so far in vertebrates [24]. However, whether

the tendency to retain Wnt subfamilies is specific to vertebrates, or rather is a feature shared by all chordates (i.e. vertebrates + urochordates + cephalochordates) remains unkown. In urochordates (tunicates), the Wnt repertoire remains unresolved: phylogenetic classifications of *Wnt* genes from partial studies in three ascidians species (i.e. *Halocynthia roretzi*, *Ciona robusta* and *Botryllus schlosseri*) have resulted in many unascribed Wnt genes (referred to as 'orphan' *Wnt* genes), and in some cases conflicting orthologies due to the high sequence divergence typical of these species [26-33]. In cephalochordates, so far only 8 *Wnt* genes (*Wnt1, 3, 4, 5, 6, 7, 8, 11*) have been studied in one amphioxus species, *Branchiostoma floridae* ([34-39], reviewed in [40]), of which 5 (*Wnt3*, *5*, *6*, *7* and *8*) have been partially characterized in another, *B. lanceolatum* [41, 42]. Consequently, the taxonomic diversity of the analyzed urochordate and cephalochordate species has been too narrow, and the phylogenetic analysis of non-vertebrate chordate *Wnt* genes too ambiguous to draw general conclusions about the evolution and function of orthologous Wnt subfamilies in the chordate phylum.

In order to provide a comprehensive view of the evolution Wnt subfamilies in chordates, we have conducted an exhaustive survey of *Wnt* genes in genomic databases of 10 ascidian (urochordate subphylum) and 3 amphioxus (cephalochordate subphylum) species, and have generated the first complete atlas of developmental expression of the Wnt family in amphioxus. Our phylogenetic analysis represents, to our knowledge, the first fully resolved reconstruction of all Wnt subfamilies in the three chordate subphyla, resolving previous ambiguous or conflictive ascriptions of orthology. Our study reveals opposite trends in *Wnt* gene losses and retentions in cephalochordates and urochordates: while amphioxus show a conservative pattern of evolution, retaining the complete ancestral repertoire of chordate Wnt subfamilies, ascidians in contrast reveal a dynamic pattern of evolution, with numerous gene losses and duplications. Our study also demonstrates for the first time the presence of *WntA* genes in chordates (both in cephalochordates and in urochordates), which implies that the absence of the WntA subfamily in vertebrates is not due to an ancestral loss in chordates as previously suggested, but to a specific gene loss occurring during the early

evolution of vertebrates. Finally, our detailed atlas of *Wnt* expression in amphioxus, including the newly identified *WntA* genes in non-vertebrate chordates as well as several cases of 'function shuffling' [9], allow us to evaluate the contributions of different Wnt subfamilies to the diversification of each chordate subphylum.

2.Results

2.1. The Wnt *gene repertoire in non-vertebrate chordates*

Our comprehensive survey of *Wnt* genes in genomic databases of 13 non-vertebrate chordate species –3 cephalochordate species and 10 urochordate species representing five ascidian families from two different orders (2 solitary Cionidae and 2 solitary Ascidiidae within Phlebobranchia; and 2 solitary Pyuridae, 3 solitary Molgulidae, and 1 colonial Styelidae within Stolidobranchia)– identified 156 *Wnt* genes (Additional file 1: Table S1), which constitutes the first comprehensive catalogue of *Wnt* genes in non-vertebrate chordates. Our phylogenetic analyses, which included a total of 247 Wnt sequences from 19 species (Sequence alignment in Additional file 2) representing all major metazoan groups, from cnidarians to vertebrates, sorted the non-vertebrate chordate Wnt sequences into 13 monophyletic groups corresponding to the 13 known Wnt subfamilies (Fig. 1).

2.1.1 Conservative Wnt evolution in cephalochordates

Our phylogenetic analyses revealed that the three *Branchiostoma* species possess one orthologue for each of the 13 ancient Wnt subfamilies (Fig. 1A and B; see Additional file 2 for sequence alignment). Our results identified 5 *Wnt* genes that had not been analyzed in cephalochordates before, and corroborated the orthology of 8 previously described amphioxus *Wnt* genes ([34-39, 41] reviewed in [40]). We further extended our Wnt survey to the transcriptome project of *Asymmetron lucayanum*, a cephalochordate distantly related to the other *Branchiostoma* species [43]. We identified 13 Wnt sequences, mostly full length (Additional file 1: Table S1), each one orthologous to one of the 13 Wnt subfamilies (Additional file 1: Figure S4; see Additional file 3 for sequence alignment). The fact that all amphioxus Wnt orthologues form a single cluster nested within vertebrate and ambulacrarian

sequences from each Wnt subfamily suggests that neither gene duplications nor gene losses affected the evolution of Wnt subfamilies in the cephalochordate subphylum. Our findings, therefore, reinforce the idea of genomic and morphological evolutionary stasis attributed to cephalochordate species ([41, 44-48]; reviewed in [49]), in spite of divergence times estimated at over tens of millions of years ago [43, 50-53].

2.1.2 Liberal Wnt evolution with losses and duplications in ascidian urochordates

Our phylogenetic analyses provided the first fully resolved orthology of ascidian *Wnt* genes, allowing us to rename previously described genes still classified as 'orphan *Wnts*' with unclear orthology, as well as to settle conflicting orthology assignments, probably caused by limited species sampling [29-32] (Fig. 1A and Additional file 1: Table S1). The phylogenetic tree showed long branches for ascidian *Wnt* genes, which rarely clustered as the sister group of vertebrate *Wnt* genes within each Wnt subfamily, as would be expected from their taxonomic relationships, likely due to artifactual 'long-branch attraction' [54]. Our results showed that ascidians, in contrast to amphioxus, do not conserve the entire Wnt repertoire and have suffered various events of gene loss, as well as gene duplications (Fig. 1). While some of the losses appeared to be ancestral, resulting in the absence of Wnt subfamilies in all ascidian species, other losses and duplications affected different families and orders heterogeneously, suggesting a more dynamic evolution of *Wnt* genes in ascidians than in the conservative amphioxus. *Wnt4* and *Wnt8*, for instance, are absent in all analyzed species, plausibly due to two ancestral gene losses that occurred prior to the ascidian radiation, and therefore relevant to our understanding of the divergence in developmental mechanisms between ascidians and other chordates (Fig. 1B). On the other hand, *Wnt1* and *Wnt11* appear to be absent in species of the Phelobobranchia suborder and *Molgula* genus, respectively, while *Wnt3* is only absent in *B. schlosseri*. It seems therefore that loss of *Wnt* genes might have contributed to the genetic divergence between different groups or even single species of ascidians.

Our results, moreover, revealed that the Wnt5 subfamily in the Stolidobranchia order had experienced extensive amplification by gene duplication, affecting all six analyzed

184 species (Fig. 1B). The fact that many of these *Wnt5* duplicates appeared to be located in the
185 same genomic regions suggested that they originated by tandem gene duplications
186 (Additional file 1: Figure S1). The complex phylogenetic reconstructions of the ascidian Wnt5
187 clade were difficult to interpret, suggesting either the occurrence of independent gene
188 duplications in different lineages after events of speciation; or alternatively, ancestral *Wnt5*
189 duplications in stem Stolidobranchia, followed by multiple gene losses and events of gene
190 conversion (Additional file 1: Figure S1). In either case, the expansion of *Wnt5* in
191 Stolidobranchia provides a singular case of Wnt subfamily amplification in non-vertebrate
192 chordates, suggesting that the evolution of this order of ascidian species has been
193 accompanied by a relaxation of the evolutionary constraints that maintain *Wnt5* genes as a
194 single copy gene in other species. This may be linked to the recruitment of new *Wnt5*
195 paralogues in biological innovations unique to this group of ascidians.
196 Overall, the dynamic pattern of gene losses and duplications of *Wnt* genes observed
197 among different orders, families or individual species of ascidians correlates with the burst of
198 morphological diversification within the urochordate subphylum, contrasting with the
199 conservative pattern of Wnt evolution and morphological stasis observed in
200 cephalochordates.
201 *2.1.3 WntA, lost and found in chordates*
202 Our analysis also led to the identification of *WntA* genes in cephalochordates and
203 urochordates (Fig. 1 and Additional file 1: Table S1). This finding was of special interest
204 since *WntA* had previously been identified in cnidarians, protostomes and ambulacrarian
205 deuterostomes (e.g. in the hemichordate *Saccoglossus kowalevskii* and the echinoderm
206 *Strongylocentrotus purpuratus*), but not in any chordate, leading to the suggestion that the
207 WntA subfamily had been lost in stem chordates [10, 24]. The identification in this study of
208 *WntA* genes in both cephalochordate and urochordate species implies therefore that *WntA*
209 was present in the last common chordate ancestor, preserved in cephalochodates and
210 urochordates, but specifically lost during the early evolution of the vertebrate lineage.
211

2.2. The complete expression atlas of cephalochordate Wnt *genes*

The apparent conservative evolutionary stasis showed in amphioxus, and the finding that amphioxus possesses a full and non-redundant Wnt repertoire make this organism a very attractive model to infer the roles of *Wnt* genes in the ancestral stem chordate, and from comparative studies, to analyze the impact of changes in the Wnt repertoire during the evolution of each chordate subphylum. In order to obtain the first comprehensive stage-matched developmental expression atlas for the cephalochordate Wnt repertoire, we performed Whole-Mount *In Situ* Hybridization (WMISH) for all *Wnt* genes in the European species *B. lanceolatum* (Fig. 2 and Fig. 3). Our results revealed that *Wnt* genes were expressed in a robust and precise tissue-specific fashion, with significant overlap among several amphioxus *Wnt* paralogues. Nevertheless, we also saw a number of differences in the expression of different paralogues, suggesting a choreographed modulation of their expression domains throughout development to generate spatio-temporally complementary patterns (a gene by gene detailed description of the expression patterns of all *Wnt* genes shown in Fig. 2 is provided in Additional file 1: Text S1, and summarised in Fig. 4). Overall, *Wnt* expression appear to be highly dynamic, spanning a broad variety of tissues derived from all three germ layers, which precisely up- or down-regulated different *Wnt* paralogues throughout development.

2.2.1 Posterior dominance of Wnt *expression*

'Posteriority' is likely the most conspicuous hallmark observed in the expression domains of the majority of amphioxus *Wnt* genes. At the blastula stage, *Wnt1, Wnt8* and *Wnt11*, which were the first *Wnt* genes to be expressed in amphioxus (class A genes in Fig. 2), showed an evident restriction of their expression domains to the posterior half of the embryo. Thus, while *Wnt1* expression clearly labelled the vegetal pole, *Wnt8* and *Wnt11* were expressed at the equator of the prospective posterior pole [55]. At the gastrula stage G3, along with the early class A genes, *Wnt3, Wnt4* and *Wnt5* expression domains appeared surrounding the blastopore (white asterisks, class B genes in Fig. 2), as well as *Wnt6* a little later (at stage G7). They were expressed in these posterior regions through neurulation (N2

and N3) until larval stages (L1). Close observation of the blastoporal view revealed some degree of overlap among gene expression patterns, but also important differences among the *Wnt* expression domains of each paralogue. For instance, while *Wnt1* signal labeled the entire circumference surrounding the blastopore (Fig. 2, G3 column), *Wnt3* demarcated a narrower outer ectodermal strip of cells into G7 (Fig. 2). *Wnt8* and *Wnt11* signals were excluded from the edges of the blastopore and reached more central endomesodermal regions (Fig. 2). At G3, *Wnt4* signal was also strongly visible in the endomesoderm near the blastopore, but rather than being restricted to the posterior region, it spanned the entire layer surrounding the gastrocoel, and was excluded entirely from the ectoderm (Fig. 2). Finally, *Wnt5* most strongly marked the endomesoderm of the dorsal blastopore lip (Fig. 2). Most of these early posterior *Wnt* expression domains were steadily maintained throughout development. The *Wnt1* expression domain, for instance, remained in the posterior wall of the neurenteric canal after elongation and closure of the blastopore until at least the early larval stage (Fig. 2, L1 column). *Wnt5* strongly marked the posterior growth zone during neurulation (Fig. 2, N2 and N3 columns), culminating in strong tailbud expression in larvae (Fig. 2, L1 column). *Wnt3* signal was observed up to L1 stage in the posterior-most ectoderm fated to become tailfin, abutting the *Wnt1* domain.

Posteriority was also observed for some *Wnt* genes with late expression onset (Class C genes in Fig. 2). Of these, *Wnt10*, for instance, showed a new ectodermal expression domain in the most posterior part of the embryo at N2 and N3 (Fig. 2), which subsequently faded concomitantly with the appearance of *Wnt11* expression in this same region. This *Wnt11* expression clearly marked the entire fin in the 1 gill-slit larval stage (Fig. 2, L1 column, and Fig. 3B). In summary, our data show that nine out of the thirteen Wnt amphioxus *Wnt* genes showed posterior expression domains (Fig. 4), highlighting posteriority as one of the main hallmarks of this gene family.

2.2.2 Mesodermal Wnt *expression and somite formation*

In addition to the propensity for posterior dominance, another important source of *Wnt* signalling was observed in the presomitic and axial mesoderm, where seven out of the

thirteen *Wnt* paralogues were expressed. Besides the aforementioned early overlapping expression domains of *Wnt8* and *Wnt11* observed in the posterior endomesoderm surrounding the blastopore at G3, new expression domains of *Wnt4*, *Wnt5*, *Wnt6* and *Wnt16* consecutively appeared in the most posterior mesoderm by G7, N2, N3 and L1, respectively, in a temporally orchestrated manner. At the 1 gill-slit larval stage, *Wnt5* and *Wnt16* expression were maintained in the posterior-most mesoderm (Fig. 2, L1 column, and Fig. 3B). In addition to the *Wnt*-positive posterior mesodermal domain, several other *Wnt* expression domains could be identified in temporally dynamic complementary (as well as overlapping) patterns, in some cases forming nested domains along the anteroposterior axis (for instance *Wnt8/16-Wnt11-Wnt8/16-Wnt4/5/6*). In other cases Wnt expression domains differed in their dorso-ventral extent within somites. For instance, *Wnt10* appeared excluded from dorsal domains compared with *Wnt16* in the N2 stage. *Wnt16* signal was observed in the last pair of formed somites by L1, while *Wnt10* appeared to be dorsally restricted in all mature somites (see cross-section in Fig. 2), and excluded from this last pair. No *Wnt* signal was observed in the anteriormost pair of somites at any stage of development.

In contrast to the paraxial mesoderm, surprisingly few *Wnt* genes appeared to be expressed in chordamesoderm or other mesoderm derivatives. For example, only *Wnt2* and *Wnt5* were expressed in the notochord. *Wnt2* signal was localized to the anterior two thirds of the chordamesoderm during neurula stages (Fig. 2), becoming restricted to the most caudal and rostral portions in 1 gill-slit larvae (Fig. 3); in contrast, *Wnt5* was conspicuously expressed in the anterior notochord only in larval stages (Fig. 2, L1 column, and Fig. 3). Finally, several non-myotomal structures of mesodermal origin appeared to express *Wnt* genes in restricted domains. *Wnt4* signal was observed in mesothelial cells in late neurula and early larval stages on the left side adjacent to the pharynx (Fig. 2). *WntA* was expressed in larval stages in mesothelial cells between the ventral endoderm of the gut and the ectoderm in both early and late larvae (Fig. 2, Fig. 3B and Fig. 4).

2.2.3 Endodermal Wnt *expression*

In addition to mesodermal expression, *Wnt* signal was also evident in endodermal derivatives. By the gastrula stage, *Wnt8* and *Wnt11* expression domains were already observed in the endodermal portion delimiting the blastopore, expression domains that persisted until N3 and L1, respectively (Fig. 2). Many of the *Wnt* genes expressed in the posterior growth zone or tailbud also labeled hindgut domains in that area (e.g. *Wnt4*; Fig. 2). During larval stages, new anterior *Wnt* expression domains became evident in the anterior region and other parts of the digestive system (Fig. 2, L1 column). Some, such as *Wnt5,* *Wnt2* and *Wnt9*, appeared to be already expressed at late neurula stages in anterior endoderm (Fig. 2, N2 and N3 columns). *Wnt* genes labeling specific derivatives by L1 included *Wnt8*, *Wnt9* and *WntA* in the mouth primordium, *Wnt2*, *Wnt5*, *Wnt8*, and *Wnt11* in Haetschek's diverticula, *Wnt8* and *Wnt9* in the endostyle, and *Wnt9* in the branchial primordium (Fig. 2, L1 column). Once the mouth was open and the first gill slit was clearly formed, *Wnt5* labeled the club-shaped gland, while *Wnt8* and *Wnt9* labeled the preoral pit, *WntA* the entire circumference of the mouth, and *Wnt9* and *WntA* the first gill slit (Fig. 2 and Fig. 3A). *Wnt4* and *Wnt2* signal appeared in a few cells of the endostyle along with *Wnt8* and *Wnt9*. *Wnt10* signal was also clearly evident in cells lining the rostral coelom (Fig. 3A). Posteriorly, *Wnt6* appeared to be expressed in a single line of cells demarcating the posterior wall of the neurenteric canal, and *Wnt9* labeled cells within the posterior gut (Fig. 3B).

2.2.4 Ectodermal Wnt *expression*

Ectodermal *Wnt* expression was detected in the epidermis as well as the nervous system. Besides the sequential coexpression of *Wnt3*, *Wnt10* and *Wnt11* in the tip of the tail, the ventral epidermis of the anteroventral region at the level of the Haetschek's diverticulum also appeared to express *Wnt11* and *Wnt5* in L1 larvae (Fig. 2, L1 column). At some stage, all *Wnt* genes –with the exception of *Wnt11*– labeled the developing central nervous system (CNS), consisting of cerebral vesicle and nerve cord (Fig. 3 and Fig. 4). Thus, the invaginating neural tube expressed *Wnt7*, *Wnt3*, *Wnt6*, *Wnt2*, *Wnt4 Wnt8* starting in neurula stages, while *Wnt5, Wnt10* and *WntA* signal appeared later in early larvae (Fig. 2). The spatio-temporal expression profiles differed among ligands, ranging from continuous *Wnt7*

expression throughout the majority of the nervous system until larval stages, to the more restricted patterns shown by *Wnt2*, *Wnt4*, *Wnt6*, *WntA, Wnt10* or *Wnt*5. Some, such as *WntA* or *Wnt10*, only labeled a few isolated cells within the cerebral vesicle in the early larval stage L1 (Fig. 2, L1 column). However, by the 1 gill-slit larval stage, regionalisation of the cerebral vesicle became apparent, with *Wnt5* expression restricted to an anterior domain encompassing the frontal eye (Fig. 3A), while genes such as *Wnt4, Wnt7*, *Wnt8, Wnt9* and *Wnt10* (and possibly *Wnt16*) marked posterior regions of the cerebral vesicle or the hindbrain (Fig. 3B).

2.3 WntA *expression in non-vertebrate chordates*

We paid special attention to further investigating the function and evolution of our newly discovered chordate *WntA* genes. We therefore analyzed the expression patterns of *WntA* during embryonic development not only in the amphioxus *B. lanceolatum*, but also in two additional chordate species, the ascidians *C. robusta* (Phelobobranchia order) and *H. roretzi* (Stolidobranchia order). *WntA* expression, however, was not detected during embryonic development in either of these two ascidian species (Additional file 1: Figure S2). This lack of expression of *WntA* was consistent with the absence of ESTs from embryonic libraries in databases of *C. robusta* and *H. roretzi*; the only existing ESTs of the *WntA* gene (FF969784 and FF969783 of *C. robusta*) came from adult animals. These results suggested, therefore, that *WntA* might be exclusively expressed at postembryonic stages or during adulthood in ascidians.

In contrast to ascidians, we found that *WntA* was expressed in a complex pattern during amphioxus embryonic development. Our results revealed that amphioxus *WntA* was the last *Wnt* gene to turn on in mid-late neurulae, with expression in only one or two cells located anteriorly on the left side under the ectoderm (Fig. 2), likely in the oral mesovesicle (OMV). The OMV is a coelomic vesicle that develops from the posterior ventral corner of the left first somite, and which has been associated with amphioxus mouth opening [56]. This expression domain persisted throughout development, accompanied by punctuated

351 expression in the posterior cerebral vesicle from the late neurula until the pre-mouth larval
352 stage (see above). *WntA* signal was also observed in cells between the ectoderm and
353 endoderm of the forming gut of early and 1 gill-slit larvae (Fig. 2, cross-section, and Fig. 3),
354 possibly in the mesothelial precursors of the “amoebocytes”, considered to be the
355 homologues of the vertebrate blood cells [57]. Strong expression was further evident all
356 around the mouth opening in these late larvae (Fig 3A). This gene therefore represents a
357 late-phase *Wnt* (class D) that plays a major post-neurulation role in differentiating structures
358 such as the mouth and cerebral vesicle, and possibly in the circulatory system of amphioxus.
359

360 **3. Discussion**

361 ***3.1. Evolutionary patterns of Wnt subfamilies in non-vertebrate chordates***

362 The identification and fully resolved phylogenetic reconstruction of the entire Wnt
363 repertoire in several species of urochordate and cephalochordate (Fig. 1 and Additional file
364 1: Figure S4) permit the first complete reconstruction of the evolution of the Wnt subfamilies
365 in chordates, revealing that each subphylum follows different evolutionary trajectories.
366 Cephalochordates show a conservative evolutionary pattern, without either apparent gene
367 duplications or gene losses since the cephalochordate lineage diverged from other chordates
368 more than half a billion years ago [43, 50]. This finding suggests that the amphioxus genome
369 has preserved the complete and prototypical chordate/deuterostome/eumetazoan Wnt
370 repertoire (Fig. 1B). Analyses of many amphioxus gene families (e.g. *Hox* cluster [58],
371 homeobox gene families [59], gene families of the steroid and retinoic acid signaling pathway
372 [60-62], the tyrosine kinase superfamily [63], DNA-methylation genes [64, 65], and the FGF
373 gene family [66]) reinforces the idea of evolutionary stasis of cephalochordate genomes [43,
374 44, 48]. The expression patterns of all *Wnt* genes analysed here in *B. lanceolatum* are
375 congruent with the expression of their 8 orthologues previously characterised in the
376 American lancelet species *B. floridae* (*Wnt1, 3, 4, 5, 6, 7, 8* and *Wnt11*; [34-39], reviewed in
377 [40]), which extends previous reports of cephalochordate stasis of morphology and

developmental expression patterns [41, 46, 47, 67], and provides further support to the idea that the ancestral chordate resembled in many respects a modern amphioxus [58, 59].

In sharp contrast to the conservative nature of the amphioxus Wnt complement, ascidians show a dynamic pattern of evolution including several gene losses and duplications (Fig. 1B). Some of the gene losses are likely ancestral, affecting the early stem of the ascidian lineage, since some paralogues are absent in all ascidian species. In contrast, other losses appear to be more recent, affecting only some groups of ascidians (i.e. *Wnt1* in the Phelobobranchia suborder and *Wnt11* in the *Molgula* genus), or limited to specific species (*Wnt3* to *B. schlosseri*). Inferring the point at which Wnt losses occurred is not only important to better understand their possible impact on the divergence of developmental mechanisms between ascidians and other chordates, but it will also help elucidate how they may have contributed to genetic and morphological divergence during the ascidian radiation. Moreover, the extensive gene duplications of the *Wnt5* subfamily of the Stolidobranchia order may have also facilitated the divergence within this order. Since this is the only case of amplification of all *Wnt* subfamilies analyzed in non-vertebrate chordates, and recurrent tandem duplications have independently affected several species of the Stolidobranchia, the developmental constraints to maintain *Wnt5* as single copy seem to have been exclusively relaxed in this order of ascidians.

In contrast to cephalochordates and urochordates, many vertebrate Wnt subfamilies consist of two ohnologues (e.g. human *Wnt* subfamilies 2, 3, 5, 7, 8, 9 and 10), which originated during the two rounds of whole genome duplications (2R-WGD) that occurred during early vertebrate evolution [4]. The differences in the retention rate of *Wnt* paralogues between vertebrate and non-vertebrate chordates are possibly due to the different modes of duplication, i.e. genome duplication *vs.* small-scale duplication (reviewed in [4] and [8]), and suggest that duplication within Wnt subfamilies in non-vertebrate chordates might be impaired by dosage-imbalance with the exception of the Wnt5 subfamily in the Stolidobranchia order. Remarkably, WntA is the only Wnt subfamily absent in vertebrates. To investigate when the *WntA* gene was lost, we analyzed the Wnt catalogue of two additional

406 vertebrate species: the lamprey *Petromyzon marinus*, an agnathan representative, and the
407 shark *Callorhinchus milii*, a cartilaginous fish. We identified 24 Wnt sequences in *P. marinus*
408 and 20 in *C. milii* databases (Additional file 1: Table S1). In our phylogenetic reconstruction,
409 *P. marinus* and *C. milii* sequences distributed among all Wnt subfamilies, with the exception
410 of the WntA subfamily (Additional file 1: Figure S4; see Additional file 3 for sequence
411 alignment), suggesting that *WntA* was lost early in vertebrate evolution, before the
412 divergence of jawless and gnathostome lineages. It can be argued, therefore, that *WntA*
413 plausibly became dispensable in the primitive vertebrate either because alternative pathways
414 or genes compensated for its function, or because environmental or physiological changes
415 made them dispensable [8]. Finally, taking advantage of genomic information on the new
416 *Wnt* genes identified in this study, we have re-evaluated the conserved synteny of previously
417 postulated *Wnt* clusters [25]. Amphioxus has 4 clusters: the *Wnt9-Wnt1-Wnt6* cluster, the
418 *Wnt2-Wnt16* cluster, the *Wnt3-Wnt10* cluster, and the *Wnt5-Wnt7* cluster (Fig. 5; Additional
419 file 1: Table S3). In ascidians, we have found evidence for a single cluster: *Wnt9-Wnt6-Wnt3*
420 in *C. robusta* and *Wnt6-Wnt3* in *C. savignyi* (Fig. 5; Additional file 1: Table S3). Lamprey
421 conserves at least 5 clusters: the *Wnt9-Wnt1-Wnt6* cluster, the *Wnt3-Wnt10* and at least 3
422 *Wnt5-Wnt7* clusters (Fig. 5; Additional file 1: Table S3). Interestingly, cluster conservation is
423 higher between lamprey and amphioxus than between either of these and human (*Wnt5-*
424 *Wnt7*, and *Wnt5-Wnt7* and *Wnt2-Wnt16*, respectively), pointing to genomic rearrangements
425 in the lineage leading to humans. In summary, our work highlights how distinct genomic
426 rearrangements and patterns of gene conservation, loss and duplication shaped differently
427 the *Wnt* repertoire in amphioxus, ascidians and vertebrates, and provides an evolutionary
428 scenario that will facilitate future investigations of how these changes relate to adaptations to
429 the environmental and physiological requirements evolved by each subphylum.
430
431 ***3.2 Comparative analysis of* Wnt *expression patterns during embryonic***
432 ***development in the three chordate subphyla***

Our assignment of all non-vertebrate chordate *Wnt* genes to the different Wnt subfamilies permits the first comparison of expression patterns of all orthologues among vertebrate and non-vertebrate chordates. With the complete cephalochordate dataset in hand, we were able to compare the expression patterns of all *B. lanceolatum Wnt* genes with the reported patterns of vertebrate *Wnt* genes and those available for ascidians (Additional file 1: Figure S3 and Text S2). Overall, our analysis revealed three main situations: first, cases of orthologous *Wnt* genes that share expression domains in homologous structures among chordate species, likely reflecting ancestral functional conservation; second, homologous structures that share *Wnt* expression domains, although the orthology of the ligands is not conserved among taxa, suggesting therefore gene function shuffling; and third, *Wnt* expression domains absent in amphioxus but present in other chordates, possibly reflecting lineage-specific *Wnt* innovations during the evolution of Olfactores (vertebrates + urochordates), or simplifications of the cephalochordate lineage.

3.2.1 Ancestral conserved Wnt functions in chordates

Amphioxus *Wnt7* is highly expressed in the CNS (Fig. 2), a feature that is shared with ascidian and vertebrate Wnt homologues ([27, 34, 68]; reviewed in Additional file 1: Fig. S3 and Text S2). The posterior expression of *Wnt5* in somite and muscle development is also conserved in amphioxus (Fig. 2) ([39] and this work), vertebrates [69], and ascidians [28, 70]. *Wnt9* expression is conserved in endodermal derivatives in both amphioxus (Fig. 2) and vertebrates (specifically gut derivatives including the vertebrate liver), as well as in the gill slits, the vertebrate homologues of amphioxus pharyngeal arches, and in the amphioxus cerebral vesicle and in the brain of vertebrates [71-74], while to our knowledge, only partial expression data have been documented for *Wnt9* orthologues in a colonial ascidian [33]. Finally, shared expression domains are observed for of *Wnt10* and *Wnt16* in the neurectoderm, *Wnt16* in somites, and *Wnt5, Wnt3, Wnt8* and possibly *Wnt11* in the tailbud (Additional file 1: Fig. S3 and associated references). Globally, comparative analyses suggest a limited conservation in the expression patterns of orthologous Wnt subfamiles in the three chordate subphyla.

3.2.2 Function shuffling among Wnt paralogues

In vertebrates, expression of *Wnt1* is essential for proper anteriorposterior axial patterning of the brain and specification of particular neuronal populations (the “mid-hindbrain organizer”), in some cases functioning redundandly with other *Wnt* genes [75]. In amphioxus, no *Wnt1* expression has been observed in the anterior neuroectoderm or the cerebral vesicle ([38] and this study). However, we have identified a different ligand, *Wnt2*, with expression in the developing cerebral vesicle at mid-neurula stages, which is compatible with a role of amphioxus *Wnt2* in cerebral vesicle/hindbrain patterning. Other *Wnt* genes, including *Wnt3*, *Wnt5*, *Wnt7*, *Wnt8*, *Wnt9*, *WntA* and *Wnt16,* are also turned on in the amphioxus brain between neurulation and larval stages (reviewed in [40] and this study). These results are consistent with the unexpectedly complex genoarchitecture in the amphioxus neural tube that is conserved with vertebrates [42], and highlights events of ‘function shuffling’ [9] among Wnt paralogues during chordate evolution. Other remarkable examples of ‘function shuffling’ can be found in the notochord, one of the defining synapomorphies of all chordates. Besides *Wnt5* and *Wnt11*, which are near-ubiquitously expressed and may play more general functions (perhaps in cell movement or polarity) across chordates, *Wnt2* is the only paralogue expressed in the amphioxus notochord, while chick *Wnt16* (and maybe *Xenopus Wnt4* and *Wnt8* according to Xenobase data) is expressed in the vertebrate structure (Fig. 2; reviewed in Additional file 1: Figure S3). In contrast, *Wnt5* in *C. robusta* and the *Wnt-5*α paralogue in *H. roretzi* are the only *Wnt* genes so far identified with expression in the asicidian notochord [26, 31, 76]. Similar examples of Wnt function shuffling are observed in early mesodermal derivatives (e.g. early paraxial mesoderm or somites), which express *Wnt10* in amphioxus (this work) but *Wnt3* in both vertebrates and ascidians (Fig. 2; reviewed in Additional file 1: Figure S3).

Comparison of the development of the posterior pole of the embryo in all three subphyla reveals a complex scenario in which a variable number of Wnt subfamilies take part in different species. While in ascidians, *Wnt5* seems to be the only ligand determining early posteriority in the primary axis, in amphioxus our study reveals a highly redundant posterior

489 *Wnt* expression system involving at least eight out of the thirteen subfamilies (i.e. *Wnt1*, *8*, *11*
490 and *3* surrounding the blastopore during gastrulation, plus *Wnt4*, *5*, *16*, and *6* in the most
491 caudal part of the embryo later during neurulation and larval stages). In vertebrates,
492 interestingly, while some species (similarly to ascidians) use a reduced number of Wnt
493 ligands for determing early posteriority (e.g. *Wnt8a* in zebrafish [77], *Wnt11* (and *Wnt5*) in
494 *Xenopus* [78, 79], and *Wnt3* in mouse ([80] reviewed in [11]), other species such as chicken
495 show a redundant posterior Wnt system, more similarly to amphioxus (e.g. [81] [82];
496 Additional File 1: Figure S3 and associated references). If the ancestral chordate relied on a
497 simple Wnt system for determing posteriority, extensive Wnt function shuffling occurred
498 during the evolution of the cephalochordate lineage as well as some vertebrate species such
499 as chicken, recruiting other Wnt subfamilies for this posterior signalling role. Evidence from
500 the direct-developing hemichordate *Saccoglossus kowalevski* further corroborates the idea
501 that posterior Wnt flexibility is a common occurrence during evolution, with a different but
502 partially shared combination of *Wnt* genes showing blastoporal expression during
503 gastrulation (*Wnt1, Wnt3, Wnt4*, *Wnt6* and *Wnt16*) [83]. It seems therefore that providing β-
504 catenin is asymmetrically localised (along the axis or in dividing daughter cells), which
505 particular Wnt ligands act upstream, or how they are spatially organised relative to one
506 another, may not be particularly important [13], making that the Wnt system is one of the
507 gene families most prone to function shuffling so far described. Importantly, the extensive
508 events of function shuffling that occurred during Wnt evolution challenge the notion of
509 establishing homologies simply based on the expression of orthologous genes, and highlight
510 the need to consider these events when analyzing cases of deep homology.

511 *3.2.3* Wnt *expression domains absent in amphioxus but present in other chordates*

512 Different *Wnt* expression domains have been observed at different stages of germline
513 formation and gonadogenesis in Olfactores, such as *Wnt5* in two ascidian species and
514 mouse, *Wnt4* and *Wnt8* in zebrafish and mouse, *Wnt1* and *Wnt3* in chick, and *Wnt11* in
515 *Xenopus* (Additional File 1: Figure S3 and associated references), while no evidence has
516 been found suggesting any specific *Wnt* expression in cephalochordate primordial germ cells

517 (PGCs) or the germline (this work; [46, 67, 84, 85]; reviewed in [49]). Similarly, mesodermal
518 cardiac-related expression domain of *Wnt9* in Olfactores –i.e. the heart endocardium of
519 vertebrates [86] and the epithelial walls of the vasculature of the colonial ascidian
520 *B.schlosseri* [33]– is not paralleled by any *Wnt* paralogue in amphioxus. Moreover, in
521 vertebrates, many *Wnt* genes are expressed in placodal derivatives and neural crest
522 (Additional File 1: Figure S3 and associated references), and complex modulation of Wnt
523 signals both within ectoderm and from other tissues is required for their specification and
524 later differentiation [87-89]. In amphioxus, none of the ectodermal *Wnt* domains seem to be
525 compatible with the presence of placode-like structures, which supports the notion that
526 placodes may have been an evolutionary innovation of Olfactores [90-92], similarly to
527 migratory cells with neural crest-like properties [93]. It seems therefore that the appearance
528 of important functional novelties during chordate diversification –e. g. germline, heart or
529 placode development– were accompained by new expression domains and functions of *Wnt*
530 genes.
531
532 ***3.3. Evolution of* WntA: *another new mouth?***
533 Our work demonstrates the presence of *WntA* orthologues in both cephalochordates
534 and urochordates, suggesting that its absence in vetebrates is likely due to a specific gene
535 loss in this lineage. Our expression analyses reveal WntA function may be linked to mouth
536 development in amphioxus, since it is clearly expressed in the region where the mouth will
537 open at neurula stages, and around the periphery of the opening mouth in late larvae. A role
538 for Wnt signaling in cephalochordate mouth formation may be further supported by the
539 expression of Wnt antagonist *Dkk1/2/4* in the region in which the dissolution of basal laminae
540 and mouth perforation occur [56].
541 After the dorso-ventral inversion of chordates (i.e. chordates are dorsoventrally
542 inverted relative to non-chordates), the chordate mouth shifted from its now dorsal position
543 [94], implying that chordates evolved a new means of mouth formation. Two possible
544 evolutionary scenarios have been proposed for the origin of the amphioxus mouth, which

545 would have had different consequences on the evolution of *WntA*. In the first scenario, the
546 amphioxus mouth would share its evolutionary origins with the ambulacrarian coelomic pore-
547 canal [56]. In this case, the absence of *WntA* expression associated with pore-canal
548 formation in *Parentrotus lividus* [95] may suggest that WntA function was coopted in the
549 cephalochordate lineage (or secondarily lost in sea urchin). In the second scenario, the
550 amphioxus mouth would represent a specialized gill slit [96] supported by the fact that both
551 structures utilize Wnt (now including WntA), Fgf and Hh signaling pathways for their
552 formation [41, 97, 98], and that uncoupling of the gene regulatory network for mouth
553 formation from the blastopore could be key deuterostome innovations. The expression of
554 *WntA* in hemichordates is of particular relevance for discriminating between these
555 hypotheses. Although evidence points to a conserved pharyngeal transcriptional network in
556 deuterostomes [99], the expression patterns reported for *WntA* in *S kowalveskii* are for
557 stages too early to properly evaluate a putative conserved role in gill slit (or mouth) formation
558 [83].

559 Regarding a role for *WntA* in ascidian mouth development, it should be noted that
560 Olfactores and amphioxus primary mouths may not be homologous structures [56, 100]. To
561 open a new mouth, Olfactores developed an anterior placode or stomodaeum, whereas
562 amphioxus utilized an alternative system probably owing to its inability to form placodes
563 [101]. With an alternative mode to open a mouth, Olfactores no longer needed *WntA* for this
564 purpose, which might favor its loss in vertebrates and redeployment in urochordates. The
565 dorsal-ventral inversion hypothesis, besides postulating changes in mouth formation
566 mechanisms, proposes associated changes in brain architecture, which may have further
567 relaxed constraints on *WntA* (and other *Wnt*) gene expression in this structure in different
568 chordate lineages. Losses or redeployments of *WntA* have indeed been frequent during
569 evolution. *WntA* has been lost in different species of Arthropods, Annelids, Platyhelminthes
570 and Cnidarians [8], or used in a variety of conserved or novel structures and processes in
571 different prototostome species [21, 102-105]. *WntA* evolution illustrates, therefore, the
572 versatility of Wnt signaling in controlling diverse biological processes in metazoans.

3.4. Functional diversication and loss of Wnt signaling during animal evolution

Function shuffling has significant consequences because it makes it difficult to predict biological function from orthology, challenging the so-called ‘’orthology-function conjecture” (reviewed in [106]). This consequence is supported by the divergent expression patterns we observe here for orthologous chordate *Wnt* genes. Far from being a specific property of the chordate Wnt family, however, substantial differences have also been reported for *Wnt* expression in many other animal species, mainly arthopods and annelids [21, 25, 102, 107-111]. A detailed analysis of many *Wnt* genes in different protostome species, for instance, leads to the conclusion that the repertoire of Wnt ligands used during segment addition has evolved differentially among arthropod lineages, that the general role of *Wnt8* in regulating posterior development has been altered in annelids [21] and onychophorans [105], and that *Wnt5* and *Wnt16* orthologues are differentially expressed in annelids [25]. Outside bilaterians, substantial differences in the expression patterns of the ctenophore *Mnemiopsis leidyi Wnt* genes render difficult comparisons with those of other metazoans, including clades such as cnidarians, placozoans and poriferans [112].

Finally, the loss of WntA in vertebrates, as well as of a number of *Wnt* genes in ascidians, illustrate another general feature of Wnt evolution: the “pervasiveness” of the loss of *Wnt* genes during animal evolution. The loss of Wnt subfamilies in ascidians might have contributed to the morphological diversification of urochordates. Likewise, the loss of *Wnt6* in hemipterans has been linked to the loss of maxillary palps in this group of insects [113], while the loss of *Wnt2* and *Wnt4* in insects [108] might be related with arthropod diversification. Thus, pervasive gene loss accompained by numerous events of function shuffling appear to be two of the main features that characterize the evolution of the Wnt family not only in chordates, but also in all branches of metazoan evolution.

4. Conclusions

Up until now, Wnt research has mainly focused on identifying 'conserved' biological functions. Here, we argue that essential information can also be gleaned from the analysis of Wnt 'differences', many of them derived from events of function shuffling and gene loss. Understanding the biological basis of such differences will help uncover how highly conserved developmental processes –such as axial patterning, germlayer specification or body segmentation– might be controlled by molecular mechanisms (e.g. Wnt signaling) with remarkable genetic and functional diversity.

5. Materials and Methods

5.1. Genome Database Searches and Phylogenetic Analyses

Protein sequences of the Wnt repertoire from vertebrate *Homo sapiens*, urochordate *C. robusta* and cephalochordate *B. floridae* were used as queries in BLASTp and tBLASTn searches in genome databases of selected species: http://amphiencode.github.io/ for *B. lanceolatum*; http://genome.bucm.edu.cn/lancelet/ for *B. belcheri*; https://blast.ncbi.nlm.nih.gov/Blast.cgi for *B. floridae*; NCBI Sequence Read Archive: SRX437623 for *Asymmetron lucayanum*; http://www.aniseed.cnrs.fr/, https://blast.ncbi.nlm.nih.gov/Blast.cgi and *http://octopus.obs-vlfr.fr/public/botryllus/blastbotryllus.php* for ascidian species (*C. robusta, C. savignyi, P. fumigata, P. mammillata, H. roretzi, H. aurantium, B. schlosseri, M. occulta, M. oculata* and *M. occidentalis*); https://genomes.stowers.org/organism/Petromyzon/marinus and https://www.ensembl.org/Petromyzon_marinus/Info/Index for *P. marinus*; and NCBI database for *C. milii*. Orthologies of the non-vertebrate chordate *Wnt* were initially assessed by reciprocal best blast hit (RBBH) approach and corroborated by phylogenetic analyses. Phylogenetic reconstructions were based on ML inferences calculated with PhyML v3.0 and automatic mode of selection of substitution model [114] using protein alignments generated with MUSCLE [115] and CLUSTALX [116] programs and reviewed by hand. Accession numbers and protein alignment for phylogenetic tree reconstruction are provided in Additional file 1 (Table S1) and Additional file 2, respectively.

628

629 ***5.2. Animal collection and gene expression analysis by WMISH***

630 *C. robusta type A* (*C. robusta*) adults were obtained from the National Bio-Resource
631 Project for Ciona (AMED, Japan). *H. roretzi* adults were purchased from fishermen in Aomori
632 and Iwate prefectures, Japan. *B. lanceolatum* adults were collected in Argelès-sur-Mer,
633 France, and induced to spawn as in [117].

634 Fragments of *Wnt* genes were PCR amplified and cloned to synthesize gene-specific
635 riboprobes for *H. roretzi* and *B. lanceolatum Wnt* genes (Additional file 1: Table S2). For *C.*
636 *robusta*, cDNA clones were obtained from the cDNA clone collections [118, 119]. WMISH
637 experiments were performed as previously described in [120] for *C. robusta*; as in [121] for
638 *H. roretzi* with minor modifications (the acetylation step was not carried out before
639 prehybridization, and after the antibody incubation the specimens were washed with PBST
640 twelve times, 20 min each, and stored overnight at 4°C); and as in [41] for *B. lanceolatum.*
641 NBT/BCIP (Roche) or BMPurple (Roche) were used for the chromogenic reaction.

642

643 ***5.3 Comparative studies of expression patterns***

644 Vertebrate and ascidian *Wnt* gene expression patterns were identified and cross- and
645 back- referenced using published literature as well as public database searches. These
646 included ANISEED for ascidians species ([122, 123]; http://www.aniseed.cnrs.fr/), ZFIN for
647 zebrafish ([124]; www.zfin.org), Xenbase for Xenopus ([125-127]; http://www.xenbase.org/,
648 RRID:SCR_003280), Geisha for chick ([81]; www.geisha.arizona.edu/geisha/) and the
649 EMAGE gene expression database for mouse ([128]; http://www.emouseatlas.org/emage/).
650 As currently no such database is available for cephalochordates, published literature images
651 were examined by eye only. In all cases, special care was taken to ensure gene name
652 nomenclature in the literature matched our results for *Wnt* gene assignation. In vertebrates,
653 the expression of paralogues was grouped for ease of comparison across subphyla, under
654 the assumption that both neo- and subfuctionalisation events would be adequately
655 represented. Please see Additional file 1: Figure S2 for additional details.

Declarations

Availability of data and material

The data that support the findings of this study are available from public databases (please, see Material and Methods section for details). All analyses undertaken in the course of this study are included in this published article and its supplementary information files (please see Additional files 1 and 2).

Competing interests

The authors declare that they have no competing interests.

Funding

I. S. was supported by the European Union Horizon 2020 research and innovation programme under grant agreement numbers 654428 (“CORBEL”). H.N. was supported by Grants in Aid from The Japan Society for The Promotion of Science (22370078, 15H04377) and The Ministry of Education, Culture, Sports, Science and Thechnology, Japan (23112714, 25113518). K. I. was supported by a Grant in Aid from the Japan Society for the Promotion of Science (26711014). H. E. was supported by ANR-16-CE12-0008-01. C. C. was supported by BFU2016-80601-P. R. A. was supported by BIO2015-67358-C2-1-P grant from Ministerio de Economía y Competitividad (Spain). C. C. and R. A. were also supported by grants BFU2010-14875 from Ministerio de Ciencia e Innovación (Spain), and SGR2014-290 and SGR2017-1665 from Generalitat de Catalunya.

Authors' contributions

I. S. and R. A. collected all amphioxus Wnt sequences from databases. I. S. and H. E. obtained the amphioxus embryos, and I. S. carried out whole-mount *in situ* hybridization experiments for expression analysis. I. S. collated all expression data and did the comparative analyses of chordate *Wnt* genes. J. M.-S, M. D.-G. and C.C., collected all ascidian Wnt sequences from databases. C.C. and R.A. performed phylogenetic calculations, with support from J. M.-S., and M. D.-G., and elaborated the evolutionary inferences. J. M.-S., K. I. and H. N. analyzed the expression of ascidian *WntA.* I. S., C.C. and

R.A. designed the study and analyzed data. I. S., C.C. and R.A. wrote the manuscript with imput from J. M.-S., H. E. and H. N.. I. S., J. M.-S., C.C. and R. A. finalized figures, tables, and text. All authors commented on the manuscript and agreed to its final version.

Acknowledgements

The authors thank all team members of the C.C. and R. A. laboratories for fruitful discussions on Wnt signaling, gene loss and evolution. The authors thank the *Phallusia* and *Halocynthia* genome sequencing consortia, as well as all other ascidian laboratories for sharing results ahead of their publication. This work was supported by the *Branchiostoma lanceolatum* genome consortium, which provided access to the *Branchiostoma lanceolatum* genome sequence. The non-human silhouettes used in Figure 5 are courtesy of Tom Barton-Owen, with modifications by I.S.

References

1. Paps J, Holland PWH: **Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty.** *Nat Commun* 2018, **9:**1730.
2. Suga H, Chen Z, de Mendoza A, Sebe-Pedros A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sanchez-Pons N, et al: **The Capsaspora genome reveals a complex unicellular prehistory of animals.** *Nat Commun* 2013, **4:**2325.
3. Taylor JS, Raes J: **Duplication and divergence: the evolution of new genes and old ideas.** *Annu Rev Genet* 2004, **38:**615-643.
4. Cañestro C, Albalat R, Irimia M, Garcia-Fernandez J: **Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates.** *Semin Cell Dev Biol* 2013, **24:**83-94.
5. Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH: **Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes.** *PLoS Genet* 2009, **5:**e1000496.
6. O'Malley MA, Wideman JG, Ruiz-Trillo I: **Losing Complexity: The Role of Simplification in Macroevolution.** *Trends Ecol Evol* 2016, **31:**608-621.
7. Olson MV: **When less is more: gene loss as an engine of evolutionary change.** *Am J Hum Genet* 1999, **64:**18-23.
8. Albalat R, Canestro C: **Evolution by gene loss.** *Nat Rev Genet* 2016, **17:**379-391.
9. McClintock JM, Carlson R, Mann DM, Prince VE: **Consequences of *Hox* gene duplication in the vertebrates: an investigation of the zebrafish *Hox* paralogue group 1 genes.** *Development* 2001, **128:**2471-2484.
10. Holstein TW: **The evolution of the Wnt pathway.** *Cold Spring Harb Perspect Biol* 2012, **4:**a007922.

11. Loh KM, van Amerongen R, Nusse R: **Generating Cellular Dive sity and Spatial Form: Wnt Signaling and the Evolution of Multicellular Animals.** *Dev Cell* 2016, **38:**643-655.
12. Guder C, Philipp I, Lengfeld T, Watanabe H, Hobmayer B, Holstein TW: **The Wnt code: cnidarians signal the way.** *Oncogene* 2006, **25:**7450-7460.
13. Petersen CP, Reddien PW: **Wnt signaling and the polarity of the primary body axis.** *Cell* 2009, **139:**1056-1068.
14. Niehrs C: **The complex world of WNT receptor signalling.** *Nat Rev Mol Cell Biol* 2012, **13:**767-779.
15. Rigo-Watermeier T, Kraft B, Ritthaler M, Wallkamm V, Holstein T, Wedlich D: **Functional conservation of Nematostella Wnts in canonical and noncanonical Wnt-signaling.** *Biol Open* 2012, **1:**43-51.
16. Andre P, Song H, Kim W, Kispert A, Yang Y: **Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation.** *Development* 2015, **142:**1516-1527.
17. Kraus Y, Aman A, Technau U, Genikhovich G: **Pre-bilaterian origin of the blastoporal axial organizer.** *Nat Commun* 2016, **7:**11694.
18. van Amerongen R, Fuerer C, Mizutani M, Nusse R: **Wnt5a can both activate and repress Wnt/beta-catenin signaling during mouse embryonic development.** *Dev Biol* 2012, **369:**101-114.
19. Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW: **Unexpected complexity of the Wnt gene family in a sea anemone.** *Nature* 2005, **433:**156-160.
20. Borisenko I, Adamski M, Ereskovsky A, Adamska M: **Surprisingly rich repertoire of Wnt genes in the demosponge Halisarca dujardini.** *BMC Evol Biol* 2016, **16:**123.
21. Janssen R, Le Gouar M, Pechmann M, Poulin F, Bolognesi R, Schwager EE, Hopfen C, Colbourne JK, Budd GE, Brown SJ, et al: **Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation.** *BMC Evol Biol* 2010, **10:**374.
22. Murat S, Hopfen C, McGregor AP: **The function and evolution of Wnt genes in arthropods.** *Arthropod Struct Dev* 2010, **39:**446-452.
23. Rentzsch F, Technau U: **Genomics and development of Nematostella vectensis and other anthozoans.** *Curr Opin Genet Dev* 2016, **39:**63-70.
24. Prud'homme B, Lartillot N, Balavoine G, Adoutte A, Vervoort M: **Phylogenetic analysis of the Wnt gene family. Insights from lophotrochozoan members.** *Curr Biol* 2002, **12:**1395.
25. Cho SJ, Valles Y, Giani VC, Jr., Seaver EC, Weisblat DA: **Evolutionary dynamics of the wnt gene family: a lophotrochozoan perspective.** *Mol Biol Evol* 2010, **27:**1645-1658.
26. Sasakura Y, Ogasawara M, Makabe KW: **HrWnt-5: A maternally expressed ascidian Wnt gene with posterior localization in early embryos.** *International Journal Of Developmental Biology May* 1998, **42:**573-579.
27. Sasakura Y, Makabe KW: **Ascidian Wnt-7 gene is expressed exclusively in the tail neural tube of tailbud embryos.** *Dev Genes Evol* 2000, **210:**641-643.
28. Miya T, Nishida H: **Isolation of cDNA clones for mRNAs transcribed zygotically during cleavage in the ascidian, Halocynthia roretzi.** *Dev Genes Evol* 2002, **212:**30-37.

29. Hino K, Satou Y, Yagi K, Satoh N: **A genomew de s vey o developmentally relevant genes in Ciona intestinalis. VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways.** *Dev Genes Evol* 2003, **213:**264-272.
30. Hotta K, Takahashi H, Ueno N, Gojobori T: **A genome-wide survey of the genes for planar polarity signaling or convergent extension-related genes in Ciona intestinalis and phylogenetic comparisons of evolutionary conserved signaling components.** *Gene* 2003, **317:**165-185.
31. Imai KS, Hino K, Yagi K, Satoh N, Satou Y: **Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks.** *Development* 2004, **131:**4047-4058.
32. Rosner A, Alfassi G, Moiseeva E, Paz G, Rabinowitz C, Lapidot Z, Douek J, Haim A, Rinkevich B: **The involvement of three signal transduction pathways in botryllid ascidian astogeny, as revealed by expression patterns of representative genes.** *Int J Dev Biol* 2014, **58:**677-692.
33. Di Maio A, Setar L, Tiozzo S, De Tomaso AW: **Wnt affects symmetry and morphogenesis during post-embryonic development in colonial chordates.** *Evodevo* 2015, **6:**17.
34. Schubert M, Holland LZ, Holland ND: **Characterization of two amphioxus Wnt genes (AmphiWnt4 and AmphiWnt7b) with early expression in the developing central nervous system.** *Dev Dyn* 2000, **217:**205-215.
35. Schubert M, Holland LZ, Holland ND: **Characterization of an amphioxus wnt gene, AmphiWnt11, with possible roles in myogenesis and tail outgrowth.** *Genesis* 2000, **27:**1-5.
36. Schubert M, Holland LZ, Holland ND, Jacobs DK: **A phylogenetic tree of the Wnt genes based on all available full-length sequences, including five from the cephalochordate amphioxus.** *Mol Biol Evol* 2000, **17:**1896-1903.
37. Schubert M, Holland LZ, Panopoulou GD, Lehrach H, Holland ND: **Characterization of amphioxus AmphiWnt8: insights into the evolution of patterning of the embryonic dorsoventral axis.** *Evol Dev* 2000, **2:**85-92.
38. Holland LZ, Holland ND, Schubert M: **Developmental expression of *AmphiWnt1*, an amphioxus gene in the *Wnt1/wingless* subfamily.** *Dev Genes Evol* 2000, **210:**522-524.
39. Schubert M, Holland LZ, Stokes MD, Holland ND: **Three amphioxus Wnt genes (AmphiWnt3, AmphiWnt5, and AmphiWnt6) associated with the tail bud: the evolution of somitogenesis in chordates.** *Dev Biol* 2001, **240:**262-273.
40. Bertrand S, Le Petillon Y, Somorjai I, Escriva H: **Developmental cell-cell communication pathways in the cephalochordate amphioxus: actors and functions.** *Int J Dev Biol* 2017, **61:**697-722.
41. Somorjai I, Bertrand S, Camasses A, Haguenauer A, Escriva H: **Evidence for stasis and not genetic piracy in developmental expression patterns of Branchiostoma lanceolatum and Branchiostoma floridae, two amphioxus species that have evolved independently over the course of 200 Myr.** *Dev Genes Evol* 2008, **218:**703-713.
42. Albuixech-Crespo B, Lopez-Blanch L, Burguera D, Maeso I, Sanchez-Arrones L, Moreno-Bravo JA, Somorjai I, Pascual-Anaya J, Puelles E, Bovolenta P, et al: **Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain.** *PLoS Biol* 2017, **15:**e2001573.

43. Yue JX, Yu JK, Putnam NH, Holland L : **The transcr ptome o an amp oxus, Asymmetron lucayanum, from the Bahamas: a window into chordate evolution.** *Genome Biol Evol* 2014, **6:**2681-2696.
44. Cañestro C, Albalat R: **Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences.** *Brief Funct Genomics* 2012, **11:**131-141.
45. Louis A, Roest Crollius H, Robinson-Rechavi M: **How much does the amphioxus genome represent the ancestor of chordates?** *Brief Funct Genomics* 2012, **11:**89-95.
46. Dailey SC, Febrero Planas R, Rossell Espier A, Garcia-Fernàndez J, Somorjai IML: **Asymmetric Distribution of pl10 and bruno2, New Members of a Conserved Core of Early Germline Determinants in Cephalochordates.** *Frontiers in Ecology and Evolution* 2016, **3**.
47. Yong LW, Bertrand S, Yu JK, Escriva H, Holland ND: **Conservation of BMP2/4 expression patterns within the clade Branchiostoma (amphioxus): Resolving interspecific discrepancies.** *Gene Expr Patterns* 2017, **25-26:**71-75.
48. Paps J, Holland PW, Shimeld SM: **A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus?** *Brief Funct Genomics* 2012, **11:**177-186.
49. Somorjai I: **Amphioxus regeneration: evolutionary and biomedical implications.** *Int J Dev Biol* 2017.
50. Cañestro C, Albalat R, Hjelmqvist L, Godoy L, Jornvall H, Gonzalez-Duarte R: **Ascidian and amphioxus Adh genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution.** *J Mol Evol* 2002, **54:**81-89.
51. Nohara M, Nishida M, Nishikawa T: **New complete mitochondrial DNA sequence of the lancelet Branchiostoma lanceolatum (Cephalochordata) and the identity of this species' sequences.** *Zoolog Sci* 2005, **22:**671-674.
52. Kon T, Nohara M, Yamanoue Y, Fujiwara Y, Nishida M, Nishikawa T: **Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences.** *BMC Evol Biol* 2007, **7:**127.
53. Igawa T, Nozawa M, Suzuki DG, Reimer JD, Morov AR, Wang Y, Henmi Y, Yasui K: **Evolutionary history of the extant amphioxus lineage with shallow-branching diversification.** *Sci Rep* 2017, **7:**1157.
54. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Systematic Zoology* 1978, **27:**401–410.
55. Holland LZ, Holland ND: **A revised fate map for amphioxus and the evolution of axial patterning in chordates.** *Integr Comp Biol* 2007, **47:**360-372.
56. Kaji T, Reimer JD, Morov AR, Kuratani S, Yasui K: **Amphioxus mouth after dorso-ventral inversion.** *Zoological Lett* 2016, **2:**2.
57. Pascual-Anaya J, Albuixech-Crespo B, Somorjai IM, Carmona R, Oisi Y, Alvarez S, Kuratani S, Munoz-Chapuli R, Garcia-Fernandez J: **The evolutionary origins of chordate hematopoiesis and vertebrate endothelia.** *Dev Biol* 2013, **375:**182-192.
58. Garcia-Fernàndez J, Holland PW: **Archetypal organization of the amphioxus Hox gene cluster.** *Nature* 1994, **370:**563-566.
59. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, et al: **The amphioxus genome**

illuminates vertebrate origins and cephalochordate biology. *Genome Res* 2008, **18:**1100-1111.
60. Cañestro C, Postlethwait JH, Gonzàlez-Duarte R, Albalat R: **Is retinoic acid genetic machinery a chordate innovation?** *Evol Dev* 2006, **8:**394-406.
61. Albalat R: **The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations.** *Mol Cell Endocrinol* 2009, **313:**23-35.
62. Albalat R, Brunet F, Laudet V, Schubert M: **Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective.** *Genome Biol Evol* 2011, **3:**985-1005.
63. D'Aniello S, Irimia M, Maeso I, Pascual-Anaya J, Jimenez-Delgado S, Bertrand S, Garcia-Fernandez J: **Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus.** *Mol Biol Evol* 2008, **25:**1841-1854.
64. Albalat R: **Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus Branchiostoma floridae.** *Dev Genes Evol* 2008, **218:**691-701.
65. Albalat R, Marti-Solans J, Canestro C: **DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates.** *Brief Funct Genomics* 2012, **11:**142-155.
66. Oulion S, Bertrand S, Escriva H: **Evolution of the FGF Gene Family.** *Int J Evol Biol* 2012, **2012:**298147.
67. Zhang QJ, Luo YJ, Wu HR, Chen YT, Yu JK: **Expression of germline markers in three species of amphioxus supports a preformation mechanism of germ cell development in cephalochordates.** *Evodevo* 2013, **4:**17.
68. Abitua PB, Wagner E, Navarrete IA, Levine M: **Identification of a rudimentary neural crest in a non-vertebrate chordate.** *Nature* 2012, **492:**104-107.
69. Blader P, Strahle U, Ingham PW: **Three Wnt genes expressed in a wide variety of tissues during development of the zebrafish, Danio rerio: developmental and evolutionary perspectives.** *Dev Genes Evol* 1996, **206:**3-13.
70. Yagi K, Satoh N, Satou Y: **Identification of downstream genes of the ascidian muscle determinant gene Ci-macho1.** *Dev Biol* 2004, **274:**478-489.
71. Curtin E, Hickey G, Kamel G, Davidson AJ, Liao EC: **Zebrafish wnt9a is expressed in pharyngeal ectoderm and is required for palate and lower jaw development.** *Mech Dev* 2011, **128:**104-115.
72. Quinlan R, Graf M, Mason I, Lumsden A, Kiecker C: **Complex and dynamic patterns of Wnt pathway gene expression in the developing chick forebrain.** *Neural Dev* 2009, **4:**35.
73. Cox AA, Jezewski PA, Fang PK, Payne-Ferreira TL: **Zebrafish Wnt9a,9b paralog comparisons suggest ancestral roles for Wnt9 in neural, oral-pharyngeal ectoderm and mesendoderm.** *Gene Expr Patterns* 2010, **10:**251-258.
74. Zhang B, Tran U, Wessely O: **Expression of Wnt signaling components during Xenopus pronephros development.** *PLoS One* 2011, **6:**e26533.
75. Lekven AC, Buckles GR, Kostakis N, Moon RT: **Wnt1 and wnt10b function redundantly at the zebrafish midbrain-hindbrain boundary.** *Dev Biol* 2003, **254:**172-187.
76. Niwano T, Takatori N, Kumano G, Nishida H: **Wnt5 is required for notochord cell intercalation in the ascidian Halocynthia roretzi.** *Biol Cell* 2009, **101:**645-659.

77. Lu FI, Thisse C, Thisse B: **Identi cation and mec anism of regulat on o the zebrafish dorsal determinant.** *Proc Natl Acad Sci U S A* 2011, **108:**15876-15880.
78. Tao Q, Yokota C, Puck H, Kofron M, Birsoy B, Yan D, Asashima M, Wylie CC, Lin X, Heasman J: **Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos.** *Cell* 2005, **120:**857-871.
79. Cha SW, Tadjuidje E, Tao Q, Wylie C, Heasman J: **Wnt5a and Wnt11 interact in a maternal Dkk1-regulated fashion to activate both canonical and non-canonical signaling in Xenopus axis formation.** *Development* 2008, **135:**3719-3729.
80. Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A: **Requirement for Wnt3 in vertebrate axis formation.** *Nat Genet* 1999, **22:**361-365.
81. Antin PB, Yatskievych TA, Davey S, Darnell DK: **GEISHA: an evolving gene expression resource for the chicken embryo.** *Nucleic Acids Res* 2014, **42:**D933-937.
82. Martin A, Maher S, Summerhurst K, Davidson D, Murphy P: **Differential deployment of paralogous Wnt genes in the mouse and chick embryo during development.** *Evol Dev* 2012, **14:**178-195.
83. Darras S, Fritzenwanker JH, Uhlinger KR, Farrelly E, Pani AM, Hurley IA, Norris RP, Osovitz M, Terasaki M, Wu M, et al: **Anteroposterior axis patterning by early canonical Wnt signaling during hemichordate development.** *PLoS Biol* 2018, **16:**e2003698.
84. Wu HR, Chen YT, Su YH, Luo YJ, Holland LZ, Yu JK: **Asymmetric localization of germline markers Vasa and Nanos during early development in the amphioxus Branchiostoma floridae.** *Dev Biol* 2011, **353:**147-159.
85. Yue JX, Li KL, Yu JK: **Discovery of germline-related genes in Cephalochordate amphioxus: A genome wide survey using genome annotation and transcriptome data.** *Mar Genomics* 2015, **24 Pt 2:**147-157.
86. Person AD, Garriock RJ, Krieg PA, Runyan RB, Klewer SE: **Frzb modulates Wnt-9a-mediated beta-catenin signaling during avian atrioventricular cardiac cushion development.** *Dev Biol* 2005, **278:**35-48.
87. Garcia-Castro MI, Marcelle C, Bronner-Fraser M: **Ectodermal Wnt function as a neural crest inducer.** *Science* 2002, **297:**848-851.
88. Steventon B, Mayor R, Streit A: **Neural crest and placode interaction during the development of the cranial sensory system.** *Dev Biol* 2014, **389:**28-38.
89. Ohyama T, Mohamed OA, Taketo MM, Dufort D, Groves AK: **Wnt signals mediate a fate decision between otic placode and epidermis.** *Development* 2006, **133:**865-875.
90. Bassham S, Postlethwait JH: **The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate Oikopleura dioica.** *Development* 2005, **132:**4259-4272.
91. Manni L, Lane NJ, Joly JS, Gasparini F, Tiozzo S, Caicci F, Zaniolo G, Burighel P: **Neurogenic and non-neurogenic placodes in ascidians.** *J Exp Zoolog B Mol Dev Evol* 2004, **302:**483-504.
92. Schlosser G: **From so simple a beginning – what amphioxus can teach us about placode evolution.** *Int J Dev Biol* 2017, **61:**633-648.
93. Abitua PB, Wagner E, Navarrete IA, Levine M: **Identification of a rudimentary neural crest in a non-vertebrate chordate.** *Nature* 2012, **492:**104-107.

62 94. Lacalli TC: **Basic featu es of the ancestral chordate b ain: a protochordate**
963 **perspective.** *Brain Res Bull* 2008, **75:**319-323.
964 95. Robert N, Lhomond G, Schubert M, Croce JC: **A comprehensive survey of wnt**
965 **and frizzled expression in the sea urchin Paracentrotus lividus.** *Genesis*
966 2014, **52:**235-250.
967 96. Ruppert EE: **Morphology of Hatschek's nephridium in larval and juvenile**
968 **stages of *Branchiostoma virginiae* (cephalochordata).** *Israel Journal of*
969 *Zoology* 1996, **42:**S161-S182.
970 97. Bertrand S, Camasses A, Somorjai I, Belgacem MR, Chabrol O, Escande ML,
971 Pontarotti P, Escriva H: **Amphioxus FGF signaling predicts the acquisition of**
972 **vertebrate morphological traits.** *Proc Natl Acad Sci U S A* 2011, **108:**9160-
973 9165.
974 98. Shimeld SM, van den Heuvel M, Dawber R, Briscoe J: **An amphioxus Gli gene**
975 **reveals conservation of midline patterning and the evolution of hedgehog**
976 **signalling diversity in chordates.** *PLoS One* 2007, **2:**e864.
977 99. Gillis JA, Fritzenwanker JH, Lowe CJ: **A stem-deuterostome origin of the**
978 **vertebrate pharyngeal transcriptional network.** *Proc Biol Sci* 2012, **279:**237-
979 246.
980 100. Christiaen L, Jaszczyszyn Y, Kerfant M, Kano S, Thermes V, Joly JS: **Evolutionary**
981 **modification of mouth position in deuterostomes.** *Semin Cell Dev Biol* 2007,
982 **18:**502-511.
983 101. Meulemans D, Bronner-Fraser M: **The amphioxus SoxB family: implications**
984 **for the evolution of vertebrate placodes.** *Int J Biol Sci* 2007, **3:**356-364.
985 102. Pruitt MM, Letcher EJ, Chou HC, Bastin BR, Schneider SQ: **Expression of the wnt**
986 **gene complement in a spiral-cleaving embryo and trochophore larva.** *Int J*
987 *Dev Biol* 2014, **58:**563-573.
988 103. Bolognesi R, Farzana L, Fischer TD, Brown SJ: **Multiple Wnt genes are required**
989 **for segmentation in the short-germ embryo of Tribolium castaneum.** *Curr*
990 *Biol* 2008, **18:**1624-1629.
991 104. Hayden L, Schlosser G, Arthur W: **Functional analysis of centipede**
992 **development supports roles for Wnt genes in posterior development and**
993 **segment generation.** *Evol Dev* 2015, **17:**49-62.
994 105. Hogvall M, Schonauer A, Budd GE, McGregor AP, Posnien N, Janssen R: **Analysis**
995 **of the Wnt gene repertoire in an onychophoran provides new insights into**
996 **the evolution of segmentation.** *Evodevo* 2014, **5:**14.
997 106. Gabaldon T, Koonin EV: **Functional and evolutionary implications of gene**
998 **orthology.** *Nat Rev Genet* 2013, **14:**360-366.
999 107. Bolognesi R, Beermann A, Farzana L, Wittkopp N, Lutz R, Balavoine G, Brown SJ,
1000 Schroder R: **Tribolium Wnts: evidence for a larger repertoire in insects with**
1001 **overlapping expression patterns that suggest multiple redundant functions**
1002 **in embryogenesis.** *Dev Genes Evol* 2008, **218:**193-202.
1003 108. Janssen R, Posnien N: **Identification and embryonic expression of Wnt2,**
1004 **Wnt4, Wnt5 and Wnt9 in the millipede Glomeris marginata (Myriapoda:**
1005 **Diplopoda).** *Gene Expr Patterns* 2014, **14:**55-61.
1006 109. Constantinou SJ, Pace RM, Stangl AJ, Nagy LM, Williams TA: **Wnt repertoire and**
1007 **developmental expression patterns in the crustacean Thamnocephalus**
1008 **platyurus.** *Evol Dev* 2016, **18:**324-341.

110. Henry JQ, Perry KJ, Wever J, Seaver E, Martindale MQ: **Beta-catenin is required for the establishment of vegetal embryonic fates in the nemertean, Cerebratulus lacteus.** *Dev Biol* 2008, **317:**368-379.
111. Henry JQ, Perry KJ, Martindale MQ: **beta-catenin and early development in the gastropod, Crepidula fornicata.** *Integr Comp Biol* 2010, **50:**707-719.
112. Pang K, Ryan JF, Mullikin JC, Baxevanis AD, Martindale MQ: **Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore Mnemiopsis leidyi.** *Evodevo* 2010, **1:**10.
113. Doumpas N, Jekely G, Teleman AA: **Wnt6 is required for maxillary palp formation in Drosophila.** *BMC Biol* 2013, **11:**104.
114. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010, **59:**307-321.
115. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC bioinformatics* 2004, **5:**113.
116. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al: **Clustal W and Clustal X version 2.0.** *Bioinformatics (Oxford, England)* 2007, **23:**2947-2948.
117. Fuentes M, Schubert M, Dalfo D, Candiani S, Benito E, Gardenyes J, Godoy L, Moret F, Illas M, Patten I, et al: **Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity.** *J Exp Zoolog B Mol Dev Evol* 2004, **302:**384-391.
118. Satou Y, Kawashima T, Shoguchi E, Nakayama A, Satoh N: **An integrated database of the ascidian, Ciona intestinalis: towards functional genomics.** *Zoolog Sci* 2005, **22:**837-843.
119. Gilchrist MJ, Sobral D, Khoueiry P, Daian F, Laporte B, Patrushev I, Matsumoto J, Dewar K, Hastings KE, Satou Y, et al: **A pipeline for the systematic identification of non-redundant full-ORF cDNAs for polymorphic and evolutionary divergent genomes: Application to the ascidian Ciona intestinalis.** *Dev Biol* 2015, **404:**149-163.
120. Satou Y: **posterior end mark 3 (pem-3), an ascidian maternally expressed gene with localized mRNA encodes a protein with Caenorhabditis elegans MEX-3-like KH domains.** *Dev Biol* 1999, **212:**337-350.
121. Wada S, Katsuyama Y, Yasugi S, Saiga H: **Spatially and temporally regulated expression of the LIM class homeobox gene *Hrlim* suggests multiple distinct functions in development of the ascidian, *Halocynthia roretzi*.** *Mech Dev* 1995, **51:**115-126.
122. Brozovic M, Martin C, Dantec C, Dauga D, Mendez M, Simion P, Percher M, Laporte B, Scornavacca C, Di Gregorio A, et al: **ANISEED 2015: a digital framework for the comparative developmental biology of ascidians.** *Nucleic Acids Res* 2016, **44:**D808-818.
123. Brozovic M, Dantec C, Dardaillon J, Dauga D, Faure E, Gineste M, Louis A, Naville M, Nitta KR, Piette J, et al: **ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets.** *Nucleic Acids Res* 2017.
124. Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SA, et al: **ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics.** *Nucleic Acids Res* 2013, **41:**D854-860.

125. Bowes JB, Snyder KA, Segerdell E, Jarabek CJ, Azam K, Zorn AM, Vize PD: **Xenbase: gene expression and improved integration.** *Nucleic Acids Res* 2010, **38:**D607-612.
126. Karpinka JB, Fortriede JD, Burns KA, James-Zorn C, Ponferrada VG, Lee J, Karimi K, Zorn AM, Vize PD: **Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes.** *Nucleic Acids Res* 2015, **43:**D756-763.
127. James-Zorn C, Ponferrada VG, Burns KA, Fortriede JD, Lotay VS, Liu Y, Brad Karpinka J, Karimi K, Zorn AM, Vize PD: **Xenbase: Core features, data acquisition, and data processing.** *Genesis* 2015, **53:**486-497.
128. Richardson L, Venkataraman S, Stevenson P, Yang Y, Moss J, Graham L, Burton N, Hill B, Rao J, Baldock RA, Armit C: **EMAGE mouse embryo spatial gene expression database: 2014 update.** *Nucleic Acids Res* 2014, **42:**D835-844.
129. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453:**1064-1071.

Additional Files:

Additional file 1 includes:

- Figures S1 to S3.
- Table S1. Non-vertebrate chordate Wnt genes.
- Table S2. *Branchiostoma lanceolatum* and *Halocynthia roretzi* primer and probe sequences.
- Table S3. Genomic coordinates (scaffold or chromosome) for cephalochordate, human and lamprey *Wnt* genes.
- Text S1. *Branchiostoma lanceolatum Wnt* expression as shown in Figure 2.
- Text S2. References for Figure S3.

Additional file 2: Wnt sequence alignement for Figure 1.

Additional file 3: Wnt sequence alignment for Figure S4.

Figure Legends

Figure 1. Wnt family evolution in chordates. (A) The ML phylogenetic tree reveals a conservative pattern of genomic evolution in cephalochordates (species names in blue), which preserve all thirteen Wnt subfamilies, contrasting with the dynamic pattern of genomic

evolution in urochordates (in red), which are characterized by several gene losses and duplications. The scale bar indicates amino-acid substitutions. Values for the approximate likelihood-ratio test (aLRT) are shown at nodes. Species abbreviations: Chordate species: Urochordates: *Botryllus schlosseri* (Bsc), *Ciona savignyi* (Csa), *Ciona robusta* (Cro; formerly *Ciona intestinallis*), *Halocynthia roretzi* (Hro), *Halocynthia aurantium* (Hau), *Mogula occulta* (Moccu), *Mogula oculata* (Mocul), *Mogula occidentalis* (Mocci), *Phallusia fumigata* (Pfu) and *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (*Bbe*), *Branchiostoma floridae* (*Bfl*), *Branchiostoma lanceolatum* (*Bla*); Vertebrates: *Danio rerio* (Dre), *Homo sapiens* (*Hsa*). Non-chordates species: hemichordate *Saccoglossus kowalevskii* (Sko), annelid *Capitella teleta* (Cte), mollusk *Lottia gigantea* (Lgi), and cnidarian *Nematostella vectensis* (Nve). (B) Schematic representation of the *Wnt* gene catalogue (*Wnt1-16* and *A*) present (white squares) or absent (red squares) in the three chordate subphyla, allowing inferrence of plausible events of gene losses (red circles; the number or letter inside the circle indicates the lost Wnt subfamily) and duplications (black squares; the number inside the square indicates the duplicated subfamily) during the evolution of different lineages. While some losses appear to be ancestral (e.g. *Wnt8* and *Wnt4* losses in stem urochordates), others only affected specific groups or species (e.g. *Wnt1* loss in *Phlebobranchia*, *Wnt11* in *Molgulas*, and *Wnt3* to *Botryllus*). In vertebrates, Wnt subfamilies expanded following the rounds of whole-genome duplications (WGD) that occurred during their early evolution [4], while in the rest of chordates all *Wnt* are conserved as single copy, with the exception of *Wnt5* in Stolidobranchia ascidians, which has suffered multiple events of tandem gene duplications (see Additional file 1: Figure S1).

Figure 2. Complete atlas of *Wnt* expression in the cephalochordate *Branchiostoma lanceolatum*. Stage-matched expression patterns of 13 amphioxus *Wnt* genes reveal complex spatio-temporal choreography throughout embryogenesis. *Wnt* genes are ordered according to their developmental timing of expression: *Wnt1, Wnt8* and *Wnt11* are first expressed in the blastula (class A genes); *Wnt3, Wnt4, Wnt5* and *Wnt6* begin in the gastrula

(class B genes); *Wnt7, Wnt16, Wnt10, Wnt2* and *Wnt9* in the early neurula (N2; class C genes); and WntA in the mid-late neurula (N3 class D gene). All *Wnt* genes are expressed through to early larval stages (L1), prior to mouth opening. Arrowheads are colour-coded to match schematics based on germ layer origins: blue, ectoderm including tail fin (*fi*); dark blue, neural tube (*nt*) and cerebral vesicle (*cv*); yellow, endoderm derivatives including foregut (*fg*), hindgut (*hg*), preoral pit (*pp*), endostyle (*es*), club-shaped gland (*csg*) and gill slit primordium (*gs*); orange, mesendoderm; red: mesoderm derivatives including somites (*so*) and mesothelial cells (*me*); fuchsia, axial mesoderm derivatives including notochord (*no*); white, tailbud structures (*tb*, boxed) including neurenteric canal and chordoneural hinge; black, mouth primordium (*mo*). Note that only some somites are represented for the sake of clarity to permit visualisation of underlying tissues. For all genes, anterior is to the left and dorsal up; developmental stages are indicated at the top and embryo schematics are shown at the bottom. Lateral and dorsal views are shown, with the exception of blastulae (blast), for which only lateral views are represented, and gastrula stage G3, for which lateral and blastopore (white asterisk) views are shown. Dotted lines indicate planes of sectioning in larvae L1. Scale bars = 50 μ m. Please see Additional file 1: Text S1 for a detailed gene-by-gene description of all amphioxus *Wnt* expression patterns.

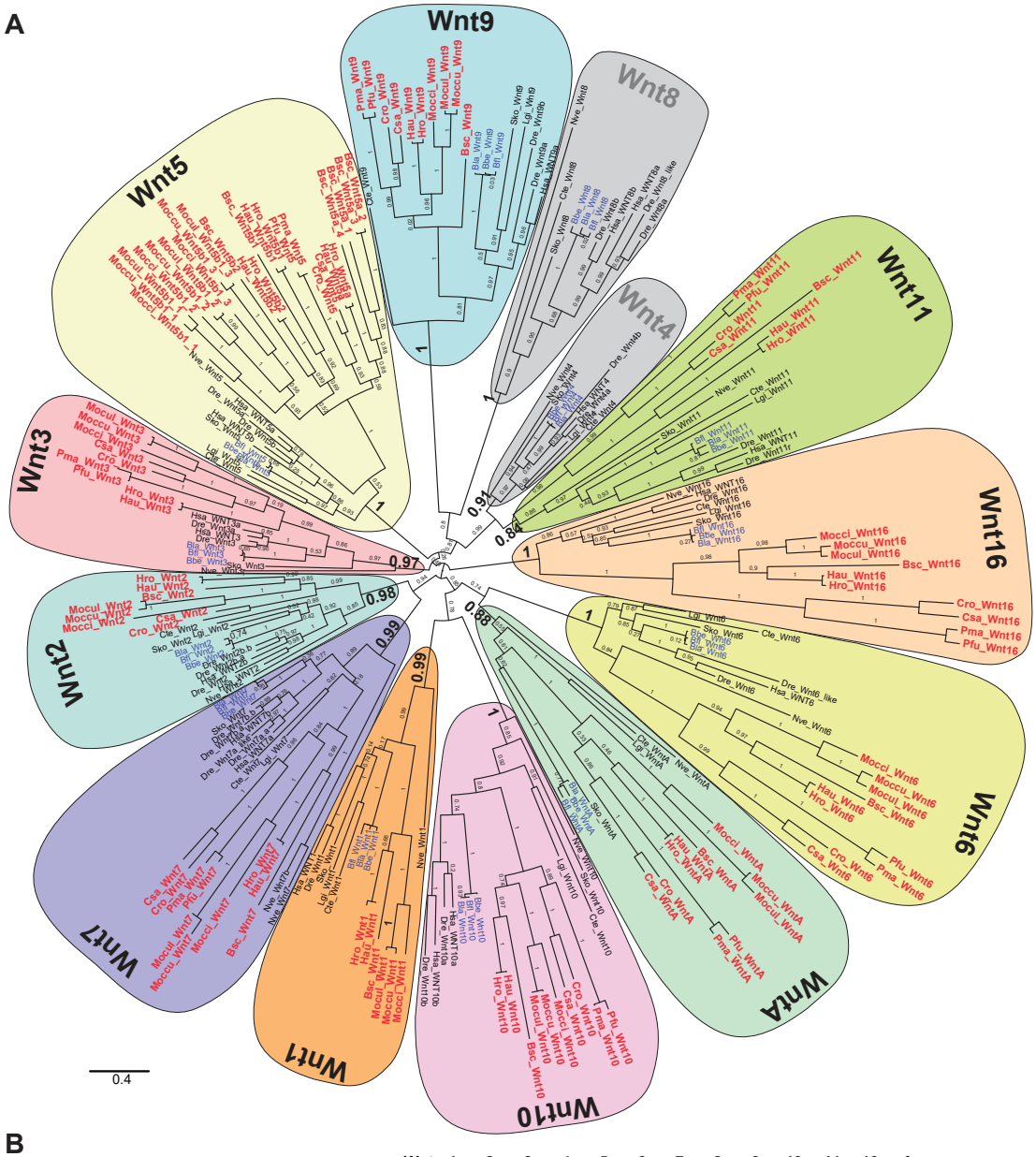
Figure 3. *Wnt* expression in 1 gill-slit amphioxus larvae. (A) Anterior expression domains of selected *Wnt* genes ordered as in Fig. 2: *Wnt8*: posterior cerebral vesicle (black arrowhead), preoral pit (pp), endostyle and around the mouth; *Wnt4:* posterior cerebral vesicle (black arrowhead) and neural tube; *Wnt5*: cerebral vesicle (black arrowhead), club-shaped gland (cg) and weakly in anterior notochord; *Wnt6*: isolated spots in neural tube (black arrowhead); *Wnt7*: posterior cerebral vesicle (black arrowhead); *Wnt16*: neural tube, including the hindbrain (black arrowheads); *Wnt10:* posterior cerebral vesicle (black arrowhead); *Wnt2*: anterior notochord (black arrow) and isolated cells of the anterior endoderm; *Wnt9*: posterior cerebral vesicle (black arrowhead), pharyngeal endoderm around the preoral pit, endostyle and the first gill slit primordium (black arrow); *WntA*: around the

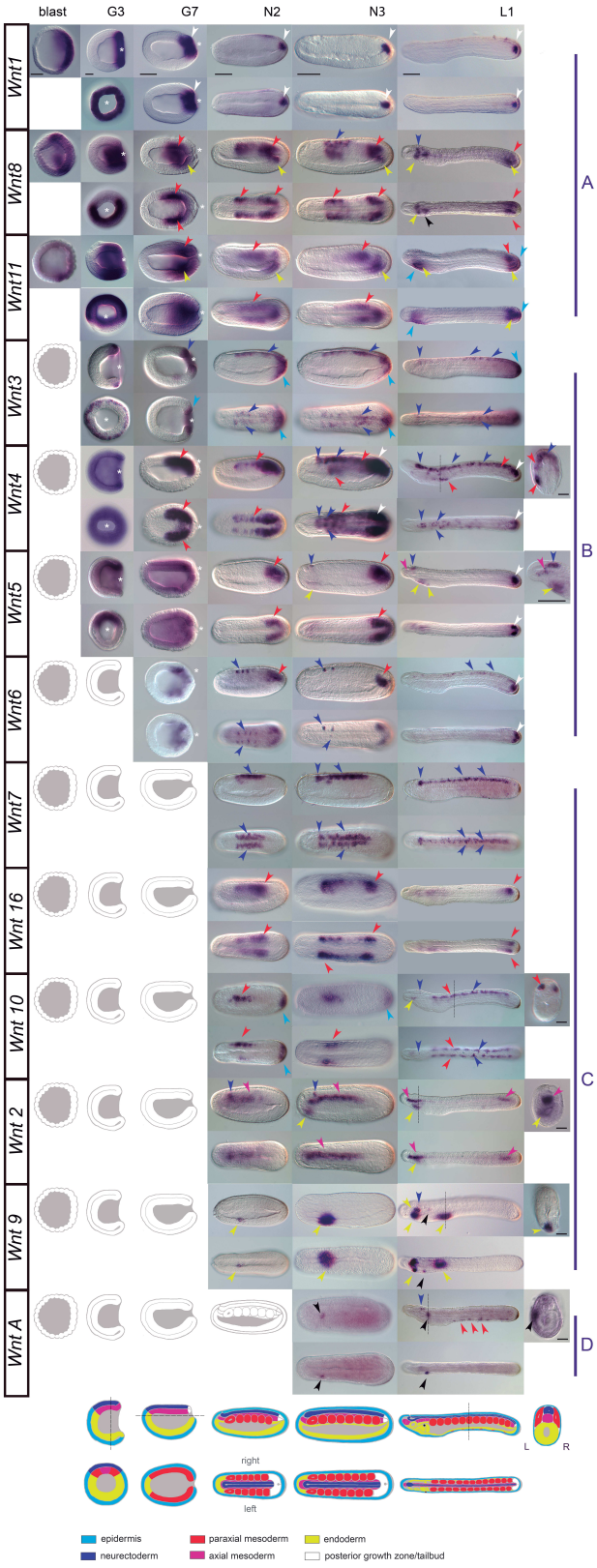
mouth and in the first forming gill slit (black arrow) and within the rostral coelom and muscle. Black asterisk: pigment spot in the forming frontal eye; white dotted line: mouth. (B) Posterior expression domains. *Wnt1*: posterior neural tube (white arrowheads) and neurenteric canal hinge (black arrowhead); *Wnt11*: tail fin (black arrowhead), *Wnt5*: tailbud region, *Wnt6*: neural tube (black arrow) and posterior wall of the neurenteric canal (black arrowhead); *Wnt16*: posterior wall of the neurenteric canal and posterior mesoderm of the last somites (black arrowhead); *Wnt2*: posterior notochord (black arrowhead); *Wnt9*: midgut endoderm (black arrowhead); *WntA* : mesothelial cells (black arrowhead). Older larvae have tails fins containing brown pigment. All views are lateral, with anterior to the left and posterior to the right.

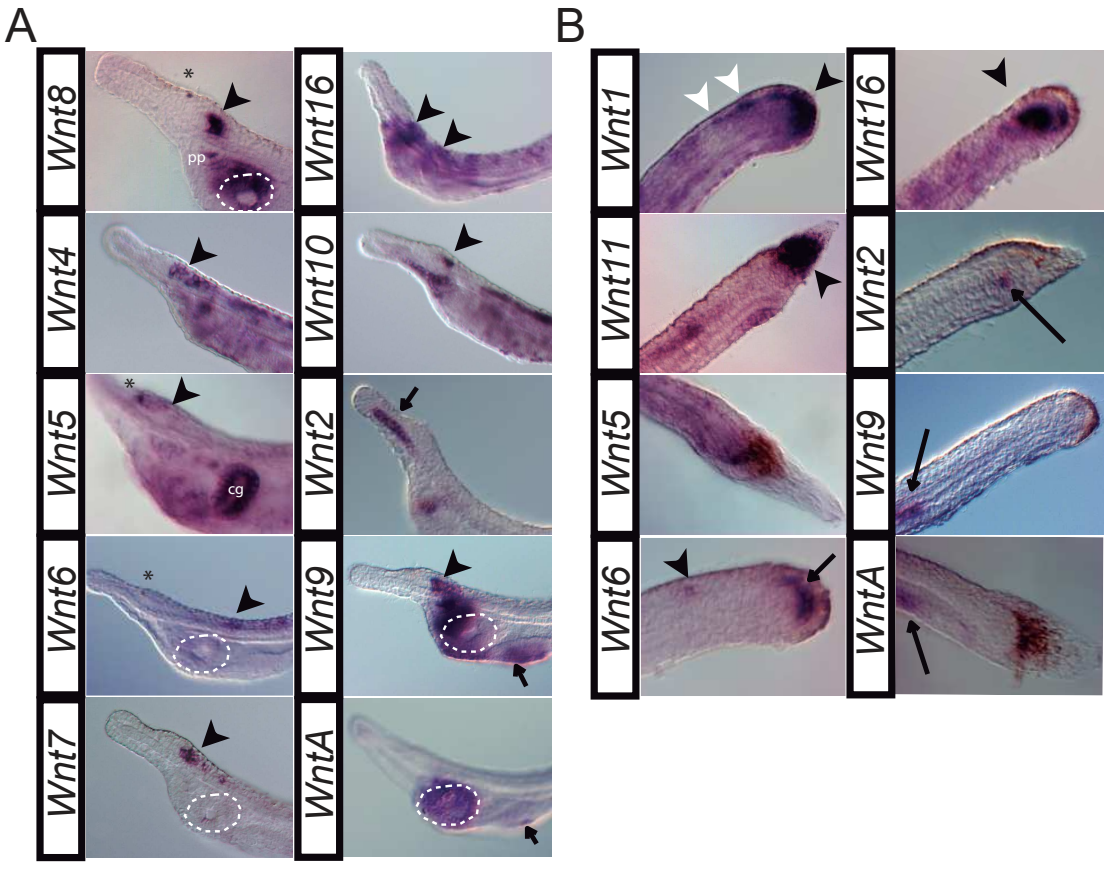
Figure 4. *Wnt* expression in *B.lanceolatum* is complex and spatio-temporally dynamic. Colour-coded horizontal bars represent the expression domains of *Wnt* genes (see left box), and refer to schematic illustrations of *B. lanceolatum* embryos from blastula to early larva (lateral view at the top and dorsal view at the bottom; anterior is to the left and posterior to the right; main structures are labelled), organised by germ layer into endoderm (yellow), mesoderm (red), chordomesoderm (magenta), and ectoderm (dark blue: neural derivatives; light blue: epidermal derivatives). Vertical dotted lines delimit antero-posterior reference domains, and underscore hypothesised Wnt functions during neural regionalisation, posterior growth and structure differentiation. # denotes minor differences with *B. floridae* expression. Abbreviations are as follows: ec=ectoderm; men=mesendoderm; ep=epidermis; ch=chordal plate; np=neural plate; no=notochord; cv=cerebral vesicle; fo=foregut; nt=neural tube; so=somites; hi=hindgut; tb=tailbud; pp=preoral pit; es=endostyle; csg=club-shaped gland; gs=gill slit primordium; my=myomere; mec=mesothelial cells; fi=fin; mo=mouth primordium. See text for further details.

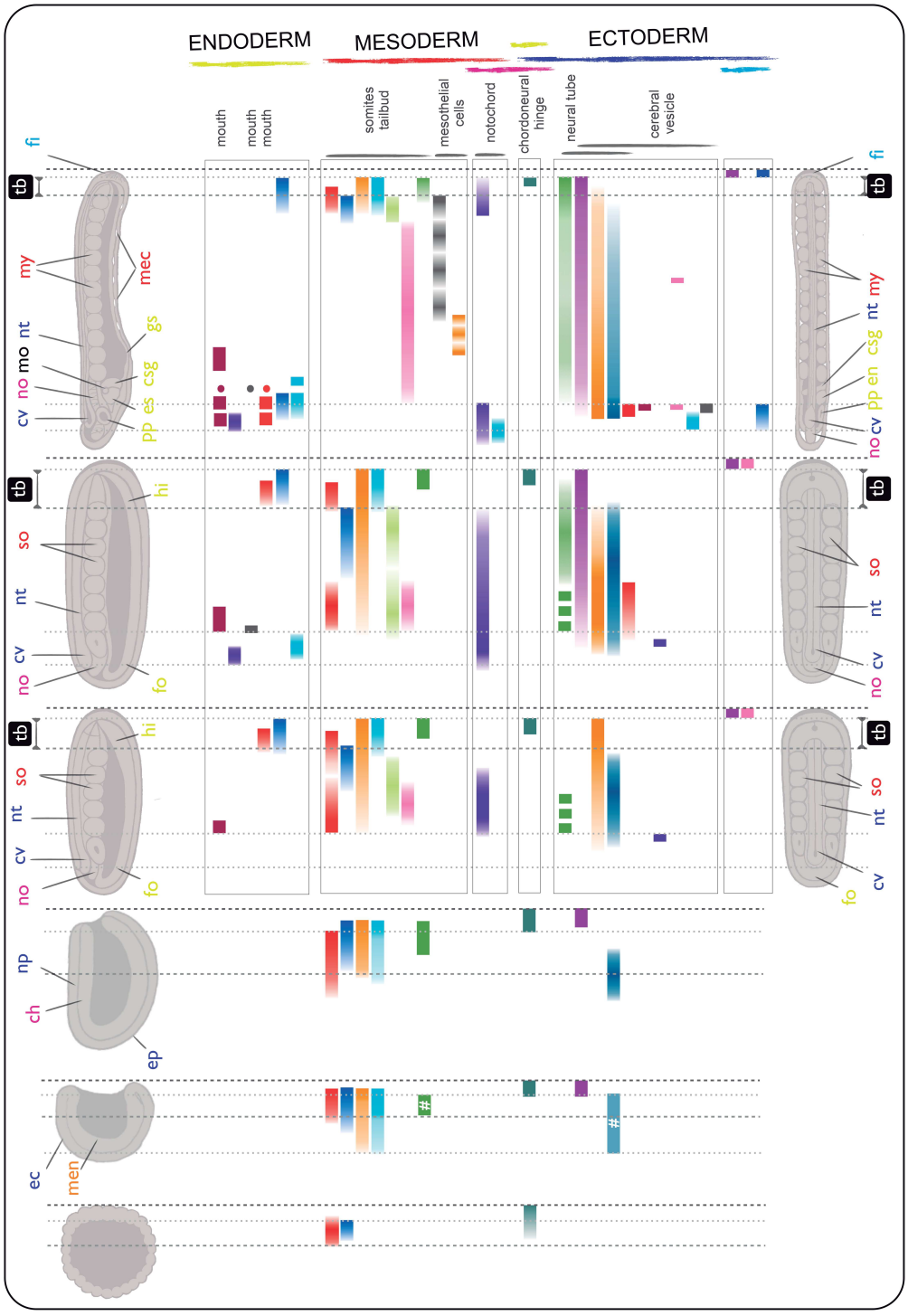
Figure 5. Synteny of *Wnt* genes in chordates. Genes and their relative position and orientation on chromosomes or scaffolds (numbers above lines) are represented by block

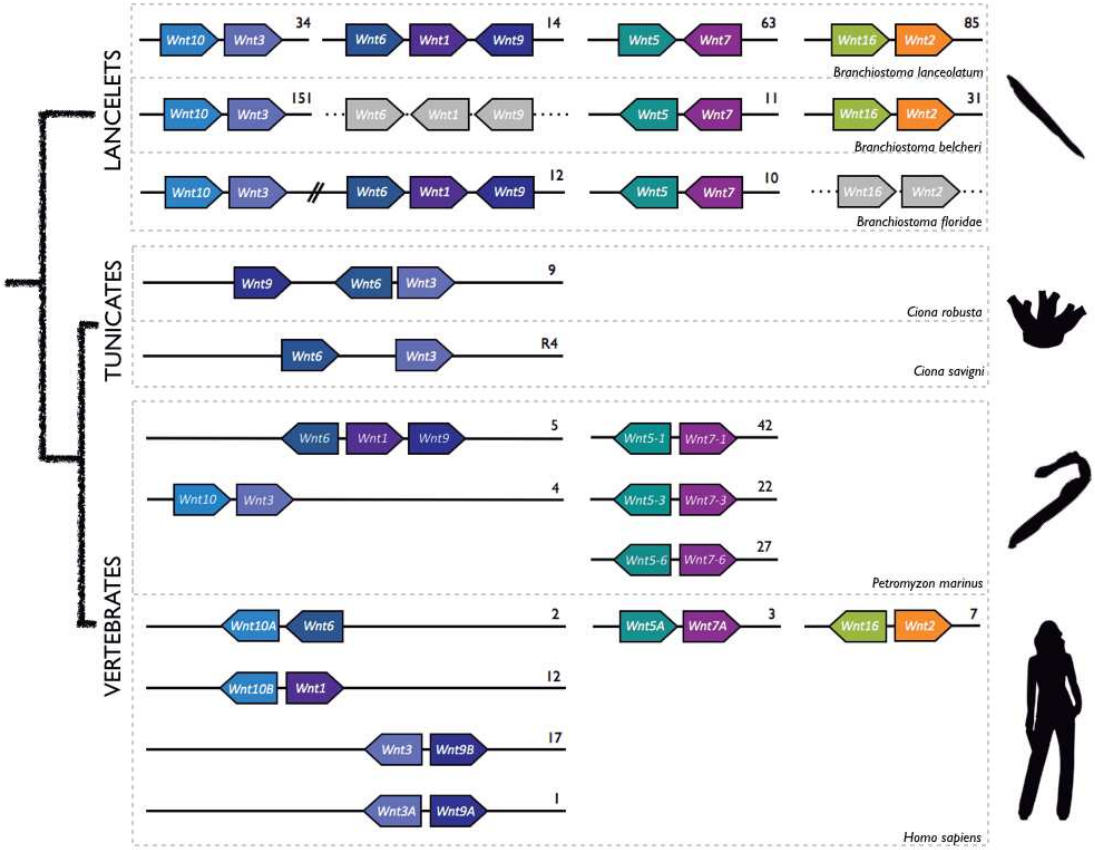
arrows, colour- coded by *Wnt* gene. Grey arrows on dotted lines indicate *Wnt* genes for which linkage has so far not been demonstrated. The phylogenetic relationship of chordate subphyla is shown on the left. Lancelets are represented by *B. floridae, B. belcheri* and *B. lanceolatum*. Among tunicates, genome arrangements in *C.savigni* and *C.robusta* are shown. *P.marinus* and *H.sapiens* are vertebrate representatives. The data presented for *B.floridae* have been modified and extended from [25, 129]; this is the only cephalochordate for which *Wnt10-Wnt3* and *Wnt6-Wnt1-Wnt9* syntenic groups are linked on the same scaffold. Representative taxa are illustrated on the right (black silhouettes).


A
Wnt9
Wnt8
Wnt5
Wnt4
Wnt11
Wnt3
Wnt16
Wnt2
Wnt6
Wnt7
WntA
Wnt1
Wnt10
0.4
B
Wnt 1 2 3 4 5 6 7 8 9 10 11 16 A Total
Vertebrates
WGD
Homo sapiens
Ciona robusta
Ciona savignyi
Phallusia fumigata
Phallusia mammillata
Halocynthia roretzi
Halocynthia aurantium
Botryllus schlosseri
Molgula occulta
Molgula oculata
Molgula occidentalis
Branchiostoma lanceolatum
Branchiostoma floridae
Branchiostoma belcheri
Chordates
Uro-chordates
Cephalochordates
Phlebobranchia
Stolidobranchia


blast
G3
G7
N2
N3
L1
Wnt1
Wnt8
Wnt11
Wnt3
Wnt4
Wnt5
Wnt6
Wnt7
Wnt 16
Wnt 10
Wnt 2
Wnt 9
Wnt A
A
B
C
D
L
R
right
left
epidermis
neurectoderm
paraxial mesoderm
axial mesoderm
endoderm
posterior growth zone/tailbud


A
Wnt8
Wnt4
Wnt5
Wnt6
Wnt7
Wnt16
Wnt10
Wnt2
Wnt9
WntA
pp
cg
B
Wnt1
Wnt11
Wnt5
Wnt6
Wnt16
Wnt2
Wnt9
WntA


ENDODERM
MESODERM
ECTODERM
mouth
mouth
mouth
somites
tailbud
mesothelial
cells
notochord
chordoneural
hinge
neural tube
cerebral
vesicle
fi
tb
my
mec
nt
mo
no
gs
csg
cv
es
pp
hi
so
nt
cv
no
fo
np
ch
ep
ec
men
#
en
Wnt1
Wnt2
Wnt3
Wnt4
Wnt5
Wnt7
Wnt8
Wnt6
Wnt9
Wnt10
Wnt11
Wnt16
WntA


LANCELETS
TUNICATES
VERTEBRATES
Wnt10
Wnt3
34
Wnt6
Wnt1
Wnt9
14
Wnt5
Wnt7
63
Wnt16
Wnt2
85
Branchiostoma lanceolatum
151
11
31
Branchiostoma belcheri
12
10
Branchiostoma floridae
9
Ciona robusta
R4
Ciona savigni
5
42
4
22
27
Petromyzon marinus
Wnt10A
Wnt10B
2
12
17
1
Wnt5A
Wnt7A
3
7
Homo sapiens

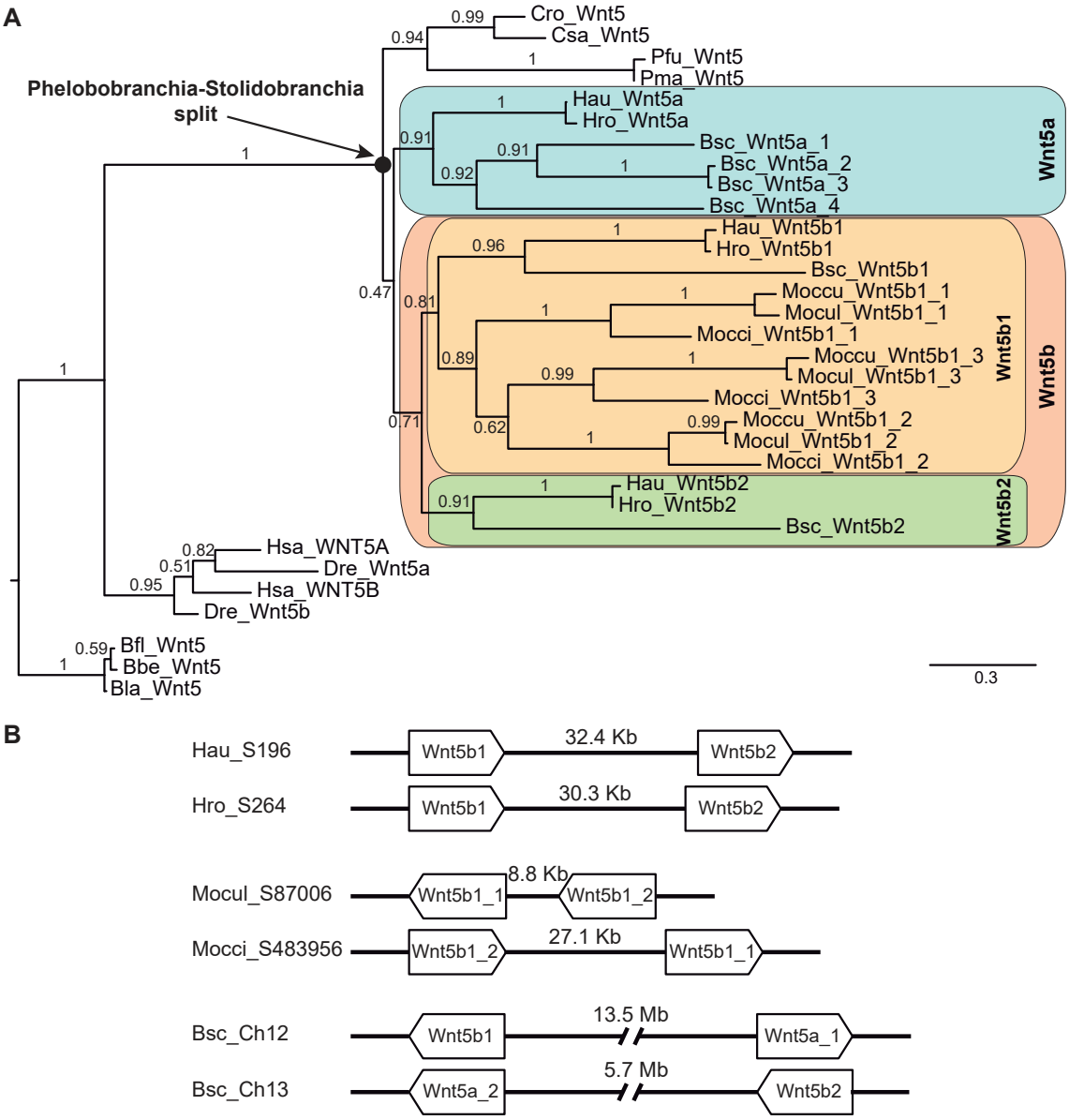


Figure S1. Evolution of *Wnt5* in ascidians. A. The ML phylogenetic tree of Wnt5 subfamily in ascidians suggests the presence of multiple paralogues (*Wnt5a* and *Wnt5b*) which originated in the Stolidobranchia clade after its split from Phlebobranchia. **B.** The fact that many of these *Wnt5* duplicates appeared to be located in the same genomic regions suggested that they originated by tandem gene duplications. Despite the fact that the number of *Wnt5* genes is the same in species of the *Molgula* and *Halocynthia* genera, the tree topology suggests that some of the duplications occurred after the splitting of the two genera (e.g. *Wnt5b1* and *Wnt5b2*), but before speciation within each group, in which further duplications independently occurred in different lineages (although ancestral *Wnt5* duplications in stem Stolidobranchia, followed by multiple gene losses and events of gene conversion in different species cannot be discarded). In any case, the expansion of *Wnt5* in Stolidobranchia provides a singular case of Wnt subfamily amplification in non-vertebrate chordates, suggesting that the evolution of this order of ascidian species has been accompanied by a relaxation of the evolutionary constraints that maintain *Wnt5* genes as single copy gene in other species. This may be linked to the recruitment of new *Wnt5* paralogues in biological innovations unique to this group of ascidians. Species abbreviations are as in Figure 1.

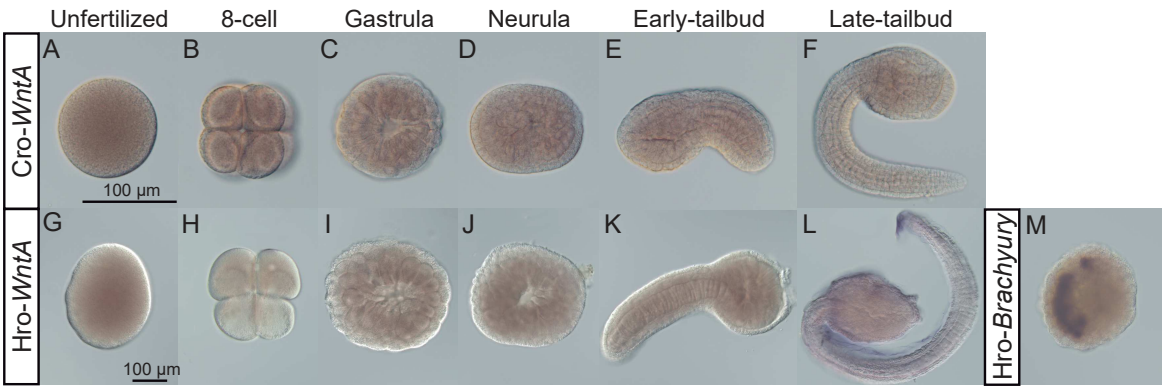


Figure S2. Expression of *WntA* in two ascidian species. Whole-mount *in situ* hybridization (WMISH) for *WntA* in *C. robusta* (Cro) (A-F) and *H. roretzi* (Hro) (G-L), and for the *Brachyury* gene in *H. roretzi* (M) as the positive WMISH control. No expression of *WntA* was observed in any selected stages –including unfertilized eggs, 8-cell, gastrula, neurula, early-tailbud and late-tailbud stages (a weak non-specific signal was observed in the tunic of *H. roretzi* late-tailbud larvae)–, whereas *Brachyury* expression was clearly detected in A-line and B-line cells (M). The lack of expression of *WntA* was consistent with the absence of ESTs from embryonic libraries in databases of *C. robusta* and *H. roretzi*.

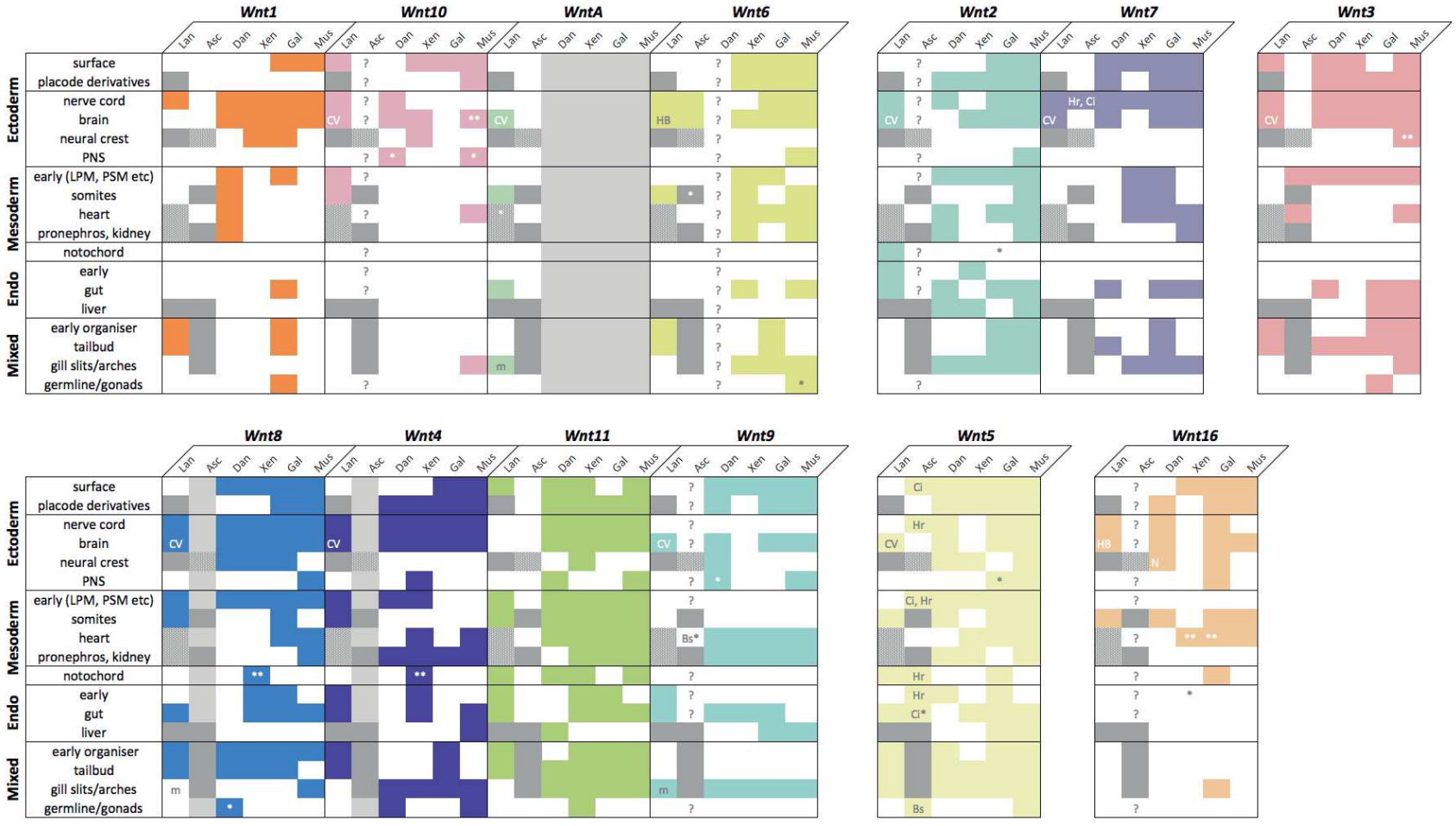

Wnt1
Wnt10
WntA
Wnt6
Wnt2
Wnt7
Wnt3
Wnt8
Wnt4
Wnt11
Wnt9
Wnt5
Wnt16
Lan
Asc
Dan
Xen
Gal
Mus
Ectoderm
Mesoderm
Endo
Mixed
surface
placode derivatives
nerve cord
brain
neural crest
PNS
early (LPM, PSM etc)
somites
heart
pronephros, kidney
notochord
early
gut
liver
early organiser
tailbud
gill slits/arches
germline/gonads
CV
HB
Hr, Ci
Ci, Hr
Bs*
Ci*
Hr
Ci
Bs
N
m

Figure S3. Chordate Wnt expression. Comparative expression for all Wnt subfamilies in chordates from data derived from our own results, published literature (see Additional file 1: Text S2 for references) and public databases (see Material and Methods section). Filled coloured boxes (Wnt subfamilies are coloured and grouped according to significant nodes in Fig. 1) denote documented expression; white filled boxes indicate absence of, or unreported expression; question marks are reserved for genes that are present, but whose expression has never been assessed; pale grey boxes represent lost genes (e. g. WntA in vertebrates or Wnt4 and Wnt8 in ascidians); dark grey boxes denote tissues/structures that are absent in the subphylum and/or at examined stages, and therefore, expression comparison is not possible (e.g. placode derivatives in amphioxus or somites in ascidians); stippled grey boxes highlight tissues/structures lacking clear homologues (e.g. heart in amphioxus or neural crest in ascidians). Expression patterns for vertebrate paralogues of the same Wnt subfamily were combined. Single asterisks (by gene order): *Wnt10* Dan: *Wnt10a* expression in neuromasts; *Wnt10* Mus: expression in ganglia; *WntA* Lan: a true heart does not exist but domains thought to be homologous have been identified with expression; *Wnt6* Asc: based on *WntE* orphan expression in muscle blocks (no somites present) in Imai et al (2004); *Wnt6* Mus: genital ridge; *Wnt2* Gal: convincing WMISH data in *Geisha* contradict a previous report of expression for which sequence ID could not be confirmed; *Wnt8* Dan: oocyte; *Wnt9* Asc (Bs): a true heart does not exist in colonial ascidians but expression in vasculature may reflect homology; *Wnt9* Dan: cranial ganglia; *Wnt5* Asc (Ci): endodermal strand; *Wnt5* Gal: ganglia; *Wnt16* Xen: hypochord. Double asterisks (by gene order): *Wnt10* Mus: various sources including knock-in expression reporter for *Wnt10a*; *Wnt3* Mus: indirect (loss of function); *Wnt8* Xen: *Xenbase* reports expression but this requires confirmation; *Wnt4* Xen: *Xenbase* reports expression but this requires confirmation; *Wnt16* Xen: suggested by Zhang et al (2011); *Wnt16* Gal: *Geisha* reports expression. Abbreviations: Lan=lancelet; Asc=ascidian; Dan=zebrafish *Danio rerio*; Xen=frog *Xenopus laevis/tropicalis*; Gal=chicken *Gallus gallus*; Mus=mouse *Mus musculus*; Ci=*Ciona intestinalis*; Hr=*Halocynthia roretzi*; Bs=*Botryllus schlosseri*; PNS (ectoderm)=peripheral nervous system; LPM/PSM (mesoderm)=lateral plate/presomitic mesoderm; CV=cerebral vesicle; HB=hindbrain; m=mouth.

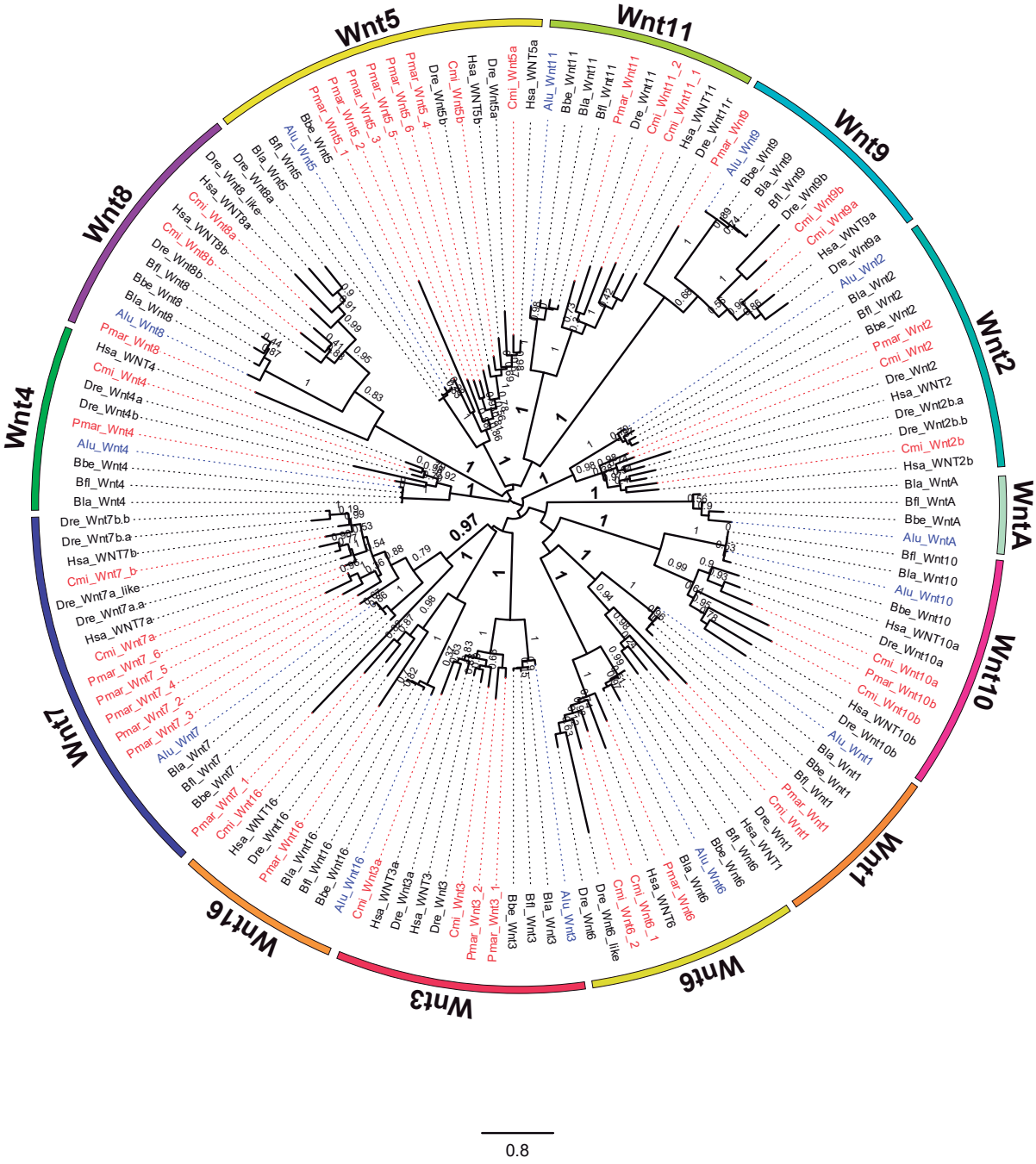

Wnt5
Wnt11
Wnt9
Wnt2
WntA
Wnt10
Wnt1
Wnt6
Wnt3
Wnt16
Wnt7
Wnt4
Wnt8
0.8

Figure S4. Wnt subfamilies in *A. lucayanum, P. marinus* and *C. milii*. The ML phylogenetic tree reveals that each Wnt sequence from the cephalochordate *A. lucayanum* (in blue) grouped inside one of the 13 Wnt cephalochordate subfamilies, including the WntA subfamily, extending further the idea of genomic evolutionary stasis of cephalochordate species. In contrast, lamprey *P. marinus* and shark *C. milii* have Wnt representatives (in red) in all Wnt subfamilies except the WntA subfamily, suggesting that *WntA* was already lost before the divergence of jawless and gnathostome lineages. Notice that *Wnt5* and *Wnt7* have duplicated extensively in *P. marinus*. The scale bar indicates amino-acid substitutions. Values for the approximate likelihood-ratio test (aLRT) are shown at nodes. Species abbreviations: Cephalochordate species: *Asymmetron lucayanum* (*Alu*), *Branchiostoma belcheri* (*Bbe*), *Branchiostoma floridae* (*Bfl*), *Branchiostoma lanceolatum* (*Bla*); vertebrate species: *Petromyzon marinus* (*Pmar*), *Callorhinchus milii* (*Cmi*), *Danio rerio* (Dre), *Homo sapiens* (*Hsa*).

Table S1. Chordate *Wnt* genes analyzed in this study.

Subphylum	Gene	Previous names	Accession number*	ESTs/cDNAs*
Cephalochordata				
B. lanceolatum	*Wnt1*		BL08374	-
	Wnt2	*Wnt2b*†	BL18396	-
	Wnt3		ACE79725 (BL00565)	EU685300
	Wnt4		BL15100	-
	Wnt5	*Wnt5a*†	ACE79726 (BL05330)	EU685301
	Wnt6		ACE79727 (BL13401)	EU685302
	Wnt7	*Wnt7b*†	ACE79728 (BL12283)	EU685303
	Wnt8	*Wnt8a*†	BL09252	-
	Wnt9	*Wnt9a*†	BL22609	-
	Wnt10	*Wnt10a*†	BL23379	-
	Wnt11		BL18893	-
	Wnt16		BL18405	-
	WntA	*Wnt4*†	BL17431	-
B. floridae	*Wnt1*		AAC80432	FE564954, BW880884
	Wnt2		XP_002601759‡	-
	Wnt3		AAL37555	BW814403
	Wnt4		AAC80431	AF061973
	Wnt5		AAL37556	BW873866, FE573306, BW890381
	Wnt6		XP_002598625	-
	Wnt7	*Wnt7b* [1]	XP_002597288	BW953131, BW847525, BW859837, AF061975
	Wnt8		AAF80559	FE571808, FE547238, FE560742
	Wnt9	*Wnt14* [2]	XP_002598627	-
	Wnt10		XP_002598506 + XP_002598508 + EST FE577715.1‡	FE577715
	Wnt11		AAF80555	FE561860, FE553982, BW840509
	Wnt16		XP_002599799 + EST BW874763.1‡	BW874763
	WntA		XP_002609873	-
B. belcheri	*Wnt1*		Bb_306640F‡	HO762011
	Wnt2	*Wnt2b*†	Bb_073170F	JZ717686
	Wnt3	*Wnt3a*†	Bb_177200F‡	-
	Wnt4	-	Bb_271880F‡	-
	Wnt5	*Wnt5b*†	Bb_039350R+ Bb_039340R	-
	Wnt6	-	Bb_258720R	-
	Wnt7	*Wnt7b*†	Bb_039420R	-
	Wnt8	*Wnt8b*†	Bb_312060F	-
	Wnt9	*Wnt9a*†	Bb_185850F	-
	Wnt10	*Wnt10a*†	Bb_177150F	-
	Wnt11	*Wnt11b*†	Bb_034950F	-
	Wnt16	-	Bb_072770F	-
	WntA	*Wnt4*†	Bb_308600R	-
A. lucayanum	*Wnt1*		-	GETC01120369
	Wnt2		-	GETC01104091
	Wnt3		-	GETC01134136
	Wnt4		-	GETC01091176
	Wnt5		-	GESY01045804
	Wnt6		-	GETC01110690
	Wnt7		-	GETC01096614
	Wnt8		-	GETC01060203
	Wnt9		-	GETC01034191
	Wnt10		-	GETC01128999
	Wnt11		-	GETC01110192
	Wnt16		-	GETC01108796

	WntA		LZCU01110230‡	-
Urochordata				
Order Enterogona				
Family Cionidae				
C. robusta (formerly *C. intestinalis*)	*Wnt2*	*Wnt2* [3] *Orphan Wnt-b* [4]	NP_001071794	BW191620 BW482592
	Wnt3	*Wnt3* [3, 4] *Wnt3*; *Wnt10*†	XP_009859675	FF728330 BW498024
	Wnt5	*Wnt5* [3, 4]	NP_001027951	BW233701 BW479449
	Wnt6	*Orphan Wnt-e* [4] *Wnt-b* [3], *Wnt11/1*†	NP_001071795	K048593 BW096149
	Wnt7	*Wnt7*† [3, 4]	XP_002128034	FF804473 BW311357
	Wnt9	*Wnt9/14/15* [3] *Wnt14/15*† [4]	XP_009859770‡	BW504071
	Wnt10	*Wnt10/12* [3] *Wnt10a/12* [4]	XP_002127850	FF855596 FF912944
	Wnt11	*Orphan Wnt-a* [4] *Wnt-a* [3], *Wnt8 putative*†	NP_001028176‡	BP006149
	Wnt16	*Wnt-c* [3] Orphan Wnt-d [4]	XP_002122330	-
	WntA	*Orphan Wnt-c* [4] *Wnt-d* [3], *Wnt6 putative*†	XP_002121451	FF969783 FF969784
C. savignyi	*Wnt2*	-	Cisavi.CG.ENS81.R54. 2835750-2851101	-
	Wnt3	-	Cisavi.CG.ENS81.R4. 332380-338472‡	BW522473 BW552601
	Wnt5	-	Cisavi.CG.ENS81.R285. 233489-241411	BW573062 BW582532
	Wnt6	-	Cisavi.CG.ENS81.R4. 1192784-1201725	-
	Wnt7	-	Cisavi.CG.ENS81.R65. 368155-376728‡	BW591084
	Wnt9	-	Cisavi.CG.ENS81.R55. 213874-219426‡	-
	Wnt10	-	Cisavi.CG.ENS81.R926. 20876-28713	-
	Wnt11	-	Cisavi.CG.ENS81.R370. 21902-27494 Cisavi.CG.ENS81.R370. 21902-27525‡	-
	Wnt16	-	R48 (708195..713372)‡	-
	WntA	-	Cisavi.CG.ENS81.R19. 2776476-2790418‡	-
Family Ascidiidae				
P. mammilata	*Wnt2*	-	phmamm.CG.MTP2014. S84.g02811	AHC0AAA87YF13_ RM1
	Wnt3	-	S291 (136784..144112)‡	-
	Wnt5	-	phmamm.CG.MTP2014. S428.g08617‡	-
	Wnt6	-	phmamm.CG.MTP2014. S211.g05449	-
	Wnt7	-	phmamm.CG.MTP2014. S528.g09785‡	-
	Wnt9	-	S498 (49716..41376)‡	-
	Wnt10	-	S537 (95161..101893)‡	-
	Wnt11	-	S667 (23720..26900)‡	-
	Wnt16	-	phmamm.CG.MTP2014. S12.g00518 + g00517‡	-
	WntA	-	S958 (26992..14509)‡	-
P. fumigata	*Wnt2*	-	S38291‡	-
	Wnt3	-	S37373 (311..151) + S21003 (2454..478) +	-

			S21494 (914..1724)‡	
	Wnt5	-	S4452 (12623..3319)‡	-
	Wnt6	-	S1558 (2785..2204) + S12084 (4220..3294) + S1416 (453..4656)‡	-
	Wnt7	-	S553 (9591..1182)‡	-
	Wnt9	-	S2907 (11436..16644) + S4488 (12402..10747)‡	-
	Wnt10	-	S770 (23692..17150)‡	-
	Wnt11	-	S937 (16989..19608)‡	-
	Wnt16	-	S6225 (4783..3870) + S12301 (557..4227)‡	-
	WntA	-	S18 (28331..15021)‡	-
Order Stolidobranchia				
Family Pyuridae				
H. roretzi	*Wnt1*	-	Harore.CG.MTP2014. S634.g12371	DB583287 DB596795
	Wnt2	-	Harore.CG.MTP2014. S441.g09992	FY853847
	Wnt3	-	Harore.CG.MTP2014. S141.g00494‡	-
	Wnt5a	-	Harore.CG.MTP2014. S15.g14311	DB582570
	Wnt5b1	*HrWnt-5* [5]	Harore.CG.MTP2014. S264.g00180	FY863897 FY857344
	Wnt5b2	*HrWnt-5b* [6]	Harore.CG.MTP2014. S264.g00833	-
	Wnt6	-	Harore.CG.MTP2014. S139.g04292‡	-
	Wnt7	*HrWnt-7* [7]	Harore.CG.MTP2014. S12.g13523‡	-
	Wnt9	-	Harore.CG.MTP2014. S44.g02020‡	-
	Wnt10	-	Harore.CG.MTP2014. S118.g10881‡	-
	Wnt11	-	Harore.CG.MTP2014. S13.g11848	-
	Wnt16	-	Harore.CG.MTP2014. S355.g12638	-
	WntA	-	Harore.CG.MTP2014. S96.g13623‡	-
H. aurantium	*Wnt1*	-	Haaura.CG.MTP2014. S1095.g07586	-
	Wnt2	-	Haaura.CG.MTP2014. S270.g03873	-
	Wnt3	-	Haaura.CG.MTP2014. S461.g05183‡	-
	Wnt5a	-	Haaura.CG.MTP2014. S607.g05926	-
	Wnt5b1	-	Haaura.CG.MTP2014. S196.g03193	-
	Wnt5b2	-	Haaura.CG.MTP2014. S196_g03196	-
	Wnt6	-	Haaura.CG.MTP2014. S15.g00465‡	-
	Wnt7	-	Haaura.CG.MTP2014. S121.g02336‡	-
	Wnt9	-	Haaura.CG.MTP2014. S742.g06492‡	-
	Wnt10	-	S325 (666452..66801) + Haaura.CG.MTP2014. S426.g04986‡	-
	Wnt11	-	Haaura.CG.MTP2014.	-

			S64.g01495	
	Wnt16	-	Haaura.CG.MTP2014.S1551.g08535‡	-
	WntA	-	Haaura.CG.MTP2014.S1507.g08466‡	-
Family Molgulidae				
M. occulta	*Wnt1*	-	S716238 (22921..26525)‡	-
	Wnt2	-	S628454 (20043..26806)‡	-
	Wnt3	-	S655179 (2543..5965)‡	-
	Wnt5b1_1	-	S335521 (2809..8148)‡	-
	Wnt5b1_2	-	S211732 (585..2328)‡	-
	Wnt5b1_3	-	S400099 (9221..14111)‡	-
	Wnt6	-	S496787 (907..4017)‡	-
	Wnt7	-	S713940 (4242..7631)‡	-
	Wnt9	-	S379508 (8681..5606)‡	-
	Wnt10	-	S669497 (5717..2510)‡	-
	Wnt16	-	S321748 (11895..19107)‡	-
	WntA	-	S694937 (23190..26531)‡	-
M. oculata	*Wnt1*	-	Moocul.CG.ELv1_2.S108579.g10924	-
	Wnt2	-	Moocul.CG.ELv1_2.S29028.g01331	-
	Wnt3	-	Moocul.CG.ELv1_2.S127564.g14939	-
	Wnt5b1_1	-	Moocul.CG.ELv1_2.S87006.g06727	-
	Wnt5b1_2	-	Moocul.CG.ELv1_2.S87006.g06728‡	-
	Wnt5b1_3	-	Moocul.CG.ELv1_2.S109151.g11115	-
	Wnt6	-	Moocul.CG.ELv1_2.S129770.g15227	-
	Wnt7	-	Moocul.CG.ELv1_2.S117709.g13989	-
	Wnt9	-	Moocul.CG.ELv1_2.S127563.g14931 + g14932‡	-
	Wnt10	-	Moocul.CG.ELv1_2.S127563.g14924‡	-
	Wnt16	-	Moocul.CG.ELv1_2.S75012.g05236‡	-
	WntA	-	S113814 (30817..34341)‡	-
M. occidentalis	*Wnt1*	-	Moocci.CG.ELv1_2.S415796.g12520	-
	Wnt2	-	S643507 (63317..53607)‡	-
	Wnt3	-	Moocci.CG.ELv1_2.S649914.g30158+ g30159‡	-
	Wnt5b1_1	-	Moocci.CG.ELv1_2.S483956.g17278‡	-
	Wnt5b1_2	-	Moocci.CG.ELv1_2.S483956.g17276	-
	Wnt5b1_3	-	Moocci.CG.ELv1_2.S478739.g16781‡	-
	Wnt6	-	Moocci.CG.ELv1_2.S482528.g17137‡	-
	Wnt7	-	Moocci.CG.ELv1_2.S262076.g06136‡	-

	Wnt9	-	Moocci.CG.ELv1_2.S494455.g18504‡	-
	Wnt10	-	Moocci.CG.ELv1_2.S386552.g11160‡	-
	Wnt16	-	Moocci.CG.ELv1_2.S645945.g29523‡	-
	WntA	-	Moocci.CG.ELv1_2.S625464.g26812‡	-
Family Styelidae				
B. schlosseri	*Wnt1*	*wnt1* [8]	Boschl.CG.Botznik2013.chrUn.g54946‡	comp513900_c5_seq1 comp560097_c2_seq1
	Wnt2	*wnt2B* [9] *wnt2b* [8]	Boschl.CG.Botznik2013.chr5.g27946‡	comp556459_c5_seq5
	Wnt5a_1	*wnt5A* [9]	chr12 (16990326..16997306)‡	comp564602_c4_seq4
	Wnt5a_2	-	Boschl.CG.Botznik2013.chr13.g54683‡	-
	Wnt5a_3	-	Boschl.CG.Botznik2013.chrUn.g52618‡	comp559388_c0_seq2
	Wnt5a_4	-	Boschl.CG.Botznik2013.chrUn.g69917‡	comp555181_c0_seq1
	Wnt5b1	-	Boschl.CG.Botznik2013.chr12.g59382‡	comp555286_c1_seq4
	Wnt5b2	*Wnt5b* [8]	Boschl.CG.Botznik2013.chr13.g39640 + g39642‡	comp560720_c1_seq3 comp560720_c1_seq2
	Wnt6	*Wnt2* [8]	Boschl.CG.Botznik2013.chr7.g45543 + Boschl.CG.Botznik2013.chrUn.g47754‡	comp542257_c0_seq comp553141_c0_seq1
	Wnt7	*wnt7B* [9] *wnt7a* [8]	Boschl.CG.Botznik2013.ChrUn.g46011‡	comp561241_c1_seq1
	Wnt9	*Wnt9B* [9]	chr12 (10811283..10817191)‡	comp566883_c1_seq16
	Wnt10	-	Boschl.CG.Botznik2013.chr6.g45564 + Boschl.CG.Botznik2013.chr6.g45565‡	comp547559_c0_seq1
	Wnt11	*Wnt4a* + *Wnt4* [8]	Boschl.CG.Botznik2013.ChrUn.g26693 + Boschl.CG.Botznik2013.ChrUn.g30622‡	comp555881_c2_seq1
	Wnt16	*Wnt16* [8]	Boschl.CG.Botznik2013.Chr9.g65512 + g65511‡	comp537474_c1_seq1 comp490118_c1_seq1 comp465456_c0_seq1
	WntA	*Wnt7b* [8]	chrUn (216959500..216964568) + Boschl.CG.Botznik2013.chrUn.g34723‡	comp557402_c0_seq1
Vertebrata				
P. marinus	*Wnt1*		scaf_00005‡: 20,134,359..20,157,546	ENSPMAT00000000452
	Wnt2		scaf_00042‡: 3,108,251-3,142,066	-
	Wnt3_1		scaf_00004‡: 12,290,358_12,311,172	ENSPMAT00000007828
	Wnt3_2		scaf_00028‡: 10,149,055..10,152,095	-
	Wnt4		scaf_00039‡: 6,074,861..6,105,243	ENSPMAT00000005487
	Wnt5_1		scaf_00042‡: 9,894,951..9,900,968	-
	Wnt5_2		scaf_05238‡: 2,865..8,297	-

	Wnt5_3		scaf_00022‡: 6,477,998..6,492,244	ENSPMAT00000003927
	Wnt5_4		scaf_00010‡: 2,668,595..2,676,578	-
	Wnt5_5		scaf_00068‡: 1,990,410..1,999,665	ENSPMAT00000010914
	Wnt5_6		scaf_00027‡: 8,182,445..8,189,929	-
	Wnt6		scaf_00005‡: 20,064,324..20,090,054	-
	Wnt7_1		scaf_00042‡: 6,386,092..6,420,339	-
	Wnt7_2		scaf_00057‡: 2,365,159..2,391,659	-
	Wnt7_3		scaf_00022‡: 7,178,423..7,194,222	-
	Wnt7_4		scaf_00049‡: 4,331,652..4,338,594	ENSPMAT00000001922
	Wnt7_5		scaf_00677‡: 7,202..37,080	-
	Wnt7_6		scaf_00027‡: 10,183,907..10,212,509	ENSPMAT00000006170
	Wnt8		scaf_00044‡: 1,493,431..1,505,124	ENSPMAT00000010173
	Wnt9		scaf_00005‡: 20,190,936..20,230,438	-
	Wnt10a		scaf_00004‡: 12,215,164..12,250,533	ENSPMAT00000007810
	Wnt11		scaf_00077‡: 1,067,696..1,071,303	-
	Wnt16		scaf_00038‡: 5,333,846..5,345,122	ENSPMAT00000009648
C. milii	*Wnt1*		XP_007907807	
	Wnt2		XP_007892784	
	Wnt2b		XP_007897604	
	Wnt3		XP_007905574	
	Wnt3a		XP_007884748	
	Wnt4		XP_007896096	
	Wnt5a		XP_007888715	
	Wnt5b		XP_007896697	
	Wnt6_1		XP_007906706	
	Wnt6_2		XP_007910098	
	Wnt7a		XP_007901542	
	Wnt7b		XP_007892915	
	Wnt8a		XP_007899559	
	Wnt8b		XP_007898078	
	Wnt9a		XP_007883762	
	Wnt9b		XP_007905575	
	Wnt10a		XP_007909548	
	Wnt10b		XP_007910099	
	Wnt11_1		XP_007909586	
	Wnt11_2		XP_007892018	
	Wnt16		XP_007903164	

*NCBI accession numbers are provided when available. Otherwise, accession numbers are from http://amphiencode.github.io/ for *B. lanceolatum*; http://genome.bucm.edu.cn/lancelet/ for *B. belcheri*; http://www.aniseed.cnrs.fr/ for ascidians; http://octopus.obs-vlfr.fr/public/botryllus/blast_botryllus.php for *B. schlosseri*'s ESTs database; https://genomes.stowers.org/organism/Petromyzon/marinus and https://www.ensembl.org/Petromyzon_marinus/Info/Index for *P. marinus*

†Name from the gene annotation in the database.

‡Sequences manually modified or annotated from the database.

References for Table S1

1. Schubert M, Holland LZ, Holland ND: Characterization of two amphioxus Wnt genes (AmphiWnt4 and AmphiWnt7b) with early expression in the developing central nervous system. *Dev Dyn* 2000, 217:205-215.
2. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al: The amphioxus genome and the evolution of the chordate karyotype. *Nature* 2008, 453:1064-1071.
3. Hotta K, Takahashi H, Ueno N, Gojobori T: A genome-wide survey of the genes for planar polarity signaling or convergent extension-related genes in Ciona intestinalis and phylogenetic comparisons of evolutionary conserved signaling components. *Gene* 2003, 317:165-185.
4. Hino K, Satou Y, Yagi K, Satoh N: A genomewide survey of developmentally relevant genes in Ciona intestinalis. VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways. *Dev Genes Evol* 2003, 213:264-272.
5. Sasakura Y, Ogasawara M, Makabe KW: HrWnt-5: A maternally expressed ascidian Wnt gene with posterior localization in early embryos. *International Journal Of Developmental Biology May* 1998, 42:573-579.
6. Miya T, Nishida H: Isolation of cDNA clones for mRNAs transcribed zygotically during cleavage in the ascidian, Halocynthia roretzi. *Dev Genes Evol* 2002, 212:30-37.
7. Sasakura Y, Makabe KW: Ascidian Wnt-7 gene is expressed exclusively in the tail neural tube of tailbud embryos. *Dev Genes Evol* 2000, 210:641-643.
8. Rosner A, Alfassi G, Moiseeva E, Paz G, Rabinowitz C, Lapidot Z, Douek J, Haim A, Rinkevich B: The involvement of three signal transduction pathways in botryllid ascidian astogeny, as revealed by expression patterns of representative genes. *Int J Dev Biol* 2014, 58:677-692.
9. Di Maio A, Setar L, Tiozzo S, De Tomaso AW: Wnt affects symmetry and morphogenesis during post-embryonic development in colonial chordates. *Evodevo* 2015, 6:17.

Table S2 *Branchiostoma lanceolatum* and *Halocynthia roretzi* primer and probe sequences

	Forward Sequence (5')	ReverseSequence (3')	Probe
Bla_Wnt1	GGAGAGAGTGCATGCTGTGA	TACAAGCACGTGTGGATGGT	ACTAGTGATTTACAAGCACGTGTGGATGGTCTTCGTCCTGACGCACTCCTC GCACTTGACCTGACAGCACCACTGGAAGGTGCAGTTACACCGCTCTTTCGT CACTTCCTGTCTAGTCGCGTAGCCTCGTCCGCAACACAGCAGGTCGCAGCC ATCCAGCCCTCGTGACGTCACGTTGCACTCCCTCCCTCTTGTACCCTCGAA GCCGAGACGGGGGTTGTTGCGACAGAAGTTGGGCGACCTCTCGTGGTAGAC CAGATCGTTGTCCGTGGGGTACTTGTGGCGGGAGTTCTTGGGGACAAGCCC GGTGACTTTCGCCCTGGAGCCCCGGTTGTTGCCGATGTCTGGCACCGCAAC TTTAGACGCGCCGTCGAACTTCTCCTTCAAGCTGTCGCCGACGTCGCGGAA GTTGGGCAGGCGCATCCAGCAGGTCTTGAGCGTGCAGGACCCCGACATGCC GTGACACTTGCACTCTCTGCGTAGGTTCTCGGCTACAGCCACTCTGCCTGC CTCGTTGTTGTGCATGTTGACCAGGTAGCGGACGGAGTCTCTCGTCTTCTC TCCCGCATCCACGAACTGCTTGGCGAACTCCTTCCCGAACTCGACGTTGTC GGAACATCCTCCCCATTCCCAGT
Bla_Wnt2	CGCACCATCTGTGACAACAT	GTCGATGATTTCGGAGCATT	ACTAGTGATTGTCGATGATTTCGGAGCATTCTTTACACCGGACGAAACAA CACCAATGGAACTTGCACTCACACTTCCGTGACCTCTTCACCCTGGTGGTG TCATAGCCCCGACCACAGCACATGATGTCGCAGCCATCAGTTCCGAGGGAC GATCGGTTACATTGGCGACCCACCGTGCCGAGGGAACCTGTGTCATCATCG GCTAAGCAGTAGTCTGGGGAGGTATCGAAGTACACCAGGTCTGACTTGGT GGGTGTCTTGTGGTTCCTGTCGGACACGGTGAGTCTGCTGCCGCCCTGATT CATCGTCACCTCAACCGCCCCATCATACCGCCGCCTCAGGTAGTCCGTCAC TTTCCTGAAATGCGACATTGCCCTCCAGCAGGTGCGGAGAGTGCAGGACCC GCTGACGCCATGACACTTACACTCCATCTTCATGTTCTTAGCGACGCCCCT CCTCCCTGCCCTGTTGTTATGCATGTTCATGGCCGCTCGGGCGTCCCTCTG GCTTCGGTCCCTCGCATCCACGAACTTCCTCGTGAAGCCCTCTCCGAACTT GATGTTGTCGCTGCAGCCTCCCCACTCGAACTCACCCGTTGAGTCCTCGCT GAATCCCCGCTTCTTGGGGTCACATG

Bla_Wnt3	AGTACTCCCGAGCGCGTTATG	TCACGTCATTTGCACGTGTGC	AGTACTCCCGAGCGCGTTATGGATTCAGTACGCTGCCAGGTGTCGGTCTTG TTCCTGTTCGTCATTTGGTCTCAGACGGCGGACGTCGGCATGGGTTCTCCC GGGTGGTGGTATCTGGCGGTGGGACCACACAGTTCACTTCACTGGCTGCTGCC AATGCGGGAGGACAGGGGAGAAGATGGCCGCTGGTCTGCAGCAGTATACC AGGTTTGGTGCCGAGACAGATCCGGTACTGCCGGAAGTTTCACGAGATCA TGCCGTTCGTTGCCGACGGCACGAAGCTGGGCATCCGGGAGTGTCAGCATC AGTTCCGCGGGCGCCGCTGGAACTGCACCACCGTCCAGGGACAGGTCTCCA TCTTCGGACCAGTCTTAGACAGAGGAAACATCAGTACCACACCAAAAACAC CTCCACGTGCGGGGGACTCCATACTCTGGCCGGCTGTCAACAGAGCTTCAA GAGAAGCTGCCTTCGTACACGCCATCACCTCGGACTGTGGCTGCGACAACA GACACAAGGGACCTCCTGGCGAGGGATGGAAGTGGGGAGGATGCAGTGAA GATGTCTTCTTCGGCACCAAGTTCTCGCGAGACTTCGTGGACGCGAGAATA CGGGGAAGGCGAGACGGGCGATCAGCCATGGACAGACACAACAATGAGGC AGGCAGACAGTCTATCATGAAGAACCTGCAGCTGAAGTGTAAGTGCCACG GCCTGTCGGGCAGCTGTGAGATCAAGACCTGCTGGTGGGCCCAGCCTGACT TCCGCACGGTGGGAAACGTGCTGAAGGACAAGTACGACTCCGCCTCCGAGA TGGCCGTGGAGCGCCACCGCGAGCCCTCCGGCATGGTCGACTCCCTCTACC CGCGGTACAGCTTCTTCAAGGCGCCCGGTAAGGACGACCTCATCTACTTCG AGAACTCGCCCAACTTCTGCGAGCCGAACAACTCCACGGGCTCGCTGGGAA CCAGGGGGAGGGAGTGTAACATCACGAGCCACGGCATCGACGGGTGCCAGC TGATGTGCTGCGGCAGGGGCTGGAACACGCGGACCGAGATGAGAACTGAG AAGTGCCACTGCCAGTTCCACTGGTGCTGCTACGTCACGTGTCAGGAGTGC CAGAAGAAGCACCAGGTGCACACGTGCAAATGACGTGA

Bla_Wnt4	TCTGCTACAGCGAACCAGTG	GTGCACCTCCACTGTCCTTC	ACTAGTGATTGTGCACCTCCACTGTCCTTCTACACGTTTTGCACTTGACAT AGCAACACCAATGGAACTTGCAACTACATCTCTCCACTACTTCCCGGGTGT GGGTGTTGTACCCTCTCCCGCAGCACAGAAGTTCGCAGCCATCGATGGCCT TGGAAGTCTTGTTGCACACCCTCCCGACTGTCCCCATCGACCCCACCTTGG TGTCCCGCACGCAAAAGTCTGGGGAAGCATCCAGATACACCAGGTCGGAGC TCGAGTGCGGTTTGAAGTCAGAATTGAGCGGCACGAGTTCTCTCCTACTGC CGATCTTTTTCTGTTGCACCTCGGTGGCGCCGTCGAATTTCTCCTTGAGCC TCGCCCCGACTTCCCGGAACGGCGGCATGGCCCGCCAGCACGTCTTCAGCT CGCAGGATCCCGAGACGCCGTGACACTTACACTCCGTCTTCATGTGGTCCA CTAGGTTCCTCCTTCCGGCTTCATTGTTGTGCAGATCCATCAGGGCCCTGC TCGAGGTCGCCGCGCCCGCTGCCGAGATGGAGTGCACGAAGGCAGCTTCCC TCGTACCTTCTTCTAGGACCTTCCCAAACACAGGGCCTCGTCTGTTGATGA GTGTGCTGCAGTTCCAGCGGCGGTGACGAAACTGGAACTGGCACTCTTCTA TAGACATGCGGGCCCCCTCCTTTACACTGTCCATCACTT

Bla_Wnt5	TGATACCACGGTCTGAAGATG	AGTTGCGCCTCTGCTATTTGC	TGATACCACGGTCTGAAGATGGCGGTACAGATGAGTCTGCGGGTACTGCG GGTGCTGGTGACACTGCTGTCCTGTTACACACACCTGGGCAGAGTCAGGGC CACCTGGTGGCAAATGGCAGTGGATTCGAGATTGTACAGCCTGTCCCGGGC GGAACTGTACATAATCGGGGCTCAGCCGCTCTGCACCTCGTTGGCGGGGCT GTCGTCGGGACAGAGAAAACTGTGTAACTTGTATCAGGACCACATGGCGT CTGTGGGCATCGGGGCCAGACAGGGGATAGAGGAGTGTCAGCACCAGTTT AGGGACCGAAGATGGAACTGTACCACGTCGGACGAGGACTCCGTCTTCGGC AGGATAGTCAACATAGGCAGTAGAGAAGCCTCCTTCACCTACGCCATAGCA GCAGCAGGTGTAGTCAACGCCGTCAGTAGAGCATGCCGGGAAGGCGAGCT GACCACATGCGGCTGCAGCAGAGCGAAGAGACCGAAGGACCTGAACAGGG ACTGGCTGTGGGGAGGGTGTGGCGACGATGTGGAATATGGATATTACTTC GCGCGAGAGTTTGTTGACGCTCAAGAGAAGGAGATCATCCCGTCGCCTGGG TCCAGTGCGCACGCGCGGCAGCTGATGAACATGCACAACAATGAAGCTGGT AGAAAGCTCACGTTCAGTAACGCGAGAGTAGCCTGCAAGTGCCACGGAGT TTCCGGCTCCTGCAGCCTGAAGACGTGCTGGCAGCAGCTGGCGGACTTCAG GACGGTCGGCAACCTGCTGAAGGACAAGTACGACGGCGCCAACGAGGTGA AACTCATCCGGAGGGGCAAGAGGTACCGCCTGGATCGCCGCAACCCGAGAT TCAACGTCTTCACGGACGAGGACCTTGTGTACCTGAATAAGTCACCCGACT ACTGCAACGCGGACCCGACCATAGGTTCGCTGGGAACACACGGCAGAGAGT GTAACAAAACCGGGCTGGGCACGGACGGGTGTAACCTCATGTGCTGCGGG AGAGGATACAACACGTTCAAACGGGAAAAGGTGGAAAGGTGCAACTGCAA GTTCCACTGGTGTTGTTACGTCAAATGCAAAAGGTGTCGGTCAATAGAAG ACGTTTACGTATGCAAATAGCAGAGGCGGCAACT

Bla_Wnt6	CGGGGCTGCGACTACCAGAT	GAACGCCTCGCTCATAAACAC	CGGGGCTGCGACTACCAGATGAAGGGGGAGAGCCCGGATGGCAGCTGGGA GTGGGGGGGCTGCGGGGACGACATCGACTTCGGCTACACAAAGTCACGTGA GTTCATGGATGCCCAGACCAGACACAGGTCCGACATCAGAACGCTCCTGAC TCTACACAACAACGAGGCGGGAAGACTGGCTGTAAAGAACTTTATGCGGA CGGAGTGCAAGTGCCACGGACTGTCCGGATCATGCGCAGTAAAGACGTGTT GGAAAAAGATGCCGATATTCCGAGAGGTCGGGGTCCGGCTAAAGGAACGG TTCAACGGCGCGTTCCAAGTCATGGGATCCAACAACGGCAAATATCTCATC CCAGTCGGAGACACTATCAAAGCCCCTACGGCTGAGGACCTCGTCTATACC AACGAGTCGCCTAATTTTTGCAAAAGGAACAGAAAAACAGGGTCGCAAGG GACCAAAGGGCGGGCTTGTAACGCCACGTCCATGGGGATAGGCGGCTGTGA CTTGTTGTGTTGTGGGAGAGGGTACAAGGAGAGACAAGTTGTCGTTGAGG AGAACTGCAAGTGTCGTTTCCACTGGTGCTGTGTAGTCAAGTGCTCCAAAT GCACGGCCGTCAAAACTGTGCACGAGTGTTTATGAGCGAGGCGTTC
Bla_Wnt7	GAAGCTGCCTTTACTTACGC	GACGTCATGATCATTTGCAGG	GAAGCTGCCTTTACTTACGCCATCAGCTCGGCGGCGCTGGTGCACGCTATC GTCACAGCCTGCAGCCAGGGTAACATCTCTGACTGCGGCTGTGACCGTACG AAGGAGGGCGATCTGAACGACGAGGGCTGGAAGTGGGGCGGCTGCTCCGC CGACGTCAAGTACGGGCTTCGCTTCTGCAAGAAGTTCGTGGACGCCCGGGA GGTGGAGCAGAACGCGCGGGCTCTGATGAACCTCCACAACAATGAGGCTGG TAGGAAGGTGATTGACCAGCACACACGTCTGGAGTGTAAGTGTCACGGTG TCTCCGGCTCCTGCACCATGAAGACATGTTGGATTACTCTCCCACGCTTCC GAGAGGTCGGCAACATCCTGAAGGAGAAGTATCACCACGATTCCCAGCTCG TGGAAGCCGTTCGCGCCAGGCGCACGCGAAGGCCGACGTTTCTGAAACTGA AAAATTCTCGGACCTTCGAGAAACCTCGCGAGATTTCCCTTGTGTACCTGC GCGGGTCACCGAACTATTGTGAGCGTGACGAGGCAACAGGGTCGCTCGGA ACGAACGGGCGGAGGTGCAACCGGACGAGCCCATATCAGGACGGGTGTGA CCTGATGTGCTGCGGGAGGGGCTACAACACGCACCAGTTCGTCAAGACGTG GCAATGTAACTGTAAGTTCCACTGGTGCTGTTACGTCAAGTGCAACCAGTG CAGTGAGCGGACCGAGGAGTATACCTGCAAATGATCATGACGTC

Bla_Wnt8	CGATTCAAATGGGTGGACTT	GTGACTTCCGTGGTGATCCT	ACTAGTGATTCTGTTTCTTCTCCGCTGCTTTGACTTTTTCCGGACGTTTTT CCTTCGGTTCTTGCTTTCCCGCTGCACACAGATGTATTTAGTGACGGTCTT GGTGCACTGCGAACACTTGACCGAGCAACACCAGTGGAACTTGCAGTTACA GCTGCTAGTGACTTCAGTCGTGATCCTTTTCGGAACGTATCCGCAGTCCTT ACAAAGGCGCTTGCAACTCTTCTTCTCGTATTTGTCCATGTTTTTCCCTCC GCGGAGGCATTCACGTCCAAGAGTTCCCCGTGACCCTACGGTGAGGTTCTG CCTACAGTAGTCGGGGGAATCCTCAAGGAACACCATGTCCTTCTTAAGCCC GGTGTTCCTCTTGGAGCTGGCGCTGTTGTTCTCCGTCAGCTGACCGCGCAC GTAGTCCACCTTGTCCGCCTTCTTGTACTTCTTCTTGAGGAAGACGCCGAT GGCGCGGAAGTCGGCCAGCTGGAGCCAGCATGTCTTGGTGGTGCAGCTTCC CGACACGCCGTGACACTTACACACCCGCTTCATGGTCTGTCTCACCGCCTT CCTGCCCACATCGTTGTTGTGTAGGTTCATGGCTGCCCGGGCGTCCTGGCC GTTCTCAACCCCGTCCGAGTACCTCTTGGAGATGCTCTCTCCGAAGGAGAT GTCGTCACTACAGCCGCCCCAAGTCCAACCGCCTTCCGCTTTTTTCCCGTT GTTGGTCTGGTCGCACCCGCATTGCTCGAAAGCGCCCTTGCTGCGGTTCCT GGTCAGGACGTACATCACACCAGCGGCGCTTATGGCGTGCACGAA

Bla_Wnt9	GGAGTGGACGACATCAACG	TCTCGTGTTGCATTCTTTGC	GACAAGCTGGAGCTCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCC CCGGGCTGCAGGAATTCACTAGTGATTTCTCGTGTTGCATTCTTTGCTCTC GTGTTGCATTCTTTGCTCTCGTGTTGCATTCTTTGCACTCGACGTAACAAC ACCATTTGACCTGGCACTGGCAAGGTCTCTTGATGATCAGGCTCTGTGTGT TGTGCCCCCTCCCGCAGCAGATGCTGTCGCAGTTCTTTTCCTTGTCGCAGA TGCGCCCTCTTGTCCCGTATGAATATTTCCCCTTTTTGCAGTAGCTCGGAG ACTCGTCGACGAACACCATGTCGTTGTTCTTCGGGCCTTTAGGCGCCTCGT TCCCGACGGGCTCGTTCATGAGCCGCGCTCGCTCCCCCGCACCGTTGGTCA TGCTGACAACCTTCATGGCCTTCTCATACTTGTTTTTCAAGACGTCTCCGA TCTCGTGGAAGGGTGCGAGTTGTCTCCAGCACGTCTGGACGCTGCAAGACC CCGAGACGCCATGACACTTACACGTCGTCTTCACCCGGGACCGCACCACTC GGATTCCCAGGTTGGTGTTATGTCTGTCCACGGCGGCTCTCAGGTCCTGCT CCTCCGCCTTGTTCATACTGAGGAACTGGCGCGCGAATTTCCGAGAATATT TCAGGTTGTCCCCGCACCCTCCCCACAGCCACGTCTCGCGGTTGTCGATGT CGTTAAAACTCTCGTCGCACGTGCACCGCTCCAGCTTCCCGCTCGAGCACG CCCGGCCCAGGCGTGGGTCAGCCCGGAGGACGAGAT
Bla_Wnt10	AAGATTCCGGAAGACCCTGT	ATTCTGTGATTCGGCACTCC	ACTAGTGATTAAGATTCCGGAAGACCCTGTGCTGAACGCCAACACCGTGTG CCGCACTTTCCCCGGGCTCAGCAAGAAGCAGCTCCAACTGTGTTACGAGTA TCCCGACGTGACGGCGGCGGCCATACAGGGCGTGCAGATCGCCATTCACGA GTGTCAGTGGCAGTTCAAGAACCACCGATGGAACTGTTCCAACCTGGAGAC CAAGAACAAGAACCCGCACCTCACACAGATCGCTAGCAAAGGTATCCGAGA GACAGCCTTCACCTTTGCGTCGGTATCCGCGGGAGTGGCTCACGCTGTGGC CAACGCCTGTAGCCTGGGTAAGCTGCACACGTGCGGCTGTGACAATGACTA CCCCACCAAGCCTCAGATCCCTACCTACTCCCCCAGCTCATTCCCAGGGGCC ATCCCTGGCGGCCCGATCCCAGGCTACCCTCCCGGGTACACACAGGGACAG ACGGCGGTACTACACACGAACAGTGGCGGCGGCAAGTCGTACAACGATTGG TCGTGGGGAGGTTGTAGCCACAACATTGAGTACGGCATCAAGTTTTCCAA GGATTTCCTTGACTCTCGCGAGACGTCGGTGGACATCTTCTCG

Bla_Wnt11	TTGCCGATGTAGGGAAGAAC	ACCGTGAAGTCGTTGTAGCC	ACTAGTGATTCCCACAGCACATTGTTTGACAGCTGTCTGGACCAACTGACG TCTTGTTGCACAACCTTCCTGTAGTTCCGTACGACCCCTTCCTGTTGTTAA CTATACAGTAGTCTGGAGACTTGTCCACGAAGATCAGATCCCCGCTGCTGT GTCGGCGGGACCGTCTGTCCTCTGGGACTAGCTGTTGTCGTGTTCCGATCT TCCTCTTCACCATCTTGATAGCGTATGAATATTTCTCGGCGAGTTCGTGAG AGATCTCTGTCAGGTCGGCTAACGACTTCCAGCATGTCTTGACGTTACATG ATCCCGAGACGCCATGACACTTGCACTTCGTCGTCATCGTGTTTTGCACGG CCAGTCTTCCAGCTCCACTGTTATGTAGGTTCATGAGTGTCTGTGTGTGAG ACCGCTTCCTCTTCATCATGGGTGCGTCCGCGAACCCGGTACCGAACTCCA GCCCGAACTTGACGTTGTCCCCACAGCCGCCCCACGTGTAGTTCCCGTCAG GCTTCTCGCCGGGTTTCCTCGCGCATGAGCAGGCCTT
Bla_Wnt16	ATGGACGGGGCTAATAAAGC	ATGATGGTCTCGCAGGTGTT	ACTAGTGATTATGATGGTCTCGCAGGTGTTGCACTTGACGTAACAGCACCA GACGAACTTGCAGCCGCCACCTCTCCACGAACCGAACCACCTGCGTGTTGT AGCCCCGCCCGCAGCACAGCAGGTTACAACTGTCCGGACCCGTGGACGTTC TGTTACACTCTCTTCCCCTGGTGCCAAAGATCCCACGTTTCTTGTCCGACT TGCAGAAGTTTGGAGATTTCTCAAGAAACACCAACTCCTCGTCTCCGATAG GTACCTTCCGGTGCATCTTCTGTTTCCGCCGTAACCGCCGTCTCGTCTTCC GGAGGATGGGGACTGACGTCATGTACTTCTTTTTCAACATATCACCAACCC GTTTGAAGCTGGGCAATGTTTTCCAACAGGTCTTTACTGCGCATGATCCGG ACACGCCGTGACACCTGCACTTG

Bla_WntA	AACGCCATCAGGTGGTCTAC	TACAAGTGACCTCGCAGCAC	ACTAGTGATTTACAAGTGACCTCGCAGCACCATTTGAATTTGCAGTTGCA ACTCTCTTCTGTGGTCTCCTCTCTCGACACGAAGTCTCTGTTACAACAGAG TAGCTGACAGCCGTCTACCCCCGCACTTGTTCTATTACATTTCCGCCCGCG CGTCCCGTACGACCCCTTGGCCACGTCGACTTCGCAGTAGTCGGGAGAATC GTCCAGATACACCACGTCGTGTGCTGTGGGGCGTTTGTGTCTTTTGTTTCT CGGACGTAATCTGGATCCCTTTTTGTTGATCTTGACGTACGTGGCTCCGTG GTACTTTCCCATGATAATGTCTCCGACCTGTCGAAAGCTCGGCATGGACTC CCAACAGACCATGGTGGCGCAAGACCCGGACACGCCGTGGCACTTGCAGGT AACCTTCATATTGGCCTTGATTGACTTTCTGCCTGCTTCGTTGTTGTGTAA GTTCATTAGTCCAGTATCGTCTGCGGCCACTTCATTAGCGTCCATAAACTC TTTGGACAGGTTGTCTCCGAACCTGATGTTATGGGAGCACCCTCCCCAATC CCAC

Hro_WntA	GTACCGGCCCCGAACTCC	TCACTTGCAATATGATTCTTCAAC	GTACCGGCCCCGAACTCCGACCAACAAACACGAAATTCAAAAGGAGCGCGA GCCTACTGCAATTCACTAAAATACCTCACAGATAAGCAGAAGAGTTTGTG CATCCACAGTCCCACGATAATGAGTAAAATTAGCCACGGTGTTATCGCTGG AATGAGAGAATGTCAATTGCAATTTCAACACAGGAGATGGAACTGTACAC CGTATCAATCAAATAATCCGAAGACCCATGCTTTCGAACCTGTGCTGAAAA ACAAGGGCCCTGAAAGAGCGTACATCAAAGCCATGCTGAGCGCAGCCGTGT CCTATAAGATAACGCGAGCGTGTTCTGGCGGTGAACTTCCTCTTGAATGCA TGTGTGCAAATATTAAACGTCCAGCGAAGGATGATCCGGAGGACCCTAAG AAAAAATTTGAATGGGGAGGATGCTCGGATGATGTTAATTTTGGAGATTC GAGATCCCGTGAATTCTTTGACGAATCGGATTTATCTCGAAATCAAGGAA GAGTACTTGACATACACAACAATGAAGCGGGAAGAAAGTCCGTAAAAATG AGCAATCGTACAACATGCAAATGCCACGGCCTTACCGGAGCCTGCACACAG AACATTTGCTGGAGATCGCTACCTCCATTGCGCAGAATAGGTAAAGAATT ATTTCAACGATATAATAGCGCTGTACGAGTAAAAACTCGAAAACTGGATA TCCAAGGAAAGGCTACAACCAGCGATTTGGTATACGTTCGAAAATCACCT TCGTTTTGCATAAGAAACAATCGTGGTTCTTACGGGACCGCCAAACGGTG GTGTAACAGGACAAGCACGGAAACTGAGGGATGCGCTCATATGTGCTGCG GTCGGGGTTACAGAACCGAAGAAAAGATAAAGATAACTGAATGCAATTGC GAATTTCAATGGTGTTGTAATCTCAATTGTTATTTCTGCGAAGAACATGT TGACGAATCATATTGCAAGTGA

Table S3. *Wnt* synteny in lancelets (*B. lanceolatum, B. belcheri* and *B. floridae*) and vertebrates (*H. sapiens* and *P. marinus*)

LANCELETS

	Branchiostoma lanceolatum#		*Branchiostoma belcheri*		*Branchiostoma floridae**	
Gene	ID	Scaffold: coordinates (strand)	ID	Scaffold: coordinates (strand)	ID	Scaffold: coordinates (strand)
Wnt1[¶]	BL08374	Sc0000014: 2,655,007-2,664,370 (-)	306640F	Sc0001290: 14,548-21,893 (+)	113720	sc0000012: 3,339,473-3,348,336 (+)
Wnt2	BL18396	Sc0000085: 1,065,381-1,083,862 (+)	073170F	Sc0000031: 837,724-858,936 (+)	145693	sc0000061: 1,676,862-1,683,402 (+)
Wnt3	BL00565	Sc0000034: 1,039,202-1,073,486 (-)	177200F	Sc0000151: 348,528-368,321 (+)	118309	sc0000012: 804,098-834,674 (+)
Wnt4	BL15100	Sc0000105: 395,585-434,158 (-)	271880F	Sc0000613: 51,484-51,936 (+)	56711	sc0000323: 231,613-265,780 (+)
Wnt5	BL05330	Sc0000129: 22,035-62,251 (+)	039340R	Sc0000011: 585,989-598,190 (-)	118160	sc0000010: 3,427,102-3,453,255 (-)
Wnt6	BL13401	Sc0000014: 2,678,147-2,708,947 (-)	258720R	Sc0000490: 126,712-138,581 (-)	57222	sc0000012: 3,308,009-3,322,220 (+)
Wnt7	BL12283	Sc0000266: 95,574-120,318 (-)	039420R	Sc0000011: 721,380-731,165 (-)	56634	sc0000010: 3,594,285-3,614,777 (-)
Wnt8	BL09252	Sc0000095: 692,325-710,019 (+)	197740F	Sc0000194: 438,989-439,780 (+)	56726	sc0000413 :614,838-634,936 (-)
Wnt9	BL22609	Sc0000014: 2,622,645-2,644,219 (+)	185850F	Sc0000168: 197,245-210,888 (+)	67024	sc0000012: 3,362,909-3,377,125 (-)
Wnt10	BL23379	Sc0000034: 1,116,089-1,170,047 (-)	177150F	Sc0000151: 250,803-287,391 (+)	113709	sc0000012: 753,835-754,254 (+)
Wnt11	BL18893	Sc00000711: 287,178-311,798 (-)	034950F	Sc0000009: 654,648-663,318 (+)	271417	sc0000460: 536,632-547,782 (+)
Wnt16	BL18405	Sc0000085: 387,208-427,572 (+)	072770F	Sc0000031: 137,433-139,798 (+)	205854	sc0000027: 1,292,590-1,325,871 (+)
WntA	BL17431	Sc0000101: 315,161-333,725 (+)	308600R	Sc0001374: 7,955-16,915 (-)	60204	sc0000201: 663,106-672,419 (-)

VERTEBRATES

	Homo sapiens
Gene	Chromosome: coordinates (strand)
Wnt1	12: 48,978,453-48,981,676 (+)
Wnt2	7: 117,276,631-117,323,289 (-)
Wnt2b/Wnt13	1: 112,466,541-112,530,165 (+)
Wnt3a	1: 228,007,051-228,061,260 (+)
Wnt3	17: 46,762,506-46,833,154 (-)
Wnt4	1: 22,117,305-22,143,969 (-)
Wnt4b	
Wnt5a	3: 55,465,715-55,490,539 (-)
Wnt5b	12: 1,529,891-1,647,243 (+)
Wnt6	2: 218,859,821-218,874,233 (+)
Wnt7a	3: 13,816,258-13,880,121 (-)
Wnt7b	22: 45,920,362-45,977,129 (-)
Wnt8a	5: 138,083,892-138,092,365 (+)
Wnt8b	10: 100,463,041-100,483,744 (+)
Wnt9a/Wnt14	1: 227,918,656-227,947,898 (-)
Wnt9b/Wnt15	17: 46,833,201-46,886,730 (+)
Wnt10a	2: 218,880,363-218,899,581(+)
Wnt10b/Wnt12	12: 48,965,340-48,971,763 (-)
Wnt11	11: 76,186,325-76,210,736 (-)
Wnt11r	
Wnt16	7:121,329,003-121,341,104 (+)

	Petromyzon marinus
Gene	Scaffold: coordinates (strand)
Wnt1	scaf_00005: 20,134,359-20,157,546 (+)
Wnt2	scaf_00042: 3,108,251-3,142,066 (+)
Wnt3_1	scaf_00004: 12,290,358-12,311,172 (+)
Wnt3_2	scaf_00028: 10,149,055-10,152,095 (-)
Wnt4	scaf_00039: 6,074,861-6,105,243 (-)
Wnt5_1	scaf_00042: 9,894,951-9,900,968 (+)
Wnt5_2	scaf_05238: 2,865-8,297 (-)
Wnt5_3	scaf_00022: 6,477,998-6,492,244 (-)
Wnt5_4	scaf_00010: 2,668,595-2,676,578 (+)
Wnt5_5	scaf_00068: 1,990,410-1,999,665 (+)
Wnt5_6	scaf_00027: 8,182,445-8,189,929 (-)
Wnt6	scaf_00005: 20,064,324-20,090,054 (-)
Wnt7_1	scaf_00042: 6,386,092-6,420,339 (-)
Wnt7_2	scaf_00057: 2,365,159-2,391,659 (+)
Wnt7_3	scaf_00022:7,178,423-7,194,222 (+)
Wnt7_4	scaf_00049: 4,331,652-4,338,594 (+)
Wnt7_5	scaf_00677: 7,202-37,080 (+)
Wnt7_6	scaf_00027: 10,183,907-10,212,509 (+)
Wnt8	scaf_00044: 1,493,431-1,505,124 (-)
Wnt9	scaf_00005: 20,190,936-20,230,438 (+)
Wnt10	scaf_00004: 12,215,164-12,250,533 (+)
Wnt11	scaf_00077: 1,067,696-1,071,303 (+)
Wnt16	scaf_00038: 5,333,846-5,345,122 (-)

scaffolds 129 and 266 are in fact linked (new scaffold 63), demonstrated through additional PacBio sequencing (Robinson-Rechavi, personal communication)

* alternate scaffolds exist in many cases due to presence of two alleles during genome sequencing, and may explain opposite orientations

[¶] Shared colour denotes *Wnt* genes on the same chromosome or scaffold

Additional file 2: Wnt sequence alignement for Figure 1, can be found at Genome Biology web page (https://genomebiology.biomedcentral.com/).

Additional file 3: Wnt sequence alignment for Figure S4., can be found at Genome Biology web page (https://genomebiology.biomedcentral.com/).

Resultats. Capítol 4

Oikopleura dioica Wnt signaling: when gene loss, duplication and function shuffling push the limits of chordate EvoDevo

Josep Martí-Solans[1], Hector Godoy[1], Miriam Diaz-Gracia[1], Takeshi Onuma[2], Hiroki Nishida[2], Ricard Albalat[1] i Cristian Cañestro[1]

[1]Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

[2]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan.

Resum

La via de senyalització de Wnt juga un paper clau en processos de desenvolupament en tots els metazous, des de l'especificació del destí cel·lular i la formació d'eixos embrionaris fins a la proliferació cel·lular. Durant l'evolució dels metazous, les famílies Wnt han sofert nombrosos guanys i pèrdues genètiques de forma específica en cada llinatge. En els cordats, per exemple, els urocordats ascidis destaquen entre la resta pel seu notable dinamisme i gran tendència a perdre i duplicar els gens Wnt. En aquest estudi, ens centrem en una altra classe d'urocordats, l'apendiculària *Oikopleura dioica*. Hem identificat el catàleg Wnt complet d'*O. dioica* i hem realitzat anàlisis filogenètiques, d'expressió i funcionals per tal d'explorar fins a quin punt els esdeveniments de pèrdua de gens, duplicació i reassignació de funcions han afectat l'evolució dels lligands Wnt en aquesta espècie. Els nostres resultats posen de relleu una evolució molt dinàmica dels lligands Wnt a *O. dioica*, on hi hagut una reducció extraordinària en el seu repertori de les subfamílies Wnt -esdevenint un dels més petits descrits fins ara en bilaterals- que ha estat acompanyada per una expansió específica d'O. dioica en altres subfamilies. Finalment, els anàlisi d'expressió i funcionals ens han permès debatre com les reassignacions de funcions (function shuffling) han convertit en prescindibles algunes subfamílies permetent-ne la seva pèrdua a *O. dioica*.

Oikopleura dioica Wnt signaling: when gene loss, duplication and function shuffling push the limits of chordate EvoDevo

Josep Martí-Solans[1], Hector Godoy[1], Miriam Diaz-Gracia[1], Takeshi Onuma[2], Hiroki Nishida[2], Ricard Albalat[1]✉ and Cristian Cañestro[1]✉

[1]Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

[2]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan.

Abstract

The Wnt signaling pathway plays key roles in developmental processes in all metazoans, ranging from cell fate specification and embryonic axis formation to cell proliferation. During metazoan evolution, Wnt families have undergone numerous gene gains and losses in a clade-specific manner. In chordates, for instance, urochordate ascidians stand out among the rest for its notable dynamism and high tendency to lose and to duplicate *Wnt* genes. In this study, we focus on another urochordate class, the appendicularian *Oikopleura dioic*a. We have identified the complete Wnt catalog of *O. dioica* and we have performed phylogenetic, expression and functional analysis in order to explore how far events of gene loss, duplication and function shuffling have affected the evolution of the Wnt ligands in this species. Our results highlight a very dynamic evolution of the Wnt ligands in *O. dioica*, in which an extraordinary reduction of the catalog of the Wnt subfamilies –being one of the smallest described so far in bilaterians– has been accompanied by a lineage-specific expansion of others. Finally, expression analysis and functional knockdown approaches by DNAi allow us to discuss how Wnt losses might be associated with gene dispensability due to either function shuffling among Wnt subfamilies or to a high degree of malleability of *O. dioica*'s developmental programs.

1. Introduction

The *Wnt* family encodes a set of secreted glycoprotein ligands that regulate key events of animal development, including embryonic axis specification, gastrulation, and the formation of limbs, heart, and neural system. (1–5). *Wnt* ligands, after binding to a Frizzled receptor, can activate two different pathways: *i*) the cell-fate pathway (a.k.a. ‘‘*Wnt*/β-catenin pathway’’ or the ‘‘canonical *Wnt* pathway’’), which is mediated by the nuclear localization of β-catenin for pathway activation; and *ii*) the cell-polarity pathway, which is mediated by several intermediate effectors acting independently of β-catenin, and includes, at least, the non-canonical planar cell polarity (PCP) pathway and the non-canonical *Wnt*/Calcium pathway (6).

The multiple metazoan *Wnt* genes are classified in 13 subfamilies –*Wnt1* to *Wnt11*, *Wnt16* and *WntA*– that emerged early in the animal evolution (7). Comparative analyses of Wnt repertoires throughout all metazoan taxa have shown

different trends to duplicate, retain or lose *Wnt* genes in different lineages (8–13) that have been associated with clade-specific features (10, 14). In poriferan species such as glass sponges, for instance, the loss of *Wnt* ligands has been linked to the atypical, largely syncytial organization of their tissues (8), whereas the loss of *Wnt6* in hemipterans has been correlated with the absence of maxillary palps in this clade (15).

Recent analysis of the *Wnt* family in chordates has revealed that *Wnt* gains and losses have distinctively impacted the three subphyla, from which cephalochordates and urochordates display opposite evolutionary patterns, and vertebrates exhibit an intermediate situation (13). Thus, cephalochordates show a conservative pattern of evolution, retaining the complete repertoire of chordate Wnt as single-member subfamilies. In contrast, ascidians reveal a dynamic pattern, with numerous gene losses and duplications that have led to the disappearance of entire subfamilies in some lineages and to the expansion of some Wnt in other species (13). Finally, vertebrates expanded Wnt subfamilies through the two rounds of whole genome duplication and lost afterward many ohnologs, although the only subfamily totally lost in this lineage is WntA, an absence that have been associated to the evolution of an alternative mode to open a mouth in vertebrates (13).

It has been argued that the complex evolutionary history of gene gains and losses in different ascidian lineages might have contributed to the morphological diversification of urochordates (13). This idea led us to wonder how far gene losses (and duplications) can push in the building of a chordate, in terms of 'which is the minimal number of *Wnt* subfamilies to build a chordate'? or 'how conserved are the *Wnt* functions in chordates'? To address these questions and to further investigate the evolvability of the *Wnt* signaling, we proposed to investigate the *Wnt* repertoire in the *Oikopleura dioica*, which belongs to another class of urochordates. *O. dioica* is a pelagic urochordate that has become an attractive animal model for studying gene loss in the field of evolutionary developmental biology thanks to several characteristics: *i*) its genome has been sequenced revealing the loss of many groups of genes present in ancestral chordates (16); *ii*) it occupies an strategic phylogenetic position within the chordate phylum, between the genetically-redundant vertebrates and the genomically-static cephalochordates; *iii*) it retains a typical chordate plan – including a notochord and a dorsal nerve tube– throughout their entirely life in

contrast to ascidians, which suffer a drastic metamorphosis loosing most of their chordate features (17, 18); and *iv*) it can be cultivated in the laboratory (19–21) and techniques for functional studies by knocking-down genes have been recently developed for this organism (22–25).

We have conducted an exhaustive survey of *Wnt* genes in genomic databases of *O. dioica* and have generated the first complete atlas of developmental expression of the Wnt family in this species. Our study reveals a very dynamic evolution of Wnt signaling in *O. dioica* that would have led to an extraordinary reduction of the number of subfamilies –with only 4 out of the 13 subfamilies, it has the smallest Wnt catalog described so far in chordates– accompanied by the expansion of the Wnt11 subfamily by lineage-specific gene duplications. Our detailed atlas of *Wnt* expression in *O. dioica* reveals that expression domains encompassed tissues derived from all three germ layers in a highly dynamic manner as well as several cases of 'function shuffling'. Finally, our study also suggests that an asymmetrically localized maternal Wnt ligand is required for axis formation. Therefore, our results allow us to evaluate the contributions of different Wnt subfamilies during *O. dioica* development and to investigate the evolutionary and functional limits of *Wnt* signaling in chordate development.

2. Results

2.1. Only four *Wnt* subfamilies have survived in *O. dioica*

We conducted an exhaustive survey of *Wnt* genes in genomic database of *O. dioica.* We identified 7 putative *Wnt* genes, which were initially classified by "best BLAST hit" approach into the different Wnt subfamilies. To corroborate this classification, we performed phylogenetic reconstructions using a total of 254 *Wnt* sequences from 20 species representing all major metazoan groups, from the cnidarian *Nematostella vectensis* to the vertebrate *Homo sapiens.* The analysis recovered the 13 *Wnt* subfamilies as monophyletic groups and distributed *O. dioica Wnt*s among these 13 subfamilies with high support values (Fig. 1). Maximum likelihood (ML) phylogenetic tree suggested that the 7 *O. dioica Wnts* belonged to 4 Wnt subfamilies -i. e., Wnt5, Wnt10, Wnt11 (4 genes) and Wnt16 subfamilies- (Fig. 1). Therefore, *O. dioica* seemed to have lost 9 Wnt subfamilies during its evolution. On the other hand, our results revealed that *O. dioica* has expanded the Wnt11 subfamily to 4 paralogs, named *Odi_Wnt11a* to *Odi_Wnt11d*. Analysis of the gene structure of the *Odi_Wnt11* paralogs showed that *Odi_Wnt11a* had 5 introns, one of

them in a conserved position in all *Wnt* genes across all metazoans (boxed black arrowhead in Fig. S1), and other intron located in a position considered a signature of Wnt11 subfamily in bilaterians (26) (green arrowhead in Fig. S1), further supporting its orthology. The other 3 *Odi_Wnt11* paralogs, namely *Odi_Wnt11b, Odi_Wnt11c* and *Odi_Wnt11d*, had no introns, pointing to a retrotranscriptional origin during the evolution of the appendicularian lineage.

The Wnt signaling pathway, far from simple, depends on the action of multiple genes (i. e. Wnt receptors, secreted Wnt inhibitors, intermediate effectors, etc.), which we wondered if they could have been also affected by the massive loss of *Wnt* ligands in *O. dioica*. To investigate this possibility and to assess the conservation of the Wnt signaling pathways in *O. dioica*, we used the KEGG Automatic Annotation Server (KAAS) (27) on *O. dioica* genomic and transcriptomic data from (28) to automatically identify the *O. dioica* orthologs to the components of the three main Wnt signaling pathways. KAAS analysis revealed conservation of the key components of the three Wnt pathways, with the exception of Axin, APC and several antagonists (Fig. S2A and Table S1). Phylogenetic analysis in *O. dioica* of Wnt receptor *Frizzled* (*Fzd*) (Fig. S2B) and Wnt antagonist *secreted Frizzled related protein* (*sFRP*) (Fig. S2C) revealed, in addition, a decrease in the diversity of these components in this species. *O. dioica* appeared to have lost 2 out of the 5 Fzd subfamilies present in ascidians and vertebrates –retaining members in the Fdz1/2/7 (1 gene), Fdz3/6 (3 genes) and Fdz5/8 (1 gene) subfamilies– and had only 1 sFRP (*Odi_sFRP2*), in contrast to the 5 and 3 sFRP subfamilies found in vertebrates and ascidians, respectively.

2.2 *Wnt*s are expressed in all germ-layers during *O. dioica* development

To explore the functional consequences of the *Wnt* losses and duplications in *O. dioica* developmental programs, we investigated the expression patterns of all *O. dioica Wnt* genes during embryogenesis and larval development by carrying out whole-mount in situ hybridizations (WISH). Results revealed that 5 out the 7 *Wnt* genes showed complex tissue-specific expression patterns that changed throughout different developmental stages (Fig. 2), whereas no signal was detected for two *Wnt11* paralogs, namely *Odi_Wnt11b* and *Odi_Wnt11c* (Fig S3), suggesting that

146 they may function later, maybe at adult stages, or they were integrated after retrotranscription in a genomic region in which no transcription is promoted.

148 Because many *Wnt* genes are maternally expressed and play a role in establishing the primary axis in several metazoan species, we paid special attention 150 to the *Wnt* expression in *O. dioica* eggs. WISH results revealed that from the 7 *Wnt* genes only *Odi_Wnt11a* was part of the maternal component (Fig 2S). This *Wnt* 152 paralog was, therefore, the best candidate to participate in establishing primary embryonic axis in *O. dioica* (see 2.3 section for details).

154 All the other *Wnt* appeared to only have zygotic expression. At approximately 64-cell stage, *Odi_Wnt11d* signal was detected in two internal domains, which could 156 be endodermal derivates precursors, such as endostyle and endodermal strand (Fig. 2AD yellow dots), while *Odi_Wnt11a* began its zygotic expression in several 158 blastomeres in the vegetal part of the embryo (Fig. 2U). The onset of *Odi_Wnt5* and *Odi_Wnt10* and *Odi_Wnt16* expressions occurred later, at tailbud stages, when new 160 expression domains appeared in a highly dynamic fashion encompassing tissues derived from all three germ layers.

162 In the mesoderm, *Wnt* expression signal was observed in a limited number of cells of the musculature and notochord. Thus, *Odi_Wnt10* appeared to be expressed 164 in the 5th muscle cell pair of late tailbuds (m5 in Fig. 2N), while *Odi_Wnt16* signal appeared in the 6th, 7th and 8th muscle cell pairs (m6-8 in Fig. 2AO). At this stage, 166 *Odi_Wnt11a* signal appeared to be restricted to the posterior third of the notochord (Fig. 2V-X magenta arrowheads) and to the meso/endoderm in the posterior part of 168 the trunk, anteriorly to the first cell of the notochord (Fig. 2W orange arrowheads). This meso/endodermal expression was maintained until mid-hatchling stages, when 170 traces of *Odi_Wnt11a* signal could be still detected in the trunk (Fig. 2 X-Z orange arrowheads).

172 In the endoderm, we also detected *Wnt* signal domains. From early tailbud to just-hatchling stages, *Wnt11d* labeled a group of posterior cells of the tail at the right 174 side of notochord that correspond to the region populated by the endodermal strand (Fig. 2AE-AF yellow arrowheads and Fig. S4 C and D). Later, *Odi_Wnt11d* signal 176 was detected in a region of the posterior part of the trunk where presumably the endodermal strand cells migrate (Fig. 2AG yellow arrowheads and Fig. S4E). After 178 hatch, the endostyle primordium showed a temporal expression of *Odi_Wnt11d* and *Odi_Wnt5* that was maintained up to late hatchling stages restricted to the most

ventral and posterior part of the organ (Fig. S4 E and Fig 2F-I yellow asterisk). At mid-hatchling stage, part of the stomach primordium was labeled by *Odi_Wnt10*, an expression domain that was maintained up to late-hatchling stages mostly restricted to the vertical intestine, to the most external part of the rectum in the opening of the anus, and to the connection between the esophagus and left stomach (Fig. 2Q-R yellow arrowheads and Fig S4A). At mid-hatchling stages, *Odi_Wnt11a* showed a faint expression in the connection between the two stomach lobes (Fig. 2Z yellow arrowhead).

We also observed *Wnt* expression signal in derivatives of the ectoderm, including the nervous system and epidermis. At tailbud stage, *Odi_Wnt11a* labeled an anterior region of the trunk, in the presumptive area where the pharynx and the anterior brain will develop (Fig. 2V light blue arrowheads). This expression persisted until the mid-hatchling stage, in which signal appeared labeling two bilateral domains, one adjacent to the sensory vesicle and the other at the outer part of the brain near to the epidermis, plausibly corresponding to the dorsal nerve secretomotor neurons (29) (Fig. 2W-Z light blue arrowheads). Also, at tailbud stages, *Odi_Wnt11d* signal was observed in the neuroectoderm dorsally located to the brain, which faded in posterior stages (Fig. 2AF light blue arrowheads), while *Odi_Wnt5* signal was observed in the posterior part of the tail, likely corresponding to the developing neural tube (Fig. 2D-E). By hatchling stages, *Odi_Wnt5* signal exhibited a scattered pattern throughout the neural tube including the caudal ganglion (Fig. 2F-H light blue asterisks), which became evenly distributed in late hatchlings reflecting the distribution of neurons throughout the nerve tube. *Odi_Wnt5* signal was also observed in the anterior brain of late hatchlings (Fig. 2I light blue asterisks and arrowheads, respectively). Besides, *Odi_Wnt10* signal was also detected in nervous system, specifically in the caudal ganglion by late tailbud stage (Fig. S4A light blue asterisk). From early hatchling to mid hatchling stages, *Odi_Wnt10* also labeled a single cell near the tip of the tail (Fig. 2P and Q green asterisk). This cell could be the same stained by the neuronal marker α-tubulin (30), although the neuronal nature of this cell needs further investigation.

Regarding the epidermis, *Wnt* signal was observed in four different types of epidermal domains. i) Domains in the trunk epidermis related to sensory or placode-related cells connected to the nervous system. These domains included the ventral organ and the lateral of the mouth for *Odi_Wnt5* signal (Fig. 2F-H dark blue

arrowheads), and the paired Langerhans receptor primordia in the posterior part of the trunk for *Odi_Wnt11a* signal (Fig. 2W-Y dark blue arrowheads). ii) Domains in the oikoplastic epithelium. The mid-dorsal domain of the oikoplast was labeled by *Odi_Wnt11a* (Fig. 2X-Z), which also labeled the posterior-dorsal domain together with *Odi_Wnt5, Odi_Wnt11d* and *Odi_Wnt16* (Fig. 2F-I, AF-AH, AS dark blue asterisks) and the ventral oikoplast, along with *Odi_Wnt5* and *Odi_Wnt16* (Fig2. G-I, X-Y and AS and Fig. S4B dark blue asterisks). iii) A domain in the lateral of the tail. A cell band in the equator of the tail together with the most distal epidermal cell of the tail was labeled by *Odi_Wnt5* and *Odi_Wnt11a* (Fig. 2F-H and W-Y, dark blue double arrowheads). And iv) a domain in the posterior-ventral epidermis of the trunk, close to the separation between the trunk and the tail, was labeled by *Odi_Wnt11d* (Fig. 2AF dark blue arrow).

2.3. Maternal *Odi_Wnt11a* transcripts are asymmetrically localized in the posterior-vegetal region of cleaving embryos.

As mentioned above, *Odi_Wnt11a* appeared as the only *Wnt* gene maternally transcribed in *O. dioica* (Fig. 2S). Detail analyses during the first five cleavages of *O. dioica* embryos revealed that *Odi_Wnt11a* transcripts became asymmetrically localized in the two most posterior-vegetal cells (Fig. 3). In eggs, both before and after fertilization, *Odi_Wnt11a* signal was uniformly distributed throughout the cytoplasm (Fig. 3A and B). After the first division, however, *Odi_Wnt11a* signal appeared mostly accumulated nearby the contact region of the cytoplasmic membranes of the two daughter cells, in the prospective vegetal-posterior pole of the embryo (Fig. 3C). After the second division, the signal was only detected in the two posterior cells, only visible in the area towards the presumptive vegetal pole (Fig. 3D-F). After the third division, at 8 cell-stage, *Odi_Wnt11a* signal continued restricted to the posterior region of a pair of posterior vegetal blastomeres, named B/B according to Delsman's nomenclature (31) (Fig. 3G-L), equivalent to B4.1/B4.1 blastomeres according to the ascidian nomenclature system (to facilitate comparisons, ascidian nomenclature is indicated in parentheses) (32, 33). After the fourth and fifth divisions, *Odi_Wnt11a* signal remained in the blastomere pairs B1/B1 (B5.2/B5.2) and B11/B11 (B6.4/B6.4), which could be recognized by being the smallest ones for each division (Fig. 3M-X). According to the fate map of *O. dioica*,

these blastomeres are internalized during gastrulation and remain cleavage-arrested until the hatchling stage (33, 34).

2.4. *Odi_Wnt11a* is necessary for axis formation

Among the seven *Wnt* genes of *O. dioica,* we focused on *Odi_Wnt11a* because the asymmetrical distribution of its maternal component suggested a possible role in the establishment of the embryonic primary axis. To investigate the function *Odi_Wnt11a* during development, we generated knockdown animals using a DNAi approach (24, 25), which consisted in the microinjection of a double strand DNA (dsDNA) against the target gene. We PCR amplified two dsDNA fragments of 206 bp and 363 bp that extended over the first exon and the 3'UTR region of the *Odi_Wnt11a*, respectively (Fig. 4A and Fig. S5). These dsDNA were co-injected with an mRNA encoding for *Lifeact*-mCherry fusion protein into the ovary of pre-spawning females (22) (see Materials and Methods in Supporting Information for a detail description of dsDNA technique). As expected from this experimental approach, a gradient of dsDNA and mCherry was generated in the syncytial gonad from the point of injection. In the clutch from an injected female we found, therefore, both fluorescent mCherry embryos (Fig 4B blue and yellow arrowheads) and non-fluorescent embryos (Fig 4B green and white arrowheads). According to previous works (24), we assumed that fluorescent embryos had incorporated the dsDNA, and therefore could show an altered phenotype, whereas non-fluorescent embryos should develop normally. The analysis of the phenotypes of animals from clutches of injected females with both dsDNA against *Odi_Wnt11a* showed that more than two thirds of mCherry-positive larvae showed similar abnormal morphologies (trunks and tails shorter than non-fluorescent siblings; Fig. 4B and C), whereas control animals injected with a mock dsDNA (against Kaede protein) did not show such malformations. These results supported the specificity and reproducibility of the phenotype caused by both *Odi_Wnt11a* dsDNA. *In situ* hybridization of *Odi_Wnt11a* revealed a strong reduction *Wnt* signal in the posterior vegetal blastomeres (B/B) of 8-cell fluorescent embryos (Fig. 4D). This reduction was, however, less strong in mid tailbud stage embryos (Fig. 4D), which suggested that the dsDNA injected in the ovary was not able to completely abolish zygotic transcription.

To understand the functional consequences in the *O. dioica* development of knockdown *Odi_Wnt11a* by dsDNA injection, we performed WISH and

immunohistochemical analyses as well as enzymatic activity assays with several tissue-specific developmental markers, including those for neural tissue, notochord, muscle, endoderm or germ line. α-Tubulin A has been established as a general neuronal marker with expression in brain nerves, cerebral and caudal ganglia and nerve cord in *O. dioica* larvae (30, 35). WISH analysis of α-Tubulin A showed similar expression level in the neurons of the trunk and tail when malformed (knockdown) and normal (control) larvae were compared (Fig. 5A-B). We next analyzed the expression of *brachyury*, as a notochord marker, and the activity of acetylcholinesterase (AchE), as a muscle marker (25, 36, 37), to evaluate the affectation of mesodermal tissues. Results showed that despite the malformations of the embryos, both notochord and muscle cells were present in knockdown larvae (Fig. 5C-F). Endodermal cells were be visualized in the inner region of the *O. dioica* trunk labeling the digestive track by the activity of alkaline phosphatase (ALP) (25, 38). ALP activity was detected in knockdown larvae, although the extension and intensity of the signal was clearly reduced (Fig. 5G-H). Finally, because B11/B11 blastomeres, in which *Odi_Wnt11a* is asymmetrically located, are the precursor cells of the primordial germ cells (PGC), we investigated whether knocking down *Odi_Wnt11a* might affect PGC determination. We analyzed the expression of *vasa*, a germ-line specific marker (39–41). Immunohistochemistry with an ascidian vasa anti-body (42), labeled two cells in the posterior part of the trunk (Fig.5I-J) in both, treated and control larvae, suggesting that maternal *Odi_Wnt11a* was not necessary for PGC determination.

In summary, the analysis of different cell- and tissue-specific markers indicated that knockdown maternal *Odi_Wnt11a* caused major morphological malformations such as shorter trunks and tails with bended notochords and some impaired endodermal structures, but it did not seem to affect germ-layer specification neither tissue differentiation. Further investigations will be needed to establish whether maternal *Odi_Wnt11a* was not relevant for these developmental processes or whether zygotic expression *Odi_Wnt11a*, which was only slightly affected by dsDNA injections, might be functionally compensating the maternal component.

3. Discussion

3.1. Evolution of the *Wnt* repertoire in *O. dioica*

The repertoire of *Wnt* ligands in deuterostomes –ambulacraria plus chordates– seems, in general, refractory to gene loss. In ambulacraria, for instance, the hemichordate *Saccoglossus kowalevskii* has retained the complete set of Wnt subfamilies (43), while echinoderms have all the Wnt subfamilies except the Wnt11 (12, 44, 45). Similarly, in two of the three chordate subphyla, the cephalochordates and vertebrates, all Wnt subfamilies have been preserved, with the exception of WntA in vertebrates (13). Our previous work on ascidian *Wnt* repertoire (13) and our present analysis in *O. dioica* indicates that, in contrast, the species of urochordate subphylum do not follow the same trend. According to our phylogenetic reconstructions, Wnt4 and Wnt8 subfamilies were lost in both ascidian and appendicularian species, and therefore, likely due to an early loss event in the urochordate ancestor. It is tempting to speculate that this loss might have been relevant for the divergence between urochordate and other chordate subphyla. In addition, Wnt1, Wnt3 and Wnt11 subfamilies were lost in some ascidian species (13), while Wnt1, Wnt2, Wnt3, Wnt6, Wnt7, Wnt9 and WntA subfamilies have been lost during *O. dioica* evolution, which has only retained members in 4 Wnt subfamilies, namely Wnt5, Wnt10, Wnt11 and Wnt16 (Fig. 1). Thus, *O. dioica* (together with *Patella vulgata* (46)) has to our knowledge the minimal Wnt repertoire among all bilaterians analyzed so far.

Interestingly, the loss of Wnt subfamilies during the evolution of the appendicularian lineage has been accompanied by the expansion of the Wnt11 subfamily up to 4 paralogs (*Odi_Wnt11a*, *Odi_Wnt11b*, *Odi_Wnt11c* and *Odi_Wnt11d*), most likely by retrotransposition events from *Odi_Wnt11a* that still retain introns (Fig S1) (47). Often, intronless retroposed gene copies have been viewed as evolutionary dead-ends with little biological relevance due to the lack of regulatory elements. Although this may be case for *Wnt11b* and *Wnt11c* (Fig. S3 and S6), *Wnt11d* clearly showed a specific and dynamic expression pattern (Fig. 2 AD-AH), suggesting that it has achieved a biological role and has recruited regulatory elements that drive its expression.

Interestingly, the reduction in the repertoire of Wnt subfamilies has been accompanied by a reduction in the number of subfamilies of Wnt receptors and of antagonists (Fig. S2B and C), which may suggest a possible parallel gene

elimination (or gene coelimination) in the Wnt activators repertoire (i. e, *Wnt* ligands and receptors) (Fig S2). In addition, the conservation and expression of most of the intermediate effectors and nuclear effectors (Fig. S2A, Fig. S6 and Table S1) would indicate that the signaling pathway is totally functional in this species.

3.2. Comparative analysis of *Wnt* expression during embryogenesis of *O. dioica* and other chordates

It is generally accepted in EvoDevo that orthologous genes in the same subfamily often play conserved functions across evolutionarily distant species (48). The comparison of the expression patterns of *O. dioica Wnt* genes with other chordate species further support this notion. *O. dioica,* however, has lost many Wnt subfamilies*,* and these losses have been possible because these subfamilies have become, somehow, dispensable. Our analysis of *O. dioica* has revealed that Wnt dispensability might be associated to synfunctionalization events (49) –i.e. one paralog acquires the expression domain of another, replacing it– leading to function shuffling when different lineages are compared (50), or caused when a Wnt subfamily has become unnecessary for the development of a given structure in *O. dioica.* Our study of *O. dioica* Wnt genes reveals, therefore, examples of three evolutionary patterns: patterns of functional conservation, patterns of functional shuffling, and patterns of functional extinction.

Regarding the first evolutionary pattern, *O. dioica* orthologs conserve ancestral functions. For instance, *Wnt11* is expressed in endodermal derivates in cephalochordates and vertebrates (13, 51–53). The expression of the *O*di_*Wnt11d* paralog in the endostyle and the endodermal strand (Fig. 2 AE-AG and Fig. S4D), suggest a functional conservation of Wnt11 signaling in the chordate endoderm. Similarly, *O*di_*Wnt11a* is expressed in the posterior part of the notochord (Fig. 2 V-X), likewise *Wnt11* is expressed in the notochord of cephalochordates and vertebrates (54–58). Besides, it has been shown that down-regulation of *Wnt11* expression produces miss elongation of the A-P axis in vertebrates (59–64). Similarly, knocking-down *Odi_Wnt11a* produces *O. dioica* larvae with shortened trunks and tails. It can be argued, therefore, that the ancestral chordate function of *Wnt11* in endoderm and notochord has been preserved in *O. dioica,* although subfunctionalized between paralogs, that is, between endodermal *Odi_Wnt11d* and notochordal *Odi_Wnt11a*.

Regarding the second evolutionary pattern, synfunctionalization of *O. dioica Wnt* genes might appear as function shuffling events when compared with other chordate species. For example, *O. dioica* has lost Wnt6 *and* Wnt7 subfamilies (Fig. 1), which are expressed in the neural crest (NC) and the central nervous system (CNS) in all other chordates (54, 65–69). *Odi_Wnt5* appears to have synfunctionalized in the *O. dioica* lineage compensating for Wnt6 and Wnt7 losses as they are expressed in the nerve cord during *O. dioica* embryogenesis (Fig. 2 D and E) (notice that *Wnt5* is not involved in the development of the neuronal tissues in cephalochordates (54), neither in ascidians (59)). Function shuffling among *Wnt5* and *Wnt6* or *Wnt7* has therefore occurred during the evolution of different chordate lineages. Function shuffling is also observed when comparing the *Wnt* genes responsible for determination of primary body axis. Whereas *Odi_wnt11a* is the main candidate for this role in *O. dioica* (Fig. 3), *Wnt5* in ascidians (55, 67), *Wnt8* in zebrafish (70) and *Wnt11* in *Xenopus* (71) have been associated to this function in these species.

Regarding the third evolutionary pattern, the development of a structure that is Wnt-dependent in chordates has become Wnt-independent in *O. dioica*, leading to the loss of *Wnt* genes. For example, the formation of the gill slits appears to be independent of Wnt signaling in *O. dioica* (i.e. none of the *O. dioica Wnt* genes were expressed in the gill slits), whereas Wnt signaling (together with retinoic acid signaling) appears necessary for the formation of the homologous structures in amphioxus (*Wnt9* and *WntA* are expressed in the gill arches; (13, 72)) and vertebrates (*Wnt2, Wnt4, Wnt5, Wnt7, Wnt9, Wnt11* and *Wn16* are expressed in the pharyngeal ectoderm or mesoderm of gill slits; (73–75)). It is tempting to speculate, therefore, that the loss of many Wnt subfamilies (as well as the loss of the retinoic acid signaling (76)) in *O. dioica* could be related with major changes in the signaling requirements of this species.

3.3. Function of the maternal *Wnt* signaling pathway in *O. dioica*

The analysis of gene expression has shown that several Wnt signaling components of *O. dioica* (e. g. *Odi_Wnt11a*, *Odi_Fzd1/2/7 like, Odi_Fzd3/6a*, *Odi_Dvl*, *Odi_GSK3*, *Odi_β-catenin*…) are maternally expressed (28) and (Fig. S6). Among them, *Odi_Wnt11a* asymmetrically localizes in the posterior-most blastomeres during the firsts cleavages of development (Fig 3) resembling the

expression of *Xenopus Wnt11* and ascidian *Wnt5* in the vegetal-posterior region (55, 77).

This maternal expression of *Odi_Wnt11a* restricted to the posterior sub-cortical region of the posterior-most blastomeres suggests a role of *Wnt* signaling in A/P axis patterning (Fig 3). The localization of this mRNA correlates with the recently described 'cytoplasm' that segregate to the posterior pole of the presumptive germ line blastomeres in *O. dioica*, called postplasm. In the *O. dioica* postplasm, maternal RNA and some morphological structures like the centrosome-attracting body (CAB) localize, reminding the postplasm of urochordate ascidians (41, 78). Interestingly, ascidian postplasm contains *Wnt5* that migrates to the posterior-most blastomeres to contribute to axis formation and cell fate determination (78, 79). Ascidian fertilized eggs injected with *Wnt5* morpholino cannot complete the gastrulation, and the asymmetric separation of the mRNAs necessary for the mesoderm endoderm segregation is also impaired (60, 80). When *Odi_Wnt11a* dsDNA were injected in the ovary of pre-mature *O. dioica* females, however, no defects in the cleavage and gastrulation was noticed, although a decrease in the amount of *Odi_Wnt11a* mRNA during the cleavage stages was observed (Fig. 4D). It could be argued that the function of maternal *O. dioica* and ascidian Wnt signaling is different, or that functional inhibition by dsDNA was not strong enough to alter cleavage and gastrulation in *O. dioica* (Fig 4 and 5A-B). Interestingly, it has been recently shown that knockdown of maternal β-catenin by dsDNA in *O. dioica* prevents the proper specification of the vegetal hemisphere (25). To fully understand the function of maternal Wnt signaling in *O. dioica* axis formation is, therefore, a challenging task that requires additional experiments.

5. Materials and Methods

Laboratory culture of *Oikopleura dioica*. *O. dioica* specimens were obtained from the Mediterranean coast of Barcelona (Catalonia, Spain). Culturing of *O. dioica* and embryo collections have been performed as previously described (21). Ovarian microinjection was performed as previously described (22) and (SI Materials and Methods).

Genome Database Searches and Phylogenetic Analyses. Protein sequences of the *Wnt* genes from vertebrate H. sapiens and urochordate *C. robusta* were used as queries in BLASTp and tBLASTn searches in *O. dioica* genome databases

(http://oikoarrays.biology.uiowa.edu/Oiko/, last accessed May 2016). Homologies of the *O. dioica* sequences were initially assessed by BRBH strategy (81). Homologies were then corroborated by phylogenetic tree analyses based on ML inferences calculated with PhyML v3.0 and automatic mode of selection of substitution model (82) using protein alignments generated with MUSCLE and reviewed by hand (83, 84). Accession numbers for O. dioica sequences and protein alignment for phylogenetic tree reconstruction are provided in Table S1 in SI Materials and Methods.

Gene Expression and Tissue Differentiation. Fragments of *O. dioica* genes were PCR amplified and cloned to synthesize gene-specific riboprobes (Table S2). To reveal *Wnt* expression and evaluate neural tissues and notochord differentiation, whole-mount in situ hybridization on fixed embryos was performed as previously described (36, 76, 85). Nuclear staining (1μM Hoeschst in PBST for 1 hour at room temperature) was included in expression analysis at tailbud stages to confirm muscle cell positions. α-Tubulin A and Brachyury were used as specific markers for neuronal tissues and notochord, respectively (30, 35, 36). Histochemical reaction of acetylcholinesterase (AChE) was used to examine the differentiation of muscle cells, while histochemical staining for alkaline phosphatase was used to monitor the differentiation of endoderm cells (25, 37, 38). For germ-line differentiation, immunohistochemistry using an antibody against *Ciona robusta* Vasa homolog was carried out as previously reported (86). The primary antibody used was affinity-purified rabbit anti-CiVH (1:500 dilution) (42) and the secondary antibody used was Alexa Fluor 594 goat anti-rabbit IgG (1:500 dilution; Life Technologies).

Bibliography

1. Hikasa H, Sokol SY (2013) Wnt signaling in vertebrate axis specification. *Cold Spring Harb Perspect Biol* 5(1):1–20.
2. Liu P, et al. (1999) Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* 22(4):361–5.
3. ten Berge D, Brugmann SA, Helms JA, Nusse R (2008) Wnt and FGF signals interact to coordinate growth with cell fate specification during limb development. *Development* 135(19):3247–57.
4. Marvin MJ, Di Rocco G, Gardiner A, Bush SM, Lassar AB (2001) Inhibition of Wnt activity induces heart formation from posterior mesoderm. *Genes Dev* 15(3):316–27.

5. Ille F, Sommer L (2005) Wnt signaling: multiple functions in neural development. *Cell Mol Life Sci* 62(10):1100–8.
6. Loh KM, van Amerongen R, Nusse R (2016) Generating Cellular Diversity and Spatial Form: Wnt Signaling and the Evolution of Multicellular Animals. *Dev Cell* 38(6):643–655.
7. Kusserow A, et al. (2005) Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433(7022):156–160.
8. Schenkelaars Q, et al. (2017) Animal multicellularity and polarity without Wnt signaling. *Sci Rep* 7(1):15383.
9. Hensel K, Lotan T, Sanders SM, Cartwright P, Frank U (2014) Lineage-specific evolution of cnidarian Wnt ligands. *Evol Dev* 16(5):259–269.
10. Janssen R, et al. (2010) Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation. *BMC Evol Biol* 10:374.
11. Riddiford N, Olson PD (2011) Wnt gene loss in flatworms. *Dev Genes Evol* 221(4):187–97.
12. Robert N, Lhomond G, Schubert M, Croce JC (2014) A comprehensive survey of wnt and frizzled expression in the sea urchin Paracentrotus lividus. *Genesis* 52(3):235–50.
13. Somorjai IML, et al. (2018) Wnt evolution and function shuffling in chordate genomes: from conservative amphioxus stasis to liberal ascidian gains and losses. *Genome Biol*.
14. Albalat R, Cañestro C (2016) Evolution by gene loss. *Nat Rev Genet* 17(7):379–391.
15. Doumpas N, Jékely G, Teleman AA (2013) Wnt6 is required for maxillary palp formation in Drosophila. *BMC Biol* 11:104.
16. Denoeud F, et al. (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330(6009):1381–5.
17. Søviknes AM, Chourrout D, Glover JC (2005) Development of putative GABAergic neurons in the appendicularian urochordate Oikopleura dioica. *J Comp Neurol* 490(1):12–28.
18. Søviknes AM, Glover JC (2008) Continued growth and cell proliferation into adulthood in the notochord of the appendicularian Oikopleura dioica. *Biol Bull* 214(1):17–28.
19. Bouquet JM, et al. (2009) Culture optimization for the emergent zooplanktonic model organism Oikopleura dioica. *J Plankton Res* 31(4):359–370.

20. Nishida H (2008) Development of the appendicularian Oikopleura dioica: Culture, genome, and cell lineages. *Dev Growth Differ* 50(SUPPL. 1):239–256.
21. Martí-Solans J, et al. (2015) Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* 53(1):183–93.
22. Omotezako T, Nishino A, Onuma TA, Nishida H (2013) RNA interference in the appendicularian Oikopleura dioica reveals the function of the Brachyury gene. *Dev Genes Evol* 223(4):261–267.
23. Mikhaleva Y, Kreneisz O, Olsen LC, Glover JC, Chourrout D (2015) Modification of the larval swimming behavior in Oikopleura dioica, a chordate with a miniaturized central nervous system by dsRNA injection into fertilized eggs. *J Exp Zool Part B Mol Dev Evol* 324(2):114–127.
24. Omotezako T, Onuma TA, Nishida H (2015) DNA interference: DNA-induced gene silencing in the appendicularian Oikopleura dioica. *Proc R Soc B* 282(1807):20150435.
25. Omotezako T, Matsuo M, Onuma TA, Nishida H (2017) DNA interference-mediated screening of maternal factors in the chordate Oikopleura dioica. *Nat Publ Gr* (March):1–10.
26. Cho SJ, Vallès Y, Giani VC, Seaver EC, Weisblat DA (2010) Evolutionary dynamics of the wnt gene family: A lophotrochozoan perspective. *Mol Biol Evol* 27(7):1645–1658.
27. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182-5.
28. Danks G, et al. (2013) OikoBase: A genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. *Nucleic Acids Res* 41(D1):1–9.
29. Olsson R, Holmberg K, Lilliemarck Y (1990) Fine structure of the brain and brain nerves ofOikopleura dioica (Urochordata, Appendicularia). *Zoomorphology* 110(1):1–7.
30. Søviknes AM, Chourrout D, Glover JC (2007) Development of the caudal nerve cord, motoneurons, and muscle innervation in the appendicularian urochordate Oikopleura dioica. *J Comp Neurol* 503(2):224–43.
31. Delsman HC (1910) Beiträge zur Entwicklungsgeschichite von Oikopleura dioica. *Verch Rijksinst Odnderz Zee* 3(2):1–24.
32. Conklin EG (1905) *The organization and cell-lineage of the ascidian egg / by Edwin G. Conklin.* ([Academy of Natural Sciences], Philadelphia :) doi:10.5962/bhl.title.4801.

33. Stach T, Winter J, Bouquet J-M, Chourrout D, Schnabel R (2008) Embryology of a planktonic tunicate reveals traces of sessility. *Proc Natl Acad Sci U S A* 105(20):7229–7234.
34. Fujii S, Nishio T, Nishida H (2008) Cleavage pattern, gastrulation, and neurulation in the appendicularian, Oikopleura dioica. *Dev Genes Evol* 218(1–2):69–79.
35. Seo H-C, et al. (2004) Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* 431(7004):67–71.
36. Bassham S, Postlethwait J (2000) Brachyury (T) Expression in Embryos of a Larvacean Urochordate, Oikopleura dioica, and the Ancestral Role of T. *Dev Biol* 220(2):322–332.
37. Nishino A, Satou Y, Morisawa M, Satoh N (2000) Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: Phylogenetic relationships among muscle tissues in the urochordates. *J Exp Zool* 288(2):135–150.
38. Imai K, Takada N, Satoh N, Satou Y (2000) (beta)-catenin mediates the specification of endoderm cells in ascidian embryos. *Development* 127(14):3009–20.
39. Ganot P, Bouquet JM, Kallesøe T, Thompson EM (2007) The Oikopleura coenocyst, a unique chordate germ cell permitting rapid, extensive modulation of oocyte production. *Dev Biol* 302(2):591–600.
40. Henriet S, et al. (2015) Embryonic expression of endogenous retroviral RNAs in somatic tissues adjacent to the Oikopleura germline. *Nucleic Acids Res* 43(7):3701–3711.
41. Olsen LC, et al. (2018) Evidence for a centrosome-attracting body like structure in germ-soma segregation during early development, in the urochordate Oikopleura dioica. *BMC Dev Biol* 18(1):4.
42. Shirae-Kurabayashi M (2006) Dynamic redistribution of vasa homolog and exclusion of somatic cell determinants during germ cell specification in Ciona intestinalis. *Development* 133(14):2683–2693.
43. Darras S, et al. (2018) Anteroposterior axis patterning by early canonical Wnt signaling during hemichordate development. *PLoS Biol* 16(1):e2003698.
44. Croce JC, et al. (2006) A genome-wide survey of the evolutionarily conserved Wnt pathways in the sea urchin Strongylocentrotus purpuratus. *Dev Biol* 300(1):121–131.
45. McCauley BS, Akyar E, Filliger L, Hinman VF (2013) Expression of wnt and frizzled genes during early sea star development. *Gene Expr Patterns* 13(8):437–

44.
46. Prud'homme B, Lartillot N, Balavoine G, Adoutte A, Vervoort M (2002) Phylogenetic analysis of the Wnt gene family. Insights from lophotrochozoan members. *Curr Biol* 12(16):1395.
47. Kaessmann H, Vinckenbosch N, Long M (2009) RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* 10(1):19–31.
48. Dolinski K, Botstein D (2007) Orthology and Functional Conservation in Eukaryotes. *Annu Rev Genet* 41(1):465–507.
49. Gitelman I (2007) Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. *Mol Biol Evol* 24(9):1912–25.
50. McClintock JM, Carlson R, Mann DM, Prince VE (2001) Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. *Development* 128(13):2471–84.
51. Schubert M, Holland LZ, Holland ND (2000) Characterization of an amphioxus wnt gene, AmphiWnt11, with possible roles in myogenesis and tail outgrowth. *Genesis* 27(1):1–5.
52. Cui S, Capecci LM, Matthews RP (2011) Disruption of planar cell polarity activity leads to developmental biliary defects. *Dev Biol* 351(2):229–41.
53. Sinner D, et al. (2006) Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development* 133(10):1955–66.
54. Schubert M, Holland LZ, Stokes MD, Holland ND (2001) Three amphioxus Wnt genes (AmphiWnt3, AmphiWnt5, and AmphiWnt6) associated with the tail bud: the evolution of somitogenesis in chordates. *Dev Biol* 240(1):262–73.
55. Sasakura Y, Ogasawara M, Makabe KW (1998) HrWnt-5: A maternally expressed ascidian Wnt gene with posterior localization in early embryos. *Int J Dev Biol* 42(4):573–579.
56. Makita R, Mizuno T, Koshida S, Kuroiwa A, Takeda H (1998) Zebrafish wnt11: Pattern and regulation of the expression by the yolk cell and No tail activity. *Mech Dev* 71(1–2):165–176.
57. Baranski M, Berdougo E, Sandler JS, Darnell DK, Burrus LW (2000) The dynamic expression pattern of frzb-1 suggests multiple roles in chick development. *Dev Biol* 217(1):25–41.
58. Andre P, Song H, Kim W, Kispert A, Yang Y (2015) Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* 142(8):1516–27.
59. Niwano T, Takatori N, Kumano G, Nishida H (2009) Wnt5 is required for

notochord cell intercalation in the ascidian Halocynthia roretzi. *Biol Cell* 101(11):645–659.

60. Nakamura Y, Makabe KW, Nishida H (2006) The functional analysis of Type I postplasmic/PEM mRNAs in embryos of the ascidian Halocynthia roretzi. *Dev Genes Evol* 216(2):69–80.

61. Rauch GJ, et al. (1997) Wnt5 is required for tail formation in the zebrafish embryo. *Cold Spring Harb Symp Quant Biol* 62:227–34.

62. Yamaguchi TP, Bradley A, McMahon AP, Jones S (1999) A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* 126(6):1211–23.

63. Tada M, Smith JC (2000) Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* 127(10):2227–38.

64. Heisenberg CP, et al. (2000) Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* 405(6782):76–81.

65. Schubert M, Holland LZ, Holland ND (2000) Characterization of two amphioxus Wnt genes (AmphiWnt4 and AmphiWnt7b) with early expression in the developing central nervous system. *Dev Dyn* 217(2):205–15.

66. Sasakura Y, Makabe KW (2000) Ascidian Wnt-7 gene is expressed exclusively in the tail neural tube of tailbud embryos. *Dev Genes Evol* 210(12):641–643.

67. Imai KS, Hino K, Yagi K, Satoh N, Satou Y (2004) Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* 131(16):4047–58.

68. García-Castro MI, Marcelle C, Bronner-Fraser M (2002) Ectodermal Wnt function as a neural crest inducer. *Science* 297(5582):848–51.

69. Martin A, Maher S, Summerhurst K, Davidson D, Murphy P Differential deployment of paralogous Wnt genes in the mouse and chick embryo during development. *Evol Dev* 14(2):178–95.

70. Lu F-I, Thisse C, Thisse B (2011) Identification and mechanism of regulation of the zebrafish dorsal determinant. *Proc Natl Acad Sci* 108(38):15876–15880.

71. Tao Q, et al. (2005) Maternal Wnt11 activates the canonical Wnt signaling pathway required for axis formation in Xenopus embryos. *Cell* 120(6):857–871.

72. Onai T, et al. (2009) Retinoic acid and Wnt/β-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus. *Dev Biol* 332(2):223–233.

73. Geetha-Loganathan P, et al. (2009) Expression of WNT signalling pathway genes during chicken craniofacial development. *Dev Dyn* 238(5):1150–65.
74. Curtin E, Hickey G, Kamel G, Davidson AJ, Liao EC Zebrafish wnt9a is expressed in pharyngeal ectoderm and is required for palate and lower jaw development. *Mech Dev* 128(1–2):104–15.
75. Choe CP, et al. (2013) Wnt-dependent epithelial transitions drive pharyngeal pouch formation. *Dev Cell* 24(3):296–309.
76. Martí-Solans J, et al. (2016) Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. *Mol Biol Evol* 33(9):2401–2416.
77. Ku M, Melton DA (1993) Xwnt-11: a maternally expressed Xenopus wnt gene. *Development* 119(4):1161–73.
78. Makabe KW, Nishida H (2012) Cytoplasmic localization and reorganization in ascidian eggs: role of postplasmic/PEM RNAs in axis formation and fate determination. *Wiley Interdiscip Rev Dev Biol* 1(4):501–18.
79. Prodon F, Yamada L, Shirae-Kurabayashi M, Nakamura Y, Sasakura Y (2007) Postplasmic/PEM RNAs: a class of localized maternal mRNAs with multiple roles in cell polarity and development in ascidian embryos. *Dev Dyn* 236(7):1698–715.
80. Takatori N, Kumano G, Saiga H, Nishida H (2010) Segregation of germ layer fates by nuclear migration-dependent localization of Not mRNA. *Dev Cell* 19(4):589–98.
81. Wall DP, Fraser HB, Hirsh AE (2003) Detecting putative orthologs. *Bioinformatics* 19(13):1710–1.
82. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–21.
83. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
84. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–7.
85. Cañestro C, Postlethwait JH (2007) Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* 305(2):522–538.
86. Onuma TA, Matsuo M, Nishida H (2017) Modified whole-mount in situ hybridisation and immunohistochemistry protocols without removal of the vitelline membrane in the appendicularian Oikopleura dioica. *Dev Genes Evol* 227(5):367–374.

87. Wilkerson MD, Ru Y, Brendel VP (2009) Common introns within orthologous genes: software and application to plants. *Brief Bioinform* 10(6):631–44.

88. Fawal N, Savelli B, Dunand C, Mathé C (2012) GECA: a fast tool for gene evolution and conservation analysis in eukaryotic protein families. *Bioinformatics* 28(10):1398–9.

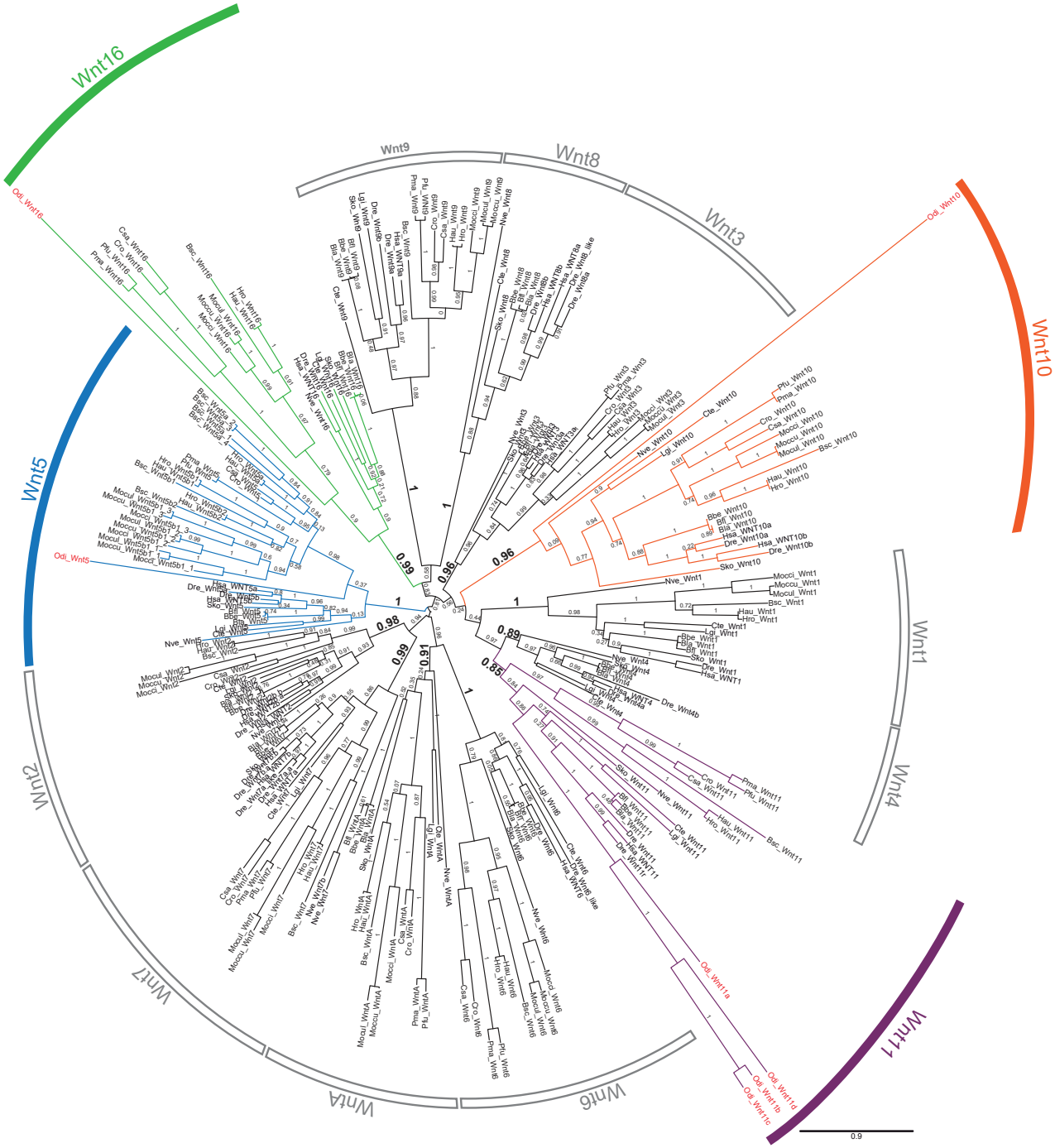

Wnt16
Wnt9
Wnt8
Wnt3
Wnt10
Wnt5
Wnt1
Wnt4
Wnt2
Wnt7
WntA
Wnt6
Wnt11
0.9

Fig. 1. Phylogenetic analysis of the *Wnt* family. ML phylogenetic tree of the *Wnt* family in chordates revealing that *Wnt* repertoire in *Oikopleura dioica* (names in red) was configured, at least, by 2 duplications and 7 losses. The scale bar indicates amino-acid substitutions. Values for the approximate likelihood-ratio test (aLRT) are shown at nodes. Species abbreviations: Chordate species: Urochordates: *Botryllus schlosseri* (Bsc), *Ciona savignyi* (Csa), *Ciona robusta* (Cro; formerly *Ciona intestinallis*), *Halocynthia roretzi* (Hro), *Halocynthia aurantium* (Hau), *Mogula occulta* (Moccu), *Mogula oculata* (Mocul), *Mogula occidentalis* (Mocci), *Phallusia fumigata* (Pfu) and *Phallusia mammillata* (Pma); Cephalochordates: *Branchiostoma belcheri* (*Bbe*), *Branchiostoma floridae* (*Bfl*), *Branchiostoma lanceolatum* (*Bla*); Vertebrates: *Danio rerio* (Dre), *Homo sapiens* (*Hsa*). Non-chordates species: hemichordate *Saccoglossus kowalevskii* (Sko), annelid *Capitella teleta* (Cte), mollusk *Lottia gigantea* (Lgi) and cnidarian *Nematostella vectensis* (Nve).

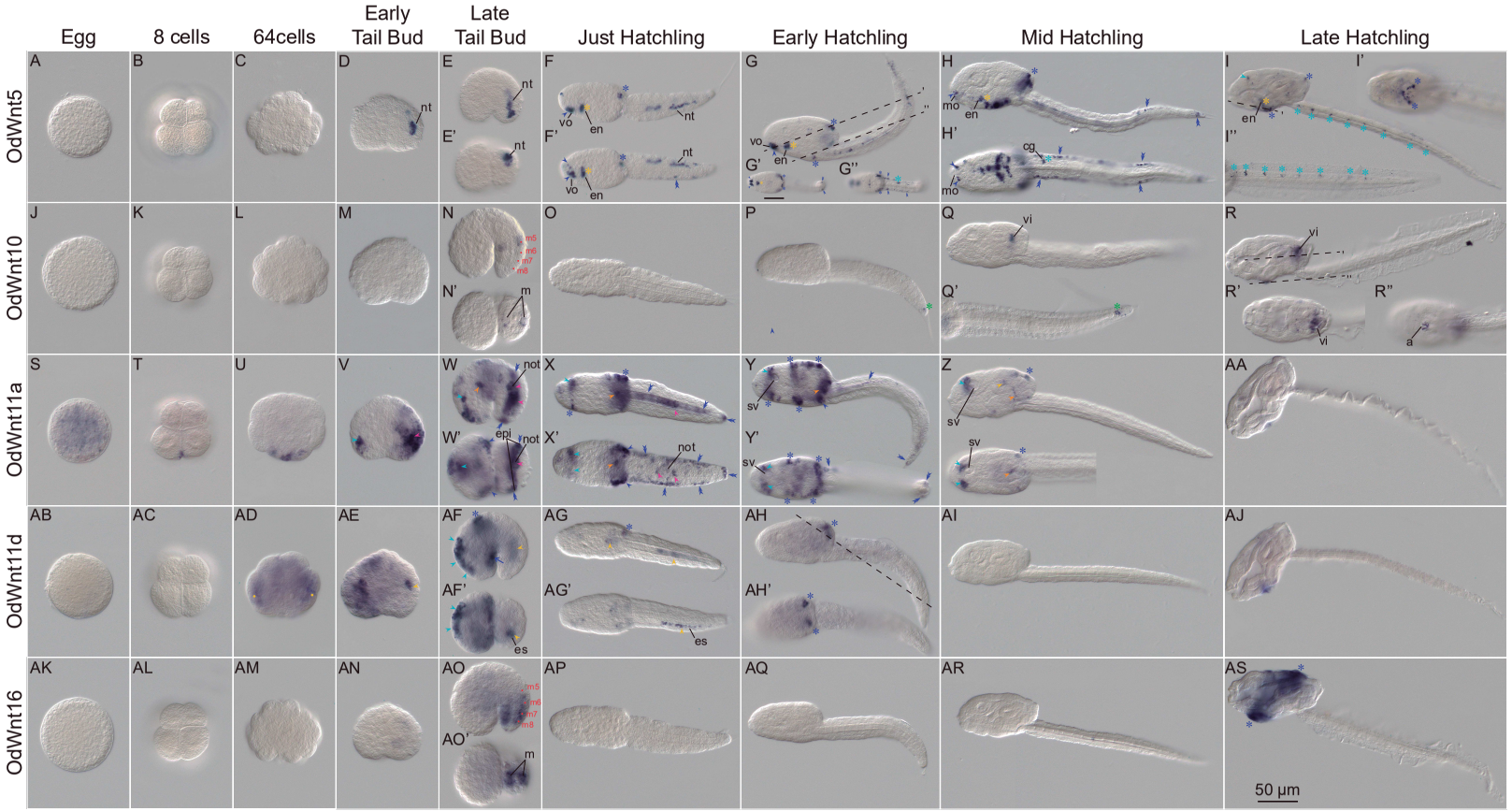


Fig. 2. Expression patterns of *Oikopleura dioica Wnt* genes during embryonic and larval development. Whole-mount in situ hybridization in O. dioica eggs (A, J, S, AB and AK), 8 cells embryos (B, K, T, AC and AL), 62 cells embryos (C, L, U, AD and AM),early tail bud embryos (D, M, V, AE and AN), late tail bud embryos (E, N, W, AF and AO), just hatchlings (F, O, X, AG and AP), early hatchlings (G, P, Y, AH and AQ), mid hatchlings (H, Q, Z, AI and AR) and late hatchlings (I, R, AA, AJ and AS). Upper image of each panel corresponds to left lateral view oriented anterior toward the left and dorsal toward the top. Bottom images (') are ventral views of optical cross sections at the levels of the dashed lines. Red dots point muscle cell nuclei; magenta arrowheads point notochord; orange arrowheads point mesoendoderm; yellow arrowheads point endodermal strand; yellow asterisks points endostyle; light blue arrowheads point neuroectoderm and anterior brain; blue light asterisks point neural tube; blue dark arrowheads point sensory epidermis; blue dark asterisks point oikoplastic epithelium; blue dark double arrowheads point caudal epithelium and blue dark arrow points posterior trunk epidermis. a, anus; cg, caudal ganglion; en, endostyle; epi, epiderms; es, endodermal strand, m, muscle cell pair; mo, mouth; not, notochord; nt, neural tube; sv, sensory vesicle; vi, vertical intestine; vo, ventral organ. Scale bar = 50 µm.

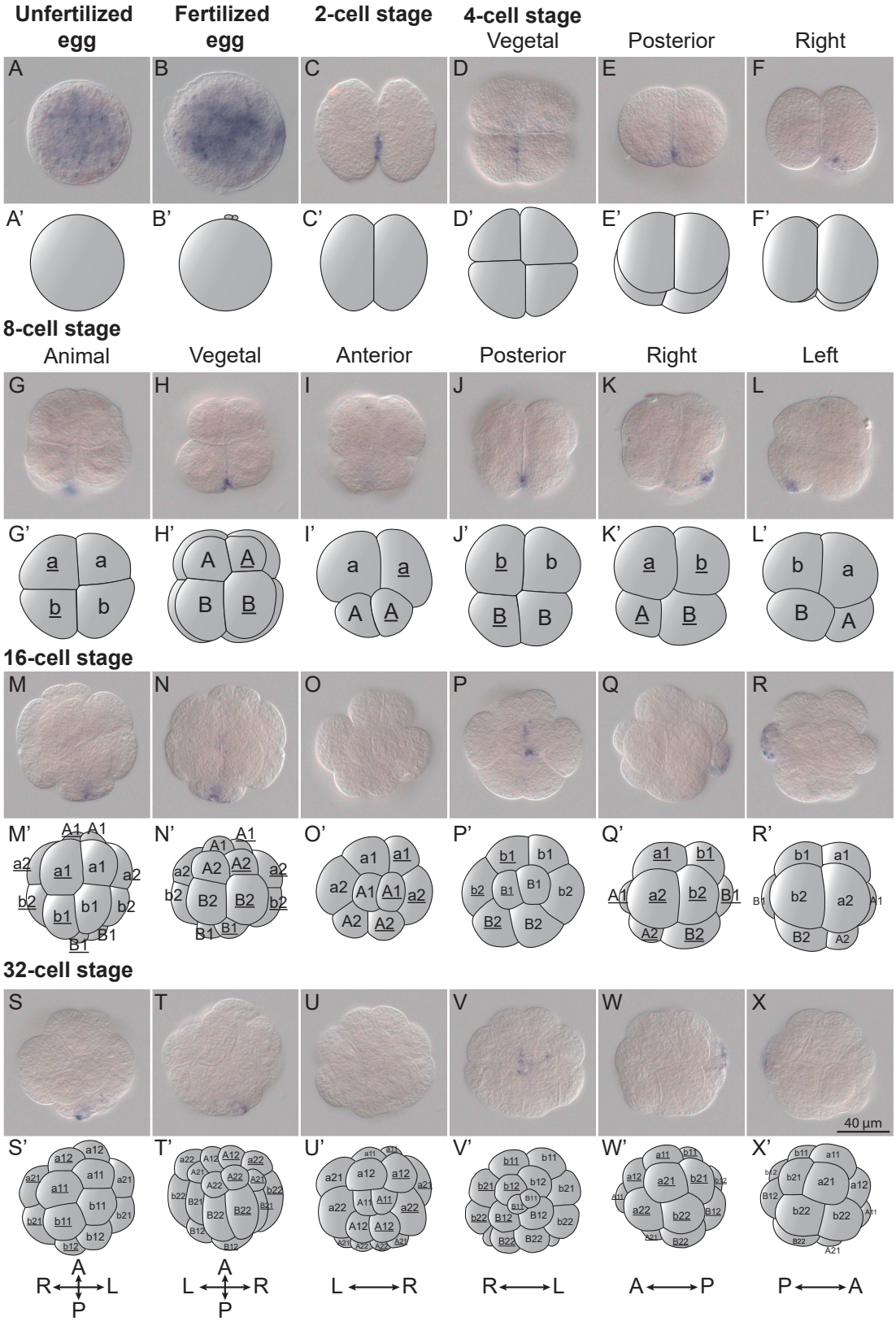


Fig. 3 *Wnt11a* expression in cleavage stages.

Fig. 3 *Wnt11a* expression in cleavage stages. Whole-mount in situ hybridization in *O. dioica* from egg up to the 32-cell stage revealing an asymmetric localization of the *Wnt11a* maternal transcripts. A. Unfertilized egg. B. Fertilized egg. C. 2 cells stage. D-F. 4 cells stage. G-L. 8 cells stage. M-R. 16 cells stage. S-X. 32 cells stage. Schematic representations (‘) adapted form (34) and blastomere names according to Delsman’s nomenclature are given. Embryos were viewed from various directions indicated at the top. Scale bar = 40 µm.

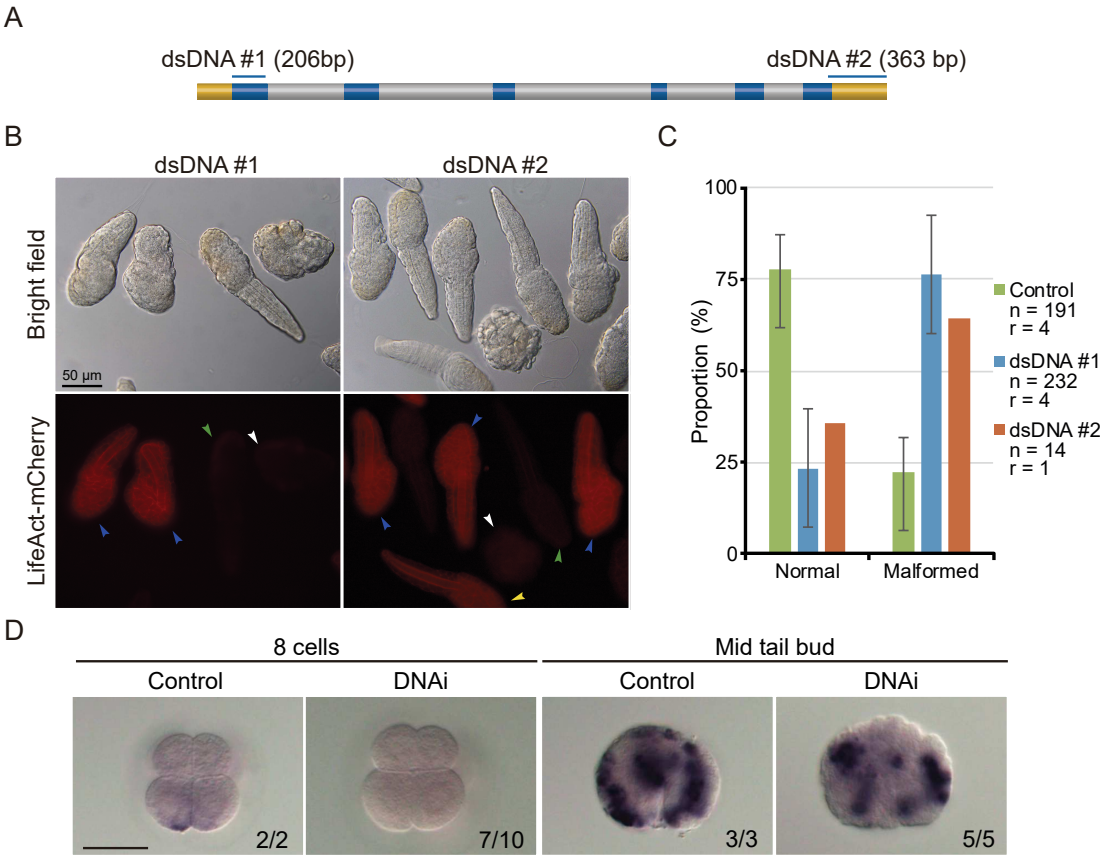


Fig. 4. *Wnt11a* knockdown phenotypes. A. Regions targeted by each of the PCR products. Upper box is a gene model of *Wnt11a* and the regions targeted by each of the PCR products are shown by blue bars above. Blue, grey and yellow boxes indicate the ORF, intron and UTR, respectively. B. Embryos co-injection with *Wnt11a* PCR products (0.2 µg/µl) and Lifeact-mCherry mRNA (0.7 µg/µl). Only mCherry positive embryos were examined. Phenotypes were categorized into two groups: normal and short. Two-thirds of the animals with fluorescence show shortened trunk and tail (blue arrowheads), although some animals with fluorescence did not show malformations (yellow arrowheads). Most of the animals without fluorescence developed normally (green arrowheads), while some embryos without fluorescence underwent developmental arrested before hatching (white arrowhead). C. Proportions of each of the phenotype of hatched larvae after injection with the three kinds of PCR products. Blue, orange and green bars show the result of DNAi #1, DNAi #2 and Control (Kaede PCR product), respectively. Number of analyzed embryos (n) and number replicas (r) are indicated in the legend. Scale bar = 50 µm.

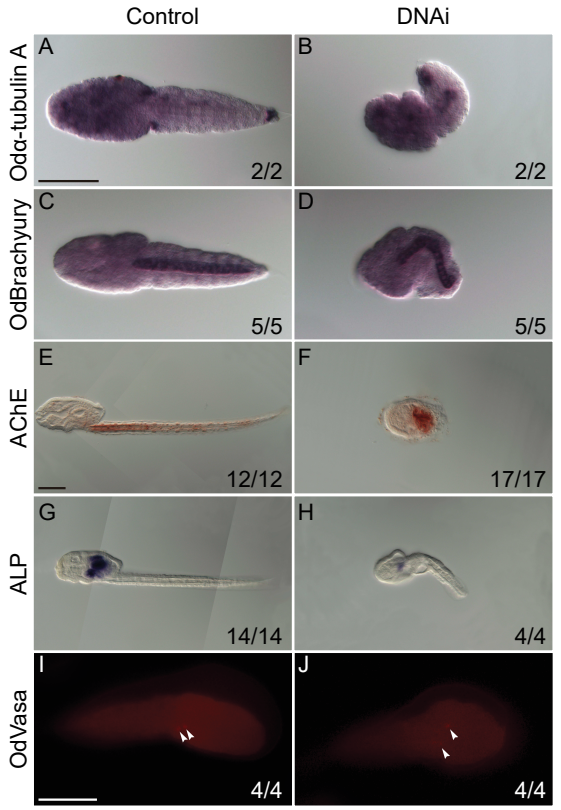


Fig. 5. Molecular characterization of *Wnt11a* knockdown. A-D. Whole-mount in situ hybridization for *Wnt11a* in 8 cells embryos (A and B) and mid tail bud stage (C and D) injected with Kaede PCR product as control and *Wnt11a* DNAi #1 PCR product (see Fig. 4A). E and F. Whole-mount in situ hybridization for the neuronal marker α-Tubulin A in just hatchlings injected with Kaede PCR product and *Wnt11a* DNAi #1 PCR product. G and H. Whole-mount in situ hybridization for the notochord marker Brachyury in just hatchlings injected with Kaede PCR product and *Wnt11a* DNAi #1 PCR product (I and J) Histochemical staining for the muscle differentiation marker acetylcholinesterase (AChE) in larvae injected with Kaede PCR product and *Wnt11a* DNAi #1 PCR product. K and L. Histochemical staining for the endoderm differentiation marker alkaline phosphatase (ALP) in larvae injected with Kaede PCR product and *Wnt11a* DNAi #1 PCR product. M and N. Immunohistochemistry for the germ line marker vasa in just hatchlings injected with Kaede PCR product and *Wnt11a* DNAi #1 PCR product. The ratio, at the bottom of each panel, indicates the proportion of embryos with the phenotype shown. Scale bar = 50 μm.

Supplementary Information for

Oikopleura dioica Wnt signaling: when gene loss, duplication and function shuffling push the limits of chordate EvoDevo

Josep Martí-Solans[1], Hector Godoy[1], Miriam Diaz-Gracia[1], Takeshi Onuma[2], Hiroki Nishida[2], Ricard Albalat[1]✉ and Cristian Cañestro[1]✉

[1]Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

[2]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

This PDF file includes:

Supplementary text
Figs. S1 to S6
Tables S1 and S2

SI Materials and Methods

Preparation of PCR products for microinjection. Each *Wnt* gene was amplified from a cDNA obtained from late hatchlings larvae with gene-specific primers (Table S2), cloned in the pCR4- TOPO vector (Invitrogen), transformed in TOP10 E. coli competent cells (Invitrogen), and sequenced using vector flanking primers. PCR products for microinjection were amplified using KOD plus (TOYOBO) and a specific inner pair of primers (Table S2). Single bands for each gene were obtained. Then, products were purified by phenol–chloroform and ethanol precipitation (adjusting the concentration of monovalent cations with sodium acetate at 0.3 M final concentration and pH 5.2). The DNA pellet was dissolved in 15 µL of water achieving a concentration of 0.5 µg/µL, approximately.

Preparation of the mRNA for microinjection. To generate the Lifeact-mCherry mRNA the pSD64TF-Lifeact-mCherry construct was used. Briefly, this construct is the Lifeact sequence amplified by PCR (forward primer, AAATTCTCGAGTCCACCATGGGTGTCGCAGATTTGAT; reverse primer, ACGTAGGGCCCTGGCGACCGGTGGATCC3) using pCMV Lifeact-TagRFP (Ibidi) as a template and subcloned into the XhoI/ApaI restriction sites of the pSD64TF-H2B-EGFP vector, which include the SP6 polymerase promoter and the 5'- and 3'-UTR sequences of the β-globin mRNA of *Xenopus laevis*. (22, 25). To generate the final Lifeact-mCherry, the EGFP sequence of the Lifeact-EGFP construct is replaced by the mCherry cDNA (22). For mRNA synthesis, the pSD64TF-Lifeact-mCherry plasmids was linearized with XbaI and used as template for in vitro transcription. Capped mRNA was synthesized with the SP6 of the mMESSAGE mMACHINE kit (Ambion) and polyadenylated with a Poly(A) Tailing Kit (Ambion). mRNA was purified by phenol-chloroform extraction and isopropanol precipitation, mRNA was dissolved in 20 µl of water achieving a concentration of 1.6 µg/µL, approximately.

Microinjection into the ovary. The ovary of *O. dioica* is a coenocyst where each pro-oocyte shares a common cytoplasm with the rest of the pro-oocytes (39). Thus, any liquid injected into the ovary spreads to a large extent of the gonad and is incorporated, in a gradient manner, into 20–30% of spawned eggs. For injection, day 4 females, i. e. before oocytes became evident in the ovary, were selected, anesthetized with 0.015 % of Ethyl 3-aminobenzoate methanesulfonate salt (Sigma-Aldrich A5040) in seawater and set down on agar in a small drop of seawater. To immobilize the females the seawater was drained with a needle. Then, using a

micropipet mounted on a micromanipulator, the nucleic acids were injected in the inner wall of the ovary at the bottom of the gonad near to the junction of the trunk and tail. Approximately 1 nl of nucleic acid solution was injected with 1 mg/ml phenol red, to visualize the volume to be microinjected. mRNA encoding Lifeact-mCherry (0.6 µg/µL final concentration), as a marker of the nucleic acid incorporation, was co-injected with PCR products (0.2 µg/µL final concentration). After injection, each animal was placed separately in a six-well multititer plate coated with gelatin. 12 hours after injection, the animals spawned the eggs. Only eggs with mCherry fluorescence were fertilized and used for analysis, although there was a continuous distribution of different intensities from strongly fluorescent to non-fluorescent eggs.

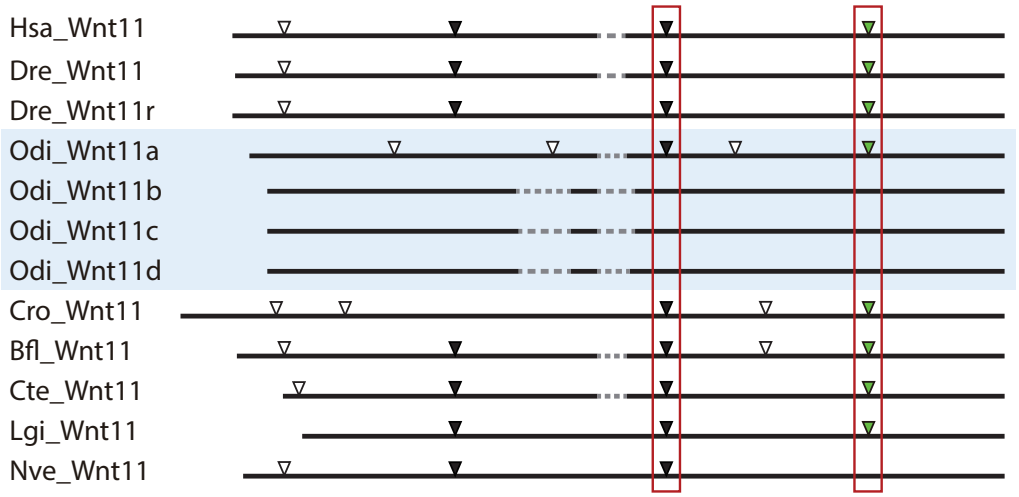


Fig. S1. Schematic comparison of intron positions between human, ascidian and *O. dioica* genes. The CIWOG program (87) from GECA package (88) was used to compare the intron/exon organization of the different orthologous genes based on position and sequence conservation in the corresponding protein alignments, with no restrictions on the proportion of identical amino acids required to define a common intron. Black horizontal lines represent aligned sequences and arrowheads represent intron positions. Black arrowheads indicate introns conserved in all Wnt subfamilies across all metazoans, while white arrowheads indicate non-conserved introns. Green arrowheads indicate Wnt11 subfamily specific intron shared between bilaterians. Red boxes indicate *O. dioica* conserved intron positions.

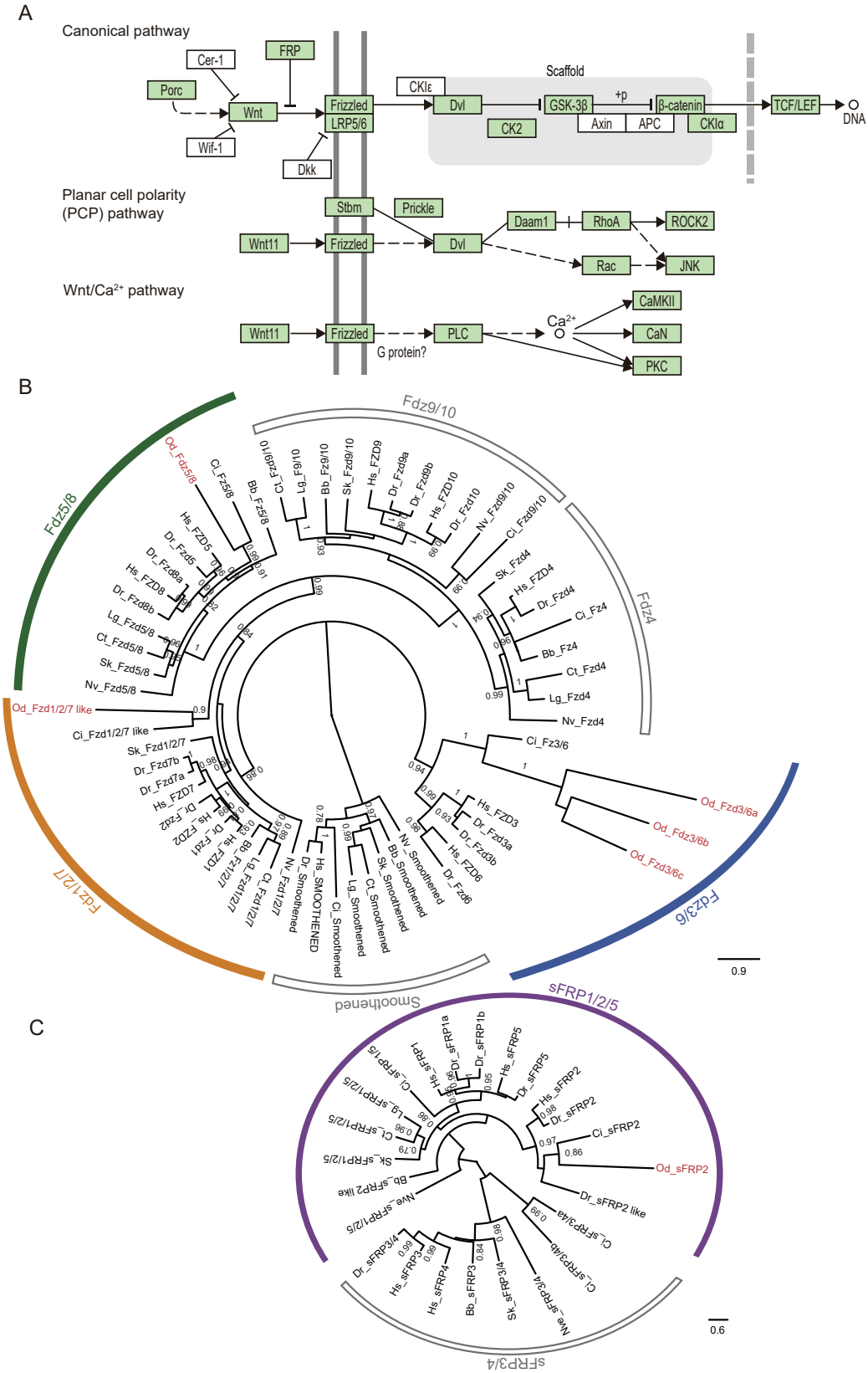


Fig. S2. Wnt signaling components in *Oikopleura dioica*.

Fig. S2. *Wnt* signaling components in *Oikopleura dioica*. A. KEGG Automatic Annotation server (KAAS) analysis of the *Wnt* pathway in *O. dioica*. BLAST program, bi-directional best hit (BBH) method and gene database of *Homo sapiens*; *Danio rerio*; *Ciona intestinalis*; *Branchiostoma floridae*; *Saccoglossus kowalevskii* and *Lottia gigantea* were used as parameters to obtain similar sequences for all *O. dioica*'s peptides and manual searches were performed to ensure the absence of the missing genes (27). Green boxes highlight the assigned KEGG Orthology (KO), while white boxes refer to the failure in assignment to the specific KOs. B. The ML phylogenetic tree of the Fzd/Smoothened family showed the presence of three Fzd orthologues in *O. dioica* genome (red), corresponding to Fzr1/2/7, Fzd 3/6 (x3) and Fzd5/8, and the loss of Fzd4, Fzd9/10 and Smoothened orthologues. Smoothened family was used as outgroup to root the tree. C. The ML phylogenetic tree of the sFRP family showed that the single sFRP orthologue in *O. dioica* genome (red) is grouped with the sFRP2 genes. Values for the approximate likelihood-ratio test (aLRT) are only shown in nodes with support values greater than 0.75. Scale bar indicates amino-acid substitutions. The protein ID (GenBank accession number) used to elaborate the trees are as follows: Oikopleura dioica: Od_Fdz1/2/7 like (GSOIDP00007729001) Od_Fzd3/6a (GSOIDP00008718001), Od_Fdz3/6b (GSOIDP00000671001), Od_Fzd3/6c (GSOIDP00006033001), Od_Fdz5/8 (GSOIDP00000721001), Od_sFRP2 like (GSOIDP00005455001); Homo sapiens: Hs_Fzd1 (Q9UP38), Hs_Fzd2 (Q14332), Hs_Fzd3 (Q9NPG1), Hs_Fzd4 (Q9NPG1), Hs_Fzd5 (Q13467), Hs_Fzd6 (O60353), Hs_Fzd7 (O75084), Hs_Fzd8 (Q9H461), Hs_Fzd9 (O00144), Hs_Fzd10 (Q9ULW2), Hs_Smoothened (Q99835); Hs_sFRP1 (NP_003003), Hs_sFRP2 (NP_003004), Hs_sFRP3 (NP_001454), Hs_sFRP4 (NP_003005), Hs_sFRP5 (AAD25052); Danio rerio: Dr_Fzd1 (NP_001124086), Dr_Fzd2 (NP_571215), Dr_Fzd3a (NP_001036226), Dr_Fzd3b (NP_001074070), Dr_Fzd4 (NP_001292398), Dr_Fzd5 (NP_571209), Dr_Fzd6 (NP_956855), Dr_Fzd7a (NP_571214), Dr_Fzd7b (NP/739569), Dr_Fzd8a (NP_570993), Dr_Fzd8b (NP_571628), Dr_Fzd9a (XP_003198734), Dr_Fzd9b (NP_571586), Dr_Fzd10 (NP_570992), Dr_Smoothened (NP_571102), Dr_sFRP1a (NP_991148), Dr_sFRP1b (NP_001077040), Dr_sFRP2 (NP_001070852), Dr_sFRP2 like (XP_003200152), Dr_sFRP3/4 (NP_571018), Dr_sFRP5 (NP_571933); Ciona intestinalis: Ci_Fzd1/2/7 (NP_001071791), Ci_Fz3/6 (NP_001071723), Ci_Fz4 (XP_018667713), Ci_Fz5/8 (XP_009859677), Ci_Fzd9/10 (XP_002125798), Ci_Smoothened (XP_002125798), Ci_sFRP1/5 (NP_001071964), Ci_sFRP2 (NP_001072004), Ci_sFRP3/4a (NP_001071813), Ci_sFRP3/4b (NP_001071812); Branchiostoma belcheri: Bb_Fzd1/2/7 (AHB53231), Bb_Fzd4 (AHB53232), Bb_Fzd5/8 (AHB53233), Bb_Fzd9/10 (AHB53234), Bb_Smoothened (XP_019641686), Bb_sFRP2 like (AED89555), Bb_sFRP3 (XP_019632159); Saccoglossus kowalevskii: Sk_Fzd1/2/7 (XP_006820151), Sk_Fzd4 (XP_002730495), Sk_Fzd5/8 (NP_001161547), Sk_Fzd9/10 (XP_006817188), Sk_Smoothened (XP_006817784), ; Capitella teleta: Ct_Fzd1/2/7 (ELT94235), Ct_Fzd4 (ELU03627), Ct_Fzd5/8 (ELT92800), Ct_Fzd9/10 (ELU07504), Ct_Smoothened (ELT97156), Ct_sFRP1/2/5 (ELU17320); Lottia gigantea: Lg_Fzd1/2/7 (XP_009047666), Lg_Fzd4 (XP_009050162), Lg_Fzd5/8 (XP_009048933), Lg_F9/10 (XP_009060024), Lg_Smoothened (XP_009064048), Lg_sFRP1/2/5 (XP_009051836); Nematostella vectensis: Nv_Fzd1/2/7 (XP_001647540), Nv_Fzd4 (XP_001622965), Nv_Fzd5/8 (XP_001634995), Nv_Fdz9/10 (XP_001630630), Nv_Smoothened (XP_001632182), Nv_sFRP1/2/5 (XP_001638620), Nv_sFRP3/4 (XP_001638660).

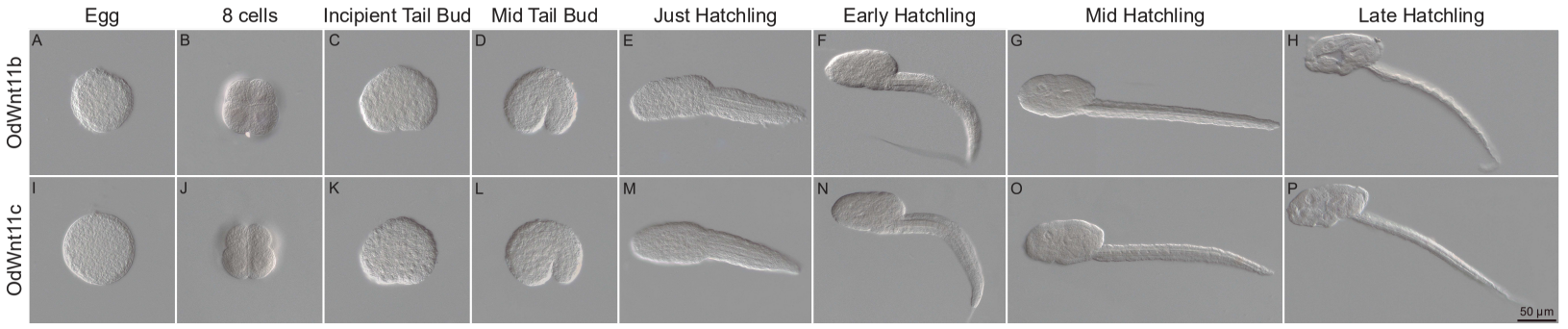


Fig. S3. Expression patterns of Oikopleura dioica *Wnt11b* and *Wnt11c* genes during embryonic and larval development. Whole-mount in situ hybridization in O. dioica eggs (A and I), 8 cells embryos (B and J), incipient tail bud embryos (C and K), mid tail bud embryos (D and L), just hatchlings (E and M), early hatchlings (F and N), mid hatchlings (G and O) and late hatchlings (H and P). Each panel corresponds to left lateral view of the animal oriented anterior toward the left and dorsal toward the top. Scale bar = 50 µm.

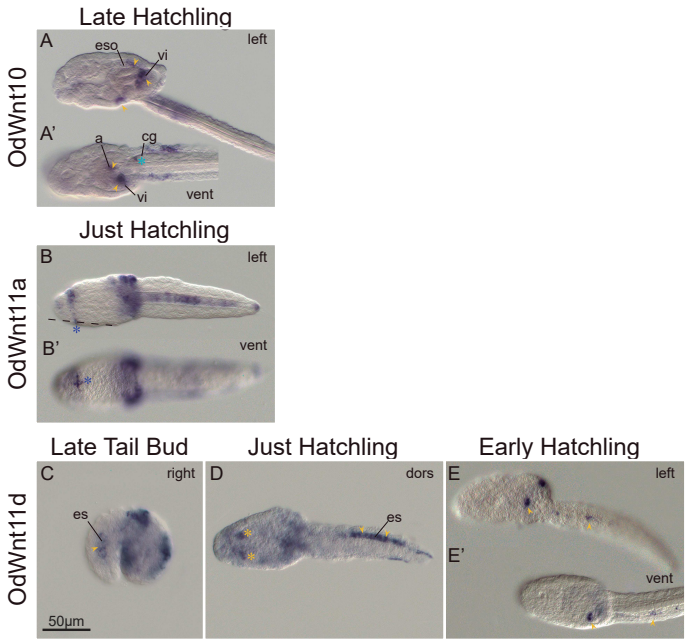


Fig. S4 Expression details of *Oikopleura dioica Wnt* genes during embryonic and larval development. Whole-mount in situ hybridization in O. dioica *Wnt10*, *Wnt11a* and *Wnt11d* genes. Embryos were viewed from various directions indicated at the top. Yellow arrowheads point endodermal strand; yellow asterisks points endostyle; blue light asterisks point caudal ganglion; blue dark asterisks point oikoplastic epithelium; a, anus, cg, caudal ganglion; eso, esophagus; es, endodermal strand, vi, vertical intestine. Scale bar = 50 µm.

OdWnt11a dsDNA #1

92% Identity

```
Spain               1  ATGAAGATTTCAGTAACCCTTTTCTCTGGATTACTCTCTGCGGCATTCGGTATCGACTGG  60
Japan (comp25579)   1  ..............C..A.......................A..C.....C.........  60

Spain              61  ATCGGCATGCACGGAAAGATGGTCGAAGACGATCTTTGCGATGGATTAAGCGA-CACGCT  119
Japan (comp25579)  61  .....A.............................C.....C...........T..-...  119

Spain             120  GCAGTATCGTCTGTGTAGTAAATTTTCAAAAAATCGGAAAACGAAAGGATTTATGGAAGC  179
Japan (comp25579) 120  ...............C...........G...G....A.......................  179

Spain             180  TATCCACACGGCCACAATCCAAACAAC  206
Japan (comp25579) 180  C...........................  206
```

OdWnt11a dsDNA #2

91% Identity

```
Spain               1  GACTGTCACTCATATATGCAAATAACTAAAACTG--TATTAATTGTACATAAACAAAACC  58
Japan (comp25579)   1  ...A..............................TT..............TC........  60

Spain              59  ATTATCTAAAAGCTTGGAAAATGAGAGAGCCTTTTTGCTATTCTCGGATTATTCTCGAGT  118
Japan (comp25579)  61  ................A.............A.AAAA.....................A...  120

Spain             119  TCAAACTGTATATTAACTCGCCCAGAGCGCAATTAACACATTGTACATGTTCGCTAGTAT  178
Japan (comp25579) 121  ............A........T..T....C...C....TC...................  180

Spain             179  AACTTAACCACTTTATCATTGAACATTCTGCCCATATTTCCTACAGAAACTGTAAAAATA  238
Japan (comp25579) 181  ............G.............T..............C..A....----------  230

Spain             239  TTTTAAACGTGAAATGAAAATTATCTTCAAAATATTTCATATTATGTATTCGATGTAAAA  298
Japan (comp25579)      ------------------------------------------------------------

Spain             299  ATTGCATAAAAAGTACATTTTATCAGACCCAAATGGGCTCGAAAAGTTTCTTGTATGACT  359
Japan (comp25579)      ------------------------------------------------------------

Spain             360  CCT  363
Japan (comp25579)      ---
```

Fig. S6. Expression levels of *Wnt* signaling components in *Oikopleura dioica* across developmental and life time. Gene expression values for *O. dioica Wnt* signaling components resulting from the KAAS analysis were extracted from the gene expression matrix of OikoBase (28). Oocyte and two to eight cell embryos are the two stages that encompass the maternal transcripts inherit by the embryos, while the following stages represents the transcriptional activity of the embryos; 1 hour post fertilization (hpf) and tailbud, larvae; hatched, early tadpole and tailshift, or adults; day 1, day 2, day 3, day 4, day 5, trunk, testis and ovary

Maternal transcripts

Pathway	Component	Oikopleura	Oocyte	2-8 cells	1HPF	Tailbud	Hatched	Early tadpole	Tailshift	Day 1	Day 2	Day 3	Day 4	Day 5	Trunk	Testis	Ovary
Wnt activators	Wnt	GSOIDG00003722001 (*Wnt5*)	0	0	0	0	0	1206	1522	626	505	418	0	0	0	0	0
		GSOIDG00011688001 (*Wnt11a*)	1262	0	669	1451	1223	0	0	0	0	0	0	0	0	0	0
		GSOIDG00008816001 (*Wnt11b*)	0	0	791	289	0	0	907	178	446	503	0	0	0	0	743
		GSOIDG00013856001 (*Wnt11c*)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		GSOIDG00013757001 (*Wnt11d*)	0	0	0	564	0	0	0	0	0	0	0	0	0	0	0
		GSOIDG00004382001 (*Wnt10*)	0	0	0	0	0	0	1027	409	228	216	0	0	0	0	0
		GSOIDG00004676001 (*Wnt16*)	0	0	0	245	0	0	0	0	0	0	0	0	0	0	0
	Frizzled	GSOIDG00007729001 (Fzd1/2/7 like)	11929	10416	15464	8889	3703	2039	2038	1011	1743	2251	1539	1539	4152	0	7039
		GSOIDG00008718001 (Fzd3/6a)	4288	5532	2796	2974	2492	3484	697	0	0	0	0	0	0	0	0
		GSOIDG00000671001 (*Fzd3/6b*)	0	0	1743	10422	5071	6025	2428	580	618	527	0	0	0	0	0
		GSOIDG00006033001/34001 (Fzd3/6c)	0	0	0	0	0	0	532	588	0	0	0	0	0	0	0
		GSOIDG00000721001 (Fzd5/8)	0	0	0	2707	2313	1466	1523	415	302	250	0	0	0	0	0
Wnt/β-catenin	LRP5/6	GSOIDP00010381001	1849	1592	1463	1421	1107	2432	3398	877	1130	1474	776	645	0	1089	2404
	Dvl	GSOIDG00009214001	4415	3253	4876	2814	1413	1025	1678	1058	1095	1017	929	675	0	0	1109
	GSK3	GSOIDG00003070001	7161	6319	10995	5708	3368	3111	4154	2349	3331	3593	3465	8293	2647	18778	6597
	CKIα	GSOIDG00014955001	4990	4641	5973	4289	2463	2988	5602	4458	3935	4885	2684	4442	3482	8489	5950
	CK2	GSOIDG00002149001	7767	6258	8428	5201	3095	3174	4471	4765	5387	6374	5445	5971	3959	3262	7617
	β-catenin	GSOIDG00011813001	4002	4715	6376	4181	3820	3478	4532	3028	2248	2324	873	368	0	0	2235
		GSOIDG00004053001	5963	7108	8853	5922	5620	2763	1953	1025	1281	1502	1155	1392	0	0	3850
	TCF/LEF	GSOIDG00012371001	8510	8669	7360	12007	8549	6690	7585	3172	2603	3255	2363	2089	3548	0	4500
Ca2+ dependent	PLC	GSOIDG00016375001	3449	2994	3812	1317	651	1512	1436	895	787	1179	1302	953	0	0	2271
	CaMKII	GSOIDG00001768001	0	0	0	1260	1179	3214	2344	1404	1103	970	727	649	0	0	0
	Calcineurin	GSOIDG00003024001	0	0	0	486	859	1301	1029	319	247	267	0	0	0	0	0
	PKC	GSOIDG00004710001	0	0	0	0	0	0	329	160	0	0	0	0	0	0	0
Planar cell polarity	Daam1	GSOIDG00009580001	6511	5882	8374	3076	1243	3115	3901	1243	1567	2033	1169	1279	0	0	4389
	RhoA	GSOIDG00010718001	11722	11420	12994	8092	4883	11507	6393	6337	7762	7612	8206	7651	11072	2704	11681
	ROCK2	GSOIDG00000429001	8917	10096	14322	7968	3802	5348	10988	3987	3453	4277	3669	4462	2235	3042	10891
	Rac	GSOIDG00007376001	11056	6601	12103	15105	10447	6069	6034	4329	5665	5207	5858	4166	5426	0	8078
	JNK	GSOIDG00004808001	3677	2954	4518	1787	873	2190	2124	768	849	870	1004	971	0	697	2299
	Strabismus	GSOIDG00009906001	13552	12385	16132	7480	6987	6044	3124	1082	1423	1420	1028	910	0	0	5025
	Prickle	GSOIDG00017601001	2280	2039	4287	6077	3708	4289	3290	1987	1014	1058	686	529	0	0	1211
Antagonists	SFRP	GSOIDG00005455001	578	0	895	649	801	0	0	0	0	0	0	0	0	0	0

Fig. S6. Expression levels of *Wnt* signaling components in *Oikopleura dioica* across developmental and life time. Gene expression values for *O. dioica Wnt* signaling components resulting from the KAAS analysis were extracted from the gene expression matrix of OikoBase (28). Oocyte and two to eight cell embryos are the two stages that encompass the maternal transcripts inherit by the embryos, while the following stages represents the transcriptional activity of the embryos; 1-hour post fertilization (hpf) and tailbud, larvae; hatched, early tadpole and tailshift, or adults; day 1, day 2, day 3, day 4, day 5, trunk, testis and ovary

Table S1. Putative *O. dioica* components of the Wnt signaling pathway. E-values of bastp analysis of *O. dioica* proteins against non-redundant *H. sapiens* reference protein database.

Component	Oikopleura	*H. sapiens* proteins with the highest E-values
Wnt/¯-catenin		
Wnt	GSOIDG00003722001	Wnt5 (3e-63; NP_001243034)
	GSOIDG00004382001	Wnt2 (7e-19; NP_003382.1)
	GSOIDG00011688001	Wnt4 (2e-51; NP_110388)
	GSOIDG00008816001	Wnt4 (6e-35; NP_110388)
	GSOIDG00013757001	Wnt4 (3e-35; NP_110388)
	GSOIDG00013856001	Wnt4 (2e-34; NP_110388)
	GSOIDG00004676001	Wnt4 (6e-35; NP_110388)
Frizzled	GSOIDG00000671001	Fdz7 (1e-64; NP_001158088)
	GSOIDG00007729001	Fdz2 (8e-129 NP_001457)
	GSOIDG00000721001	Fzd5 (2e-153; NP_003459)
	GSOIDG00008718001	Fzd6 (4e-25; NP_003497)
	GSOIDG00006033001/34001	Fzd7 (1e-66; NP_003498)
LRP5/6	GSOIDP00010381001	LRP5 (1e-60; XP_011543332)
Dishevelled	GSOIDG00009214001	Dvl (3e-129; NP_004414)
GSK3	GSOIDG00003070001	GSK-3¯ (0.0; NP_001139628)
APC	Absent	
Axin	Absent	
CKIK	GSOIDG00014955001	CKIK(0.0; NP_001883)
CKIγ	Absent	
CK2	GSOIDG00002149001	CK2 (0.0; NP_001886)
¯-catenin	GSOIDG00011813001	CTNNB1 (0.0; NP_001317658)
	GSOIDG00004053001	CTNNB1 (1e-108; NP_001317658)
TCF/LEF	GSOIDG00012371001	TCF7L2 (2e-52; XP_011538413)
Planar cell polarity		
Daam1	GSOIDG00009580001	DAAM2 (0.0; NP_056160)
RhoA	GSOIDG00010718001	RhoA (3e-102; NP_001655)
ROCK2	GSOIDG00000429001	ROCK1 (1e-37; NP_005397)
Rac	GSOIDG00007376001	RAC1 (5e-114; NP_008839)
JNK	GSOIDG00004808001	JNK3 (0.0; XP_016863916)
Strabismus/van Gogh	GSOIDG00009906001	VANGL1 (2e-118; NP_001165882)
Prickle	GSOIDG00017601001	PRICKLE2 (4e-122; XP_011531742)
Ca2+ dependent		
Phospholipase C	GSOIDG00016375001	PLCB4 (1e-155; XP_011527556)
CaMKII	GSOIDG00001768001	CaMKII (0.0; XP_006714393)
Calcineurin	GSOIDG00003024001	CALN (0.0; NP_001124163)
Protein kinase C	GSOIDG00004710001	PKC (0.0; NP_997700)
Antagonists		
Dickkopf	Absent	
WIF	Absent	
Cerberus	Absent	
sFRP	GSOIDG00005455001	sFRP2 (1e-27; NP_003004)

Table S2. Primer sequences used in this study.

Primers for riboprobe synthesis		
	Forward Sequence (5'->3')	Reverse Sequence (5'->3')
Wnt5	ATGGCGTCAAAAAACGCTCTTCAAG	GCGAGGCTATTTGCATTTGTAGATTTT
Wnt10	ATGTCGAAGGAAACGAAAAATGAG	CTAGGTACAAACGCCGTATGTTAC
Wnt11a	ATGAAGATTTCAGTCACCCTTTTCTCTG	GTTATTTGCATATATGAGTGACAGTCG
Wnt11b	ATGAGAATCTCCAACATTCTCTTCGC	GATTATCGGCATGTGCTTGTGG
Wnt11c	ATGAAAATCTTTAACGTCCTCTTTGC	TCAACGGCAGGTGCTCGTCG
Wnt11d	ATGAAAATAACTTTCTTTTCTCTTTTCGC	GGTTTAACGGCAAGAATGTGAAAT
Wnt16	ATGGCAATGCAACAATGTGAAACTG	GCTTATATACAGTAGTACTTCGTCT
Primers for dsDNA injection		
Wnt11a dsDNA #1	ATGAAGATTTCAGTCACCCTTTTCTCTG	GTTGTTTGGATTGTGGCCGTGTG
Wnt11a dsDNA #2	GTATGTATGCGGCCGCGACTGTCACTCATATATGCAAATAA	GAGGAGTCATACAAGAAACTTTTCG

Discussió

Pèrdues gèniques afectant a RA i Wnt a O. dioica *com a cas estudi per entendre el seu impacte en l'EvoDevo*

El creixent nombre de genomes seqüenciats està posant de manifest que la pèrdua gènica és un fenomen recurrent que pot haver generat diversitat al llarg de la història evolutiva dels diferents grups d'organismes (Albalat and Cañestro, 2016; O'Malley et al., 2016). En aquest sentit, un dels grans reptes en el camp de la EvoDevo és entendre l'impacte de la pèrdua gènica en l'evolució dels mecanismes del desenvolupament animal considerant diferents escenaris evolutius. En escenaris de selecció positiva, la pèrdua hauria comportat canvis morfològics o fisiològics que haurien permès l'adaptació de les espècies a noves condicions ambientals. En escenaris de selecció neutre (o quasi neutre), la pèrdua hauria afectat a gens que haurien esdevingut prescindibles per l'existència de robustesa mutacional o perquè les condicions ambientals haurien canviat. Saber si una pèrdua gènica ha estat adaptativa o neutre és generalment molt complex ja que molts cops es difícil establir relacions directes de causalitat entre la pèrdua d'un gen i l'aparició d'una nova característica evolutivament avantatjosa. Malgrat aquesta dificultat, en aquesta tesi doctoral vàrem voler estudiar l'impacte de la pèrdua gènica en l'EvoDevo a partir d'analitzar l'evolució de les vies de senyalització de l'àcid retinoic (RA) i de Wnt a *O. dioica* com a sistema model de referència.

Les vies de senyalització del RA i de Wnt són vies essencials pel desenvolupament de tots els cordats (Escriva et al., 2002; Fujiwara, 2006; Imai et al., 2000; Olivera-Martinez and Storey, 2007; Onai et al., 2009). Les dues vies estan interrelacionades, ja que l'efecte antagònic dels gradients oposats d'RA i FGF al llarg del eix anteroposterior es modulen per Wnt (Diez del Corral et al., 2003; Irving and Mason, 2000; Mercader et al., 2000; Olivera-Martinez and Storey, 2007), i de forma global determinen la regionalització dels diferents compartiments d'estructures derivades de les tres capes germinals (i. e. endoderma, mesoderma i ectoderma) al llarg del eix anteroposterior (Stern et al., 2006). Així doncs, s'ha vist que en els cordats, RA i Wnt, ja sigui interactuant de forma directa o indirecta, són essencials pel desenvolupament de

la part posterior del cos, la formació de estructures neuronals o dels arcs faringis (Graham et al., 2014; Onai et al., 2009; Pasini et al., 2012), essent considerades fonamentals per l'establiment del pla corporal. Sorprenentment, però, els resultats d'aquest treball de Tesis Doctoral revelen nombroses pèrdues gèniques en ambdues vies a *O. dioica*, malgrat aquesta espècie hagi mantingut el pla corporal típic dels cordats.

En referència a l'RA, estudis en cefalocordats i en vertebrats han demostrat que la senyalització de l'RA participa en l'establiment del patró corporal en els embrions, actuant a través de la regulació de l'expressió temporal dels gens del clúster Hox per establir identitats de posició al llarg de l'eix anteroposterior (Marlétaz et al., 2006). Els resultats d'aquesta Tesis Doctoral revelen per primer cop que *O. dioica* ha patit un fenomen de coeliminació gènica que ha afectat la majoria dels gens que codifiquen per les proteïnes de la via clàssica de metabolisme i senyalització del RA, i que els que ha sobreviscut a la coelimanció no són capaços de sintetitzar RA per una via metabólica alternativa (Capítol 2:(Martí-Solans et al., 2016)). Així doncs, podem assegurar que el desballestament de la via del RA a *O. dioica* s'ha donat en un escenari d'absència de robustesa mutacional, i que les pèrdues gèniques que ha sofert han anat acompanyades de la pèrdua de la funció de senyalització del RA en aquest organisme. Aquest descobriment porta a preguntar-nos com és possible que aquesta pèrdua funcional no hagi tingut un impacte més gran en el pla corporal, típicament de cordats, que té *O. dioica*. La resposta a aquesta paradoxa podria venir del fet que *O. dioica*, juntament amb els ascidis, ha adoptat un desenvolupament embrionari determinat i ràpid, el qual molt probablement ha portat a una disminució en la necessitat de senyals extracel·lulars com l'RA per la regulació temporalment col·lineal dels seus gens, incloent els gens *Hox* (Ferrier and Holland, 2002). Aquest fet hauria comportat una relaxació de la necessitat de mantenir el clúster *Hox* intacte, permetent la seva desintegració, tal i com ha passat a *O. dioica* o a *C. intestinalis* (Ikuta et al., 2004; Seo et al., 2004). Es raonable doncs pensar que el canvi a un sistema de desenvolupament embrionari determinat i ràpid va permetre la desintegració del clúster *Hox* en urocordats, fent dispensable la senyalització de l'RA en l'establiment de l'eix anteroposterior en aquest subfílum, i per tant possibilitant un procés de pèrdua dels components de la via. En aquest procés, però, s'ha de considerar que l'RA podia tenir més d'una funció en l'urocordat ancestral, les quals poden haver-se mantingut de forma diferent en el llinatge dels ascidis i les apendiculàries. Així doncs, en ascidis, l'ús de la senyalització de l'RA per a la reproducció asexual (Hara et al., 1992; Kawamura et al., 1993), la regeneració de l'estomac (Kaneko et al., 2010), la

formació dels palps (Yagi and Makabe, 2002), i la regulació del procés de diferenciació de l'epidermis (Kanda et al., 2009; Kanda et al., 2013; Sasakura et al., 2012) hauria comportat la preservació de la via. Per contra, les apendiculàries com *O. dioica*, només es reprodueixen sexualment, aparentment no regeneren, no tenen palps (ja que el seu estil de vida és pelàgic i no necessiten adherir-se a un substrat per fer la metamorfosis com els ascidis), i la seva l'epidermis ha estat evolutivament transformada per donar lloc al oikoplast, un epiteli que a través de mecanismes de desenvolupament totalment diferents als dels ascidis, ha adquirit la capacitat de sintetitzar les "cases". Per tant, es possible teoritzar que durant l'evolució, la via de senyalització de l'RA hauria esdevingut totalment prescindible per a *O. dioica*, permetent el desballestament de la seva maquinària genètica, això és, la coeliminació dels gens *Rdh10, Rdh16, Bco1, Aldh1a, Cyp26* i *RAR* en aquesta espècie. És interessant també destacar que les nostres dades han mostrat que alguns gens clàssicament associats al metabolisme de l'RA, com els gens *Aldh8*, *RdhE2* i *Bco*, han sobreviscut a la coeliminació i alguns d'ells fins i tot s'han duplicat. Aquest fet indicaria que aquests gens tindrien altres funcions, be fruit d'innovacions evolutives de les apendiculàries, o bé funcions ancestrals compartides per altres cordats que fins ara ens havien passat desapercebudes per la seva pliotropia funcional. Així, l'enzim Aldh8 podria formar part del sistema del defensoma d'*O. dioica* que s'activa en resposta a agents tòxics externs, mentre que els enzims RdhE2 i Bco serien candidats a participar en el metabolisme dels carotenoids i retinoides que es troben en abundància en la dieta d'*O. dioica*. En resum, *O. dioica* hauria perdut tots els gens de la via de senyalització de l'RA excepte aquells que tenen altres funcions. Si aquesta pèrdua va ser adaptativa facilitant canvis biològics importants per les apendiculàries o, per contra, va ser neutre com a conseqüència de la relaxació en la necessitat de produir RA, és impossible de saber amb certesa.

En referencia a la senyalització de Wnt, els nostres resultats mostren que *O. dioica* ha perdut 9 de les 13 subfamílies de lligands Wnt presents als metazous, esdevenint un dels bilaterals amb menys subfamílies (Capítol 4). A diferència de la via de senyalització de l'RA on es va donar un fenomen de coeliminació de la majoria de components de la via, l'extensa pèrdua de subfamílies de lligands Wnt no ha portat a la coeliminació de la via de transducció del senyal (Capítol 4). A més, els nostres resultats indiquen que les quatre subfamílies Wnt supervivents a *O. dioica* s'expressen de forma específica en les diferents capes embrionàries i recapitulen quasi totes les expressions dels gens *Wnt* en els altres cordats. Semblaria doncs que les pèrdues de determinades subfamílies Wnt podrien haver estat afavorides per esdeveniments de reassignació de funcions

(*function shuffling*) en els que un paràleg d'una subfamília hauria adquirit el domini d'expressió d'un altre. Així doncs, la robustesa mutacional als sistemes, no només l'aporten les duplicacions gèniques o de la redundància funcional per processos d'evolució convergent, sinó que els nostres resultats mostren el *function shuffling* hauria augmentat la redundància en els dominis d'expressions dels diferents *Wnt* proporcionant la robustesa mutacional necessària per a permetre la pèrdua de gens sense implicar canvis funcionals dràstics en el fenotip de *O. dioica* (Gitelman, 2007). A més, els resultats d'aquesta Tesi Doctoral mostren que el *function shuffling* no és una particularitat dels Wnt d'*O. dioica,* sinó que s'ha donat repetides vegades al llarg de la evolució de les subfamílies en els metazous (Janssen et al., 2010), incloent diferents espècies de cordats (Somorjai et al., 2018, Capítol 4). Així per exemple, l'expressió de Wnt2 a amfiox i Wnt16 a pollastre han estat reassignades al notocordi (Fokina and Frolova, 2006; Somorjai et al., 2018); diferents gens *Wnt* han estat reassignats per establir l'eix corporal primari a diverses espècies: *Wnt8* a peix zebra (Lu et al., 2011), *Wnt11* a *Xenopus* i *O. dioica* (Tao et al., 2005 i Capítol 4), i *Wnt5* a ascidis (Imai et al., 2004; Sasakura et al., 1998). Així doncs, els exemples descrits en aquesta Tesi Doctoral de *function shuffling* (Capítols 3 i 4) indiquen que aquest seria un mecanisme evolutiu que pot facilitar la pèrdua de gens al incrementar la redundància funcional, proporcionant robustesa mutacional a una xarxa gènica. Per tant, conèixer els mecanismes genètics i les situacions evolutives que afavoreixin el *function shuffling* entre gens paràlegs ajudaria a entendre millor l'evolució de les famílies gèniques –guanys i pèrdues– en els organismes.

O. dioica *com a model per investigar l'impacte de la pèrdua gènica en l'EvoDevo*

Les nostres anàlisis dels processos que han seguit les pèrdues gèniques en les vies de l'RA (coeliminació gènica) i de Wnt (*function shuffling*) posen de manifest que *O. dioica* és un sistema model atractiu per investigar esdeveniments de pèrdues gèniques i el seu impacte evolutiu. Per establir *O. dioica* com a model animal, però, a més de les característiques biològiques que la fan atractiva per aquests tipus estudis –desenvolupament embrionari i cicle de vida extremadament ràpid, simplicitat i transparència corporal, genoma reduït i totalment seqüenciat, possibilitat de manipulació gènica per estudis funcionals– ha estat fonamental desenvolupar unes instal·lacions per cultivar-la en el

laboratori, així com protocols de manteniment assequibles (low-cost) per qualsevol grup d'investigació.

Considerem, doncs, que els esforços que es recullen al Capítol 1, no solament han estat fonamentals per poder realitzar tots els projectes de recerca realitzats en el resta de la tesis, sinó que són de gran valor per qualsevol laboratori que vulgui estudiar *O. dioica* com a model *knockout* evolutiu del gen que sigui del seu interès. En el nostre laboratori, per exemple, els resultats d'aquesta Tesis que han demostrat que *O. dioica* és un *knockout* evolutiu del RA han possibilitat nous projectes que s'estan desenvolupant actualment. Per exemple, de forma similar al nostre estudi de com ha evolucionat la família Wnt en absència de RA (Capítol 4), s'està estudiant l'evolució de la via de senyalització per Fgf. En aquest mateix sentit, en el nostre grup també s'està estudiant els canvis que s'hagin pogut donar en el procés de cardiogènesis d'*O. dioica*, ja que l'RA, Wnt i Fgf són vies de senyalització fonamentals per la formació del cor en la resta de cordats (Fujiwara, 2006; Lin et al., 2010; Marlétaz et al., 2006; Simões-Costa et al., 2005).

Finalment, un darrer projecte generat a partir del nostre descobriment que *O. dioica* és un *knockout* evolutiu per l'RA, projecte en el qual jo he participat durant la Tesis Doctoral (Apèndix 2), és l'estudi dels mecanismes genètics del desenvolupament embrionari alterats per aldehids poliinsaturats (PUAs) produïts per afloraments de diatomees. Malgrat que s'ha descrit que els PUAs poden alterar el desenvolupament embrionari de molts organismes marins (e.g. copèpodes, eriçons de mar, ascidis), els mecanismes moleculars d'acció dels PUAs encara estan per revelar (Caldwell, 2009). Una de les hipòtesis que s'han proposat és que degut a la gran similitud entre els PUAs i el retinaldehid, es podria establir una competició pel domini catalític de les Aldh1a afectant així la síntesi de RA, que en darrera instància afectaria el desenvolupament embrionari (Bchini et al., 2013). El fet que *O. dioica* sigui un *knockout* evolutiu per l'RA ens ha permès testar aquesta hipòtesi. Els nostres resultats mostren que els PUAs poden alterar de forma dràstica el desenvolupament embrionari de una forma independent a l'RA, en contra de la hipòtesi inicial (Apèndix 2). En aquest estudi, a més, hem tingut la oportunitat de demostrar que alguns dels components que han sobreviscut al desballestament de la via metabòlica de l'RA, com l'*Aldh8*, són components del "*defensoma*", un conjunt de gens que tenen la capacitat d'activar la seva expressió com a resposta contra substàncies tòxiques com els PUAs (Apèndix 2).

En resum, els resultats d'aquesta tesi doctoral posen de manifest que *O. dioica* és un model animal atractiu per estudiar tant aspectes bàsics de l'impacte

de les pèrdues gèniques en l'evolució dels mecanismes del desenvolupament, com en aspectes aplicats en que certes pèrdues confereixen a *O. dioica* la condició de *knockout* evolutiu que pot ser interessant per l'estudi de mecanismes moleculars concrets com la toxicitat de les PUAs en el desenvolupament embrionari d'organismes marins.

Conclusions

Conclusions

L'anàlisi global dels resultats obtinguts en aquesta Tesi Doctoral condueix a les següents conclusions:

1. Hem establert un sistema de cultiu d'*O. dioica* que és fiable, flexible i sobretot econòmicament viable per qualsevol laboratori que desitgi utilitzar *O. dioica* com a model animal.

2. El nostre sistema de cultiu és capaç de mantenir línies saludables amb característiques similars als animals salvatges. Els nostres resultats suggereixen que, malgrat eventualment l'endogàmia pot afectar la fertilitat de les poblacions del laboratori, l'entrada esporàdica de nova variabilitat genètica ajuda a superar aquesta limitació.

3. Els nostres resultats demostren que *O. dioica* és un *knockout* evolutiu per la via de senyalització de l'àcid retinoic, un cas singular entre els cordats.

4. Les pèrdues gèniques que han portat al desballestament de la via de senyalització de l'RA (és a dir, la pèrdua de Rdh10, Rdh16, Bco1, Aldh1a, Cyp26 i RAR) exemplifiquen un patró de coeliminació.

5. Les anàlisis dels gens que han sobreviscut al desballestament de la via metabòlica del RA (és a dir Aldh8, Cco i RdhE2) mostren que aquests no constitueixen una via de síntesi alternativa d'RA a *O. dioica*, com es podria pensar per la funció dels seus homòlegs en vertebrats.

6. Els nostre resultats suggereixen que el gens que han sobreviscut al desballestament de la via metabòlica del RA ho han fet bé perquè són pliotròpics i han mantingut funcions ancestrals (p. ex. la funció detoxificant de l'Aldh8 com a part del *defensoma*), o bé perquè han experimentat possibles processos de neofuncionalització (p. ex la funció de les Cco i RdhE2 en el metabolisme de caroteinods i d'altres retinoids).

7. El nostre treball ens permet concloure que les pèrdues gèniques que han afectat a la via de l'RA a *O. dioica* han tingut lloc en un sistema genètic no robust. No obstant, el desballestament de la via no han tingut un impacte dramàtic en el pla corporal típic de cordats que preserva aquest organisme.

8. Els nostres estudis de les subfamílies Wnt a cordats demostren que diferents subfila han seguit diferents escenaris evolutius: des de l'estasi genòmica dels

"conservadors" cefalocordats, fins als patrons evolutius dinàmics de pèrdua i duplicació de gens en els "liberals" urocordats, passant per un escenari intermedi de duplicacions i pèrdues gèniques als "moderats" vertebrats.

9. Les anàlisis comparatives entre els subfila de cordats ens han permès deduir funcions Wnt ancestrals per tots els cordats, casos de *function shuffling* entre paràlegs de Wnt, així com innovacions funcionals de cada llinatge.

10. L'estudi de la subfamília de WntA a cordats permet fer una hipòtesis sobre un possible rol d'aquest lligand en la formació de la boca a cefalocordats.

11. Seguint la tendència "liberal" dels urocordats, l'apendiculària *O. dioica* presenta el patró evolutiu encara més "radical" havent perdut 9 de les 13 subfamílies de Wnt.

12. De les 4 subfamílies de Wnt que han sobreviscut, cal destacar el procés d' expansió i subfuncionalització sofert per la subfamília Wnt11.

13. Malgrat les extenses pèrdues de subfamílies Wnt a *O. dioica*, la via de transducció de senyal continua sent funcional, conservant la majoria dels seus elements.

14. El nostres resultats mostren que la majoria d'estructures que expressen algun homòleg de Wnt en altres cordats, també expressen un Wnt en un domini homòleg a *O. dioica*.

15. Els patrons d'expressió dels Wnt a *O. dioica* poden explicar-se sota tres escenaris d'evolució diferents: conservació de la funció ancestral (p. ex. Wnt11 a l'endoderma); *function shuffling* entre paràlegs (p. ex. Wnt5 al tub nerviós); o pèrdua de la funció ancestral (p. ex absència d'expressió de Wnt a les fenedures branquials).

16. El nostre treball ens permet concloure que a més de la duplicació gènica o de la redundància funcional per processos d'evolució convergent, els fenòmens de *function shuffling* augmenten també la robustesa mutacional i faciliten les pèrdues gèniques en les famílies de gens.

Bibliografia

Acuña, J. L. (2001). Pelagic Tunicates: Why Gelatinous? *Am. Nat.* **158**, 100–107.

Albalat, R. (2009). The retinoic acid machinery in invertebrates: Ancestral elements and vertebrate innovations. *Mol. Cell. Endocrinol.* **313**, 23–35.

Albalat, R. and Cañestro, C. (2009). Identification of Aldh1a, Cyp26 and RAR orthologs in protostomes pushes back the retinoic acid genetic machinery in evolutionary time to the bilaterian ancestor. *Chem. Biol. Interact.* **178**, 188–196.

Albalat, R. and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391.

Albalat, R., Martí-Solans, J. and Cañestro, C. (2012). Dna methylation in amphioxus: From ancestral functions to new roles in vertebrates. *Brief. Funct. Genomics* **11**, 142–155.

Altucci, L. and Gronemeyer, H. (2001). The promise of retinoids to fight against cancer. *Nat. Rev. Cancer* **1**, 181–93.

Aravind, L., Watanabe, H., Lipman, D. J. and Koonin, E. V (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11319–24.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008.

Bassham, S. (2005). The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate Oikopleura dioica. *Development* **132**, 4259–4272.

Bassham, S. and Postlethwait, J. (2000). Brachyury (T) Expression in Embryos of a Larvacean Urochordate, Oikopleura dioica, and the Ancestral Role of T. *Dev. Biol.* **220**, 322–332.

Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. and Rahuel-Clermont, S. (2013). Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. *Chem. Biol. Interact.* **202**, 78–84.

Belyaeva, O. V and Kedishvili, N. Y. (2002). Human pancreas protein 2 (PAN2) has a retinal reductase activity and is ubiquitously expressed in human tissues. *FEBS Lett.* **531**, 489–93.

Belyaeva, O. V., Korkina, O. V., Stetsenko, A. V., Kim, T., Nelson, P. S. and Kedishvili, N. Y. (2005). Biochemical Properties of Purified Human Retinol Dehydrogenase 12 (RDH12): Catalytic Efficiency toward Retinoids and C 9 Aldehydes and Effects of Cellular Retinol-Binding Protein Type I (CRBPI) and Cellular Retinaldehyde-Binding Protein (CRALBP) on the Ox. *Biochemistry* **44**, 7035–7047.

Belyaeva, O. V., Korkina, O. V., Stetsenko, A. V. and Kedishvili, N. Y. (2008). Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. *FEBS J.* **275**, 138–147.

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

Bertrand, S., Aldea, D., Oulion, S., Subirana, L., de Lera, A. R., Somorjai, I. and Escriva, H. (2015). Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. *PLoS One* **10**, e0136587.

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S. and Van de Peer,

Y. (2006). The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43.

Bollner, T., Holmberg, K. and Olsson, R. (1986). A Rostral Sensory Mechanism in Oikopleura dioica (Appendicularia). *Acta Zool.* **67**, 235–241.

Bouquet, J. M., Spriet, E., Troedsson, C., Otter, H., Chourrout, D. and Thompson, E. M. (2009). Culture optimization for the emergent zooplanktonic model organism Oikopleura dioica. *J. Plankton Res.* **31**, 359–370.

Burighel, P. and Brena, C. (2001). Gut ultrastructure of the appendicularian Oikopleura dioica (Tunicata). *Invertebr. Biol.* **120**, 278–293.

Burke, M., Scholl, E. H., Bird, D. M., Schaff, J. E., Colman, S. D., Crowell, R., Diener, S., Gordon, O., Graham, S., Wang, X., et al. (2015). The plant parasite Pratylenchus coffeae carries a minimal nematode genome. *Nematology* **17**, 621–637.

Caldwell, G. S. (2009). The influence of bioactive oxylipins from marine diatoms on invertebrate reproduction and development. *Mar. Drugs* **7**, 367–400.

Cañestro, C. and Postlethwait, J. H. (2007). Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305**, 522–538.

Cañestro, C., Bassham, S. and Postlethwait, J. (2005). Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. *Dev. Biol.* **285**, 298–315.

Cañestro, C., Postlethwait, J. H., Gonzàlez-Duarte, R. and Albalat, R. (2006). Is retinoic acid genetic machinery a chordate innovation? *Evol. Dev.* **8**, 394–406.

Cañestro, C., Yokoi, H. and Postlethwait, J. H. (2007). Evolutionary developmental biology and genomics. *Nat. Rev. Genet.* **8**, 932–942.

Cañestro, C., Bassham, S. and Postlethwait, J. H. (2008). Evolution of the thyroid: Anterior-posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. *Dev. Dyn.* **237**, 1490–1499.

Carter, C. J., Rand, C., Mohammad, I., Lepp, A., Vesprini, N., Wiebe, O., Carlone, R. and Spencer, G. E. (2015). Expression of a retinoic acid receptor (RAR)-like protein in the embryonic and adult nervous system of a protostome species. *J. Exp. Zool. Part B Mol. Dev. Evol.* **324**, 51–67.

Cha, S.-W., Tadjuidje, E., Tao, Q., Wylie, C. and Heasman, J. (2008). Wnt5a and Wnt11 interact in a maternal Dkk1-regulated fashion to activate both canonical and non-canonical signaling in Xenopus axis formation. *Development* **135**, 3719–29.

Chang, E. S., Neuhof, M., Rubinstein, N. D., Diamant, A., Philippe, H., Huchon, D. and Cartwright, P. (2015). Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proc. Natl. Acad. Sci.* **112**, 14912–14917.

Chavali, S., Morais, D. A. de L., Gough, J. and Babu, M. M. (2011). Evolution of eukaryotic genome architecture: Insights from the study of a rapidly evolving metazoan, Oikopleura dioica: Non-adaptive forces such as elevated mutation rates may influence the evolution of genome architecture. *BioEssays* **33**, 592–601.

Chelstowska, S., Widjaja-Adhi, M. A. K., Silvaroli, J. A. and Golczak, M. (2016). Molecular Basis for Vitamin A Uptake and Storage in Vertebrates. *Nutrients* **8**,.

Chioda, M., Eskeland, R. and Thompson, E. M. (2002). Histone gene complement, variant expression, and mRNA processing in a urochordate Oikopleura dioica that undergoes extensive polyploidization. *Mol Biol Evol* **19**, 2247–2260.

Cima, F., Brena, C. and Burighel, P. (2002). Multifarious activities of gut epithelium in an appendicularian (Oikopleura dioica: Tunicata). *Mar. Biol.* **141**, 479–490.

Collins, M. D. and Mao, G. E. (1999). Teratology of Retinoids. *Annu. Rev. Pharmacol. Toxicol.* **39**, 399–430.

Croce, J. C. and Holstein, T. W. (2014). The Wnt's Tale. In *Wnt Signaling in Development and Disease* (ed. Hoppler, S.) and Moon, R. T.), pp. 161–176. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Croce, J. C. and McClay, D. R. (2009). Evolution of the Wnt Pathways. In *Wnt Signaling* (ed. Vincan, E.), pp. 3–18. Totowa, NJ: Humana Press.

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

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., et al. (2008). A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. *Mol. Syst. Biol.* **4**, 174.

De Luca, L. M. (1991). Retinoids and their receptors in differentiation, embryogenesis, and neoplasia. *FASEB J.* **5**, 2924–33.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., et al. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE. *Science* **273**, 1856–62.

Delsman, H. C. (1910). Beiträge zur Entwicklungsgeschichite von Oikopleura dioica. *Verch. Rijksinst. Odnderz. Zee* **3**, 1–24.

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

Deng, W. and Chourrout, D. (2017). Establishment of CRISPR based genome editing in Oikopleura for gene functional study.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.-M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L. C., Jubin, C., Cañestro, C., et al. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* **330**, 1381–5.

Deutscher, D., Meilijson, I., Kupiec, M. and Ruppin, E. (2006). Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* **38**, 993–8.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* **448**, 151–6.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M. and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79.

Dmetrichuk, J. M., Carlone, R. L. and Spencer, G. E. (2006). Retinoic acid induces neurite outgrowth and growth cone turning in invertebrate neurons. *Dev. Biol.* **294**, 39–49.

Duester, G. (2013). Retinoid signaling in control of progenitor cell differentiation during mouse development. *Semin. Cell Dev. Biol.* **24**, 694–700.

Edvardsen, R. B., Lerat, E., Maeland, A. D., Flåt, M., Tewari, R., Jensen, M. F.,

Lehrach, H., Reinhardt, R., Seo, H. C. and Chourrout, D. (2004). Hypervariable and highly divergent intron-exon organizations in the chordate Oikopleura dioica. *J. Mol. Evol.* **59**, 448–457.

Edvardsen, R. B., Seo, H. C., Jensen, M. F., Mialon, A., Mikhaleva, J., Bjordal, M., Cartry, J., Reinhardt, R., Weissenbach, J., Wincker, P., et al. (2005). Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr. Biol.* **15**, 12–13.

Engberg, N., Kahn, M., Petersen, D. R., Hansson, M. and Serup, P. (2010). Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent nodal signaling. *Stem Cells* **28**, 1498–509.

Escriva, H., Holland, N. D., Gronemeyer, H., Laudet, V. and Holland, L. Z. (2002). The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. *Development* **129**, 2905–16.

Essenberg, C. E. (1922). The Seasonal Distribution of the Appendicularia in the Region of San Diego, California. *Ecology* **3**, 55–64.

Estephane, D. and Anctil, M. (2010). Retinoic acid and nitric oxide promote cell proliferation and differentially induce neuronal differentiation in vitro in the cnidarian Renilla koellikeri. *Dev. Neurobiol.* **70**, 842–852.

Fenaux, R. (1976). Cycle vital d'un appendiculaire Oikopleura dioica Fol, 1872 description et chronologie. *Ann. l'Institut Oceanogr.* **52**, 89–101.

Fenaux, R. (1986). The house of Oikopleura dioica (Tunicata, Appendicularia): Structure and functions. *Zoomorphology* **106**, 224–231.

Fenaux, R. (1998a). Life history of the Appendicularia. In *The Biology of Pelagic Tunicates* (ed. Bone, Q.), pp. 151–159. Oxford: Oxford University Press.

Fenaux, R. (1998b). Anatomy and functional morphology of the Appendicularia. In *The Biology of Pelagic Tunicates* (ed. Bone, Q.), pp. 25–34. Oxford: Oxford University Press.

Fenaux, R. and Gorsky, G. (1979). Techniques d'élevage des Appendiculaires. *Ann Inst Ocean.* **55**, 195–200.

Fenaux, R. and Gorsky, G. (1985). Nouvelle Technique d'élevage des appendicularies. *Rapp. Procès-Verbaux des Réunions - Comm. Int. pour l'Exploration Sci. la Mer Méditerranée* **29**, 291–292.

Fenaux, R., Bone, Q. and Deibel, D. (1998). Appendicularian distribution and zoogeography. In *The Biology of Pelagic Tunicates* (ed. Bone, Q.), pp. 251–264. Oxford: Oxford University Press.

Ferrier, D. E. . and Holland, P. W. . (2002). Ciona intestinalis ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol. Phylogenet. Evol.* **24**, 412–417.

Flood, R. and Deibel, D. (1998). The appendicularian house. In *The Biology of Pelagic Tunicates* (ed. Bone, Q.), pp. 105–124. Oxford: Oxford University Press.

Fokina, V. M. and Frolova, E. I. (2006). Expression patterns of Wnt genes during development of an anterior part of the chicken eye. *Dev. Dyn.* **235**, 496–505.

Fu, X., Adamski, M. and Thompson, E. M. (2008). Altered miRNA repertoire in the simplified chordate, Oikopleura dioica. *Mol. Biol. Evol.* **25**, 1067–1080.

Fujii, S., Nishio, T. and Nishida, H. (2008). Cleavage pattern, gastrulation, and neurulation in the appendicularian, Oikopleura dioica. *Dev. Genes Evol.* **218**, 69–79.

Fujiwara, S. (2006). Retinoids and nonvertebrate chordate development. *J. Neurobiol.* **66**, 645–652.

Galant, R. and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**, 910–913.

Galt, C. P. and Fenaux, R. (1990). Urochordata-Larvacea.pdf. In *Reproductive biology of invertebrates. Vol.4., Part B.* (ed. Adiyodi, K. G.) and Adiyodi, R. G.), pp. 471–500. Wiley.

Galvani, A. P. and Novembre, J. (2005). The evolutionary history of the CCR5-Delta32 HIV-resistance mutation. *Microbes Infect.* **7**, 302–9.

Ganot, P., Bouquet, J. M. and Thompson, E. M. (2006). Comparative organization of follicle, accessory cells and spawning anlagen in dynamic semelparous clutch manipulators, the urochordate Oikopleuridae. *Biol Cell* **98**, 389–401.

Ganot, P., Bouquet, J. M., Kallesøe, T. and Thompson, E. M. (2007). The Oikopleura coenocyst, a unique chordate germ cell permitting rapid, extensive modulation of oocyte production. *Dev. Biol.* **302**, 591–600.

Ganot, P., Moosmann-Schulmeister, A. and Thompson, E. M. (2008). Oocyte selection is concurrent with meiosis resumption in the coenocystic oogenesis of Oikopleura. *Dev. Biol.* **324**, 266–276.

Gitelman, I. (2007). Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. *Mol. Biol. Evol.* **24**, 1912–25.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**, 481–487.

Graham, A., Butts, T., Lumsden, A. and Kiecker, C. (2014). What can vertebrates tell us about segmentation? *Evodevo* **5**, 24.

Grandel, H., Lun, K., Rauch, G.-J., Rhinn, M., Piotrowski, T., Houart, C., Sordino, P., Küchler, A. M., Schulte-Merker, S., Geisler, R., et al. (2002). Retinoic acid signalling in the zebrafish embryo is necessary during pre-segmentation stages to pattern the anterior-posterior axis of the CNS and to induce a pectoral fin bud. *Development* **129**, 2851–65.

Gutierrez-Mazariegos, J., Schubert, M. and Laudet, V. (2014). Evolution of Retinoic Acid Receptors and Retinoic Acid Signaling.pp. 55–73.

Halder, G., Callaerts, P. and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**, 1788–92.

Hale, F. (1933). Pigs born without eyeballs. *J. Hered.* **24**, 105–106.

Handberg-Thorsager, M., Gutierrez-Mazariegos, J., Arold, S. T., Kumar Nadendla, E., Bertucci, P. Y., Germain, P., Tomançak, P., Pierzchalski, K., Jones, J. W., Albalat, R., et al. (2018). The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. *Sci. Adv.* **4**, eaao1261.

Hara, K., Fujiwara, S. and Kawamura, K. (1992). Retinoic Acid can Induce a Secondary Axis in Developing Buds of a Colonial Ascidian, Polyandrocarpa misakiensis. *Dev. Growth Differ.* **34**, 437–445.

Heisenberg, C. P., Tada, M., Rauch, G. J., Saúde, L., Concha, M. L., Geisler, R., Stemple, D. L., Smith, J. C. and Wilson, S. W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81.

Henriet, S., Sumic, S., Doufoundou-Guilengui, C., Jensen, M. F., Grandmougin, C., Fal, K., Thompson, E., Volff, J. N. and Chourrout, D. (2015). Embryonic

expression of endogenous retroviral RNAs in somatic tissues adjacent to the Oikopleura germline. *Nucleic Acids Res.* **43**, 3701–3711.

Hikasa, H. and Sokol, S. Y. (2013). Wnt signaling in vertebrate axis specification. *Cold Spring Harb. Perspect. Biol.* **5**, 1–20.

Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–5.

Holland, L. Z. (2016). Tunicates. *Curr. Biol.* **26**, R146–R152.

Holland, L. Z., Gorsky, G. and Fenaux, R. (1988). Fertilization in Oikopleura dioica (Tunicata, Appendicularia): Acrosome reaction, cortical reaction and sperm-egg fusion. *Zoomorphology* **108**, 229–243.

Holmberg, K. (1984). A transmission electron microscopic investigation of the sensory vesicle in the brain of Oikopleura dioica (Appendicularia). *Zoomorphology* **104**, 298–303.

Holstein, T. W. (2012). The Evolution of the Wnt Pathway. *Cold Spring Harb. Perspect. Biol.* **4**, a007922–a007922.

Ikuta, T., Yoshida, N., Satoh, N. and Saiga, H. (2004). Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15118–23.

Imai, K., Takada, N., Satoh, N. and Satou, Y. (2000). (beta)-catenin mediates the specification of endoderm cells in ascidian embryos. *Development* **127**, 3009–20.

Imai, K. S., Hino, K., Yagi, K., Satoh, N. and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047–58.

Irving, C. and Mason, I. (2000). Signalling by FGF8 from the isthmus patterns anterior hindbrain and establishes the anterior limit of Hox gene expression. *Development* **127**, 177–186.

Jacob, F. (1977). Evolution and tinkering. *Science (80-.).* **196**, 1161–1166.

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

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231–7.

Kanda, M., Wada, H. and Fujiwara, S. (2009). Epidermal expression of Hox1 is directly activated by retinoic acid in the Ciona intestinalis embryo. *Dev. Biol.* **335**, 454–63.

Kanda, M., Ikeda, T. and Fujiwara, S. (2013). Identification of a retinoic acid-responsive neural enhancer in the Ciona intestinalis Hox1 gene. *Dev. Growth Differ.* **55**, 260–9.

Kaneko, N., Katsuyama, Y., Kawamura, K. and Fujiwara, S. (2010). Regeneration of the gut requires retinoic acid in the budding ascidian Polyandrocarpa misakiensis. *Dev. Growth Differ.* **52**, 457–68.

Kawaguchi, R., Zhong, M., Kassai, M., Ter-Stepanian, M. and Sun, H. (2015).

Vitamin A Transport Mechanism of the Multitransmembrane Cell-Surface Receptor STRA6. *Membranes (Basel).* **5**, 425–453.

Kawamura, K., Hara, K. and Fujiwara, S. (1993). Developmental role of endogenous retinoids in the determination of morphallactic field in budding tunicates. *Development* **117**, 835–845.

Kedishvili, N. Y., Chumakova, O. V., Chetyrkin, S. V., Belyaeva, O. V., Lapshina, E. A., Lin, D. W., Matsumura, M. and Nelson, P. S. (2002). Evidence That the Human Gene for Prostate Short-chain Dehydrogenase/Reductase (PSDR1) Encodes a Novel Retinal Reductase (RalR1). *J. Biol. Chem.* **277**, 28909–28915.

Kestler, H. A. and Kuhl, M. (2008). From individual Wnt pathways towards a Wnt signalling network. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 1333–1347.

Kilian, B., Mansukoski, H., Barbosa, F. C., Ulrich, F., Tada, M. and Heisenberg, C. P. (2003). The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. *Mech. Dev.* **120**, 467–476.

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.* **28**, 617–623.

King, M. and Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. *Science (80-.).* **188**, 107–116.

Kishi, K., Hayashi, M., Onuma, T. A. and Nishida, H. (2017). Patterning and morphogenesis of the intricate but stereotyped oikoplastic epidermis of the appendicularian, Oikopleura dioica. *Dev. Biol.* **428**, 245–257.

Kohler, R. E. (1994). *Lords of the Fly: Drosophila Genetics and the Experimental Life*. chicago: University of Chicago Press.

Körner, W. F. (1952). Untersuchungen über die gehäusebildung bei appendicularien (Oikopleura dioicafol). *Zeitschrift für Morphol. und Ökologie der Tiere* **41**, 1–53.

Korona, R. (2011). Gene dispensability. *Curr. Opin. Biotechnol.* **22**, 547–51.

Kugler, J. E., Kerner, P., Bouquet, J.-M., Jiang, D. and Di Gregorio, A. (2011). Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian Ciona and the larvacean Oikopleura. *BMC Evol. Biol.* **11**, 21.

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

Lei, Z., Chen, W., Zhang, M. and Napoli, J. L. (2003). Reduction of all-trans-retinal in the mouse liver peroxisome fraction by the short-chain dehydrogenase/reductase RRD: induction by the PPAR alpha ligand clofibrate. *Biochemistry* **42**, 4190–6.

Lemaire, P. and Piette, J. (2015). Tunicates: exploring the sea shores and roaming the open ocean. A tribute to Thomas Huxley. *Open Biol.* **5**, 150053.

Lin, S.-C., Dollé, P., Ryckebüsch, L., Noseda, M., Zaffran, S., Schneider, M. D. and Niederreither, K. (2010). Endogenous retinoic acid regulates cardiac progenitor differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9234–9.

Liu, L., Suzuki, K., Nakagata, N., Mihara, K., Matsumaru, D., Ogino, Y., Yashiro, K., Hamada, H., Liu, Z., Evans, S. M., et al. (2012). Retinoic acid signaling regulates sonic hedgehog and bone morphogenetic protein signalings during genital tubercle development. *Birth Defects Res. B. Dev. Reprod. Toxicol.* **95**, 79–

88.

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

Lombard, F., Sciandra, A. and Gorsky, G. (2009). Appendicularian ecophysiology. II. Modeling nutrition, metabolism, growth and reproduction of the appendicularian Oikopleura dioica. *J. Mar. Syst.* **78**, 617–629.

López-Urrutia, Á. and Acuña, J. L. (1999). Gut throughput dynamics in the appendicularian Oikopleura dioica. *Mar. Ecol. Prog. Ser.* **191**, 195–205.

Lu, F.-I., Thisse, C. and Thisse, B. (2011). Identification and mechanism of regulation of the zebrafish dorsal determinant. *Proc. Natl. Acad. Sci.* **108**, 15876–15880.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat. Rev. Neurosci.* **8**, 755–65.

Maden, M. and Hind, M. (2003). Retinoic acid, a regeneration-inducing molecule. *Dev. Dyn.* **226**, 237–244.

Makino, T. and McLysaght, A. (2012). Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. *Genome Res.* **22**, 2427–35.

Makita, R., Mizuno, T., Koshida, S., Kuroiwa, A. and Takeda, H. (1998). Zebrafish wnt11: Pattern and regulation of the expression by the yolk cell and No tail activity. *Mech. Dev.* **71**, 165–176.

Marlétaz, F., Holland, L. Z., Laudet, V. and Schubert, M. (2006). Retinoic acid signaling and the evolution of chordates. *Int. J. Biol. Sci.* **2**, 38–47.

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

Martinucci, G., Brena, C., Cima, F. and Burighel, P. (2005). Synchronous spermatogenesis in appendicularians. In *Response of Marine Ecosystems to Global Change: Ecological Impact of Appendicularians* (ed. Gorsky, G.), Youngbluth, M. J.), and Deibel, D.), pp. 113–123. Paris: Éditions des Archives Contemporaines.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. and Gehring, W. J. (1984). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–8.

McGregor, A. P., Pechmann, M., Schwager, E. E., Feitosa, N. M., Kruck, S., Aranda, M. and Damen, W. G. M. (2008). Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider. *Curr. Biol.* **18**, 1619–23.

Melhus, H., Michaëlsson, K., Kindmark, A., Bergström, R., Holmberg, L., Mallmin, H., Wolk, A. and Ljunghall, S. (1998). Excessive dietary intake of vitamin A is associated with reduced bone mineral density and increased risk for hip fracture. *Ann. Intern. Med.* **129**, 770–8.

Mercader, N., Leonardo, E., Piedra, M. E., Martínez-A, C., Ros, M. A. and Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961–70.

Metallo, C. M., Ji, L., de Pablo, J. J. and Palecek, S. P. (2008). Retinoic acid and bone morphogenetic protein signaling synergize to efficiently direct epithelial

differentiation of human embryonic stem cells. *Stem Cells* **26**, 372–80.
Mikels, A. J. and Nusse, R. (2006). Purified Wnt5a Protein Activates or Inhibits β-Catenin–TCF Signaling Depending on Receptor Context. *PLoS Biol.* **4**, e115.
Mikhailov, K. V., Slyusarev, G. S., Nikitin, M. A., Logacheva, M. D., Penin, A. A., Aleoshin, V. V. and Panchin, Y. V. (2017). The Genome of Intoshia linei Affirms Orthonectids as Highly Simplified Spiralians. *Curr. Biol.* **26**, 1768–1774.
Mikhaleva, Y., Kreneisz, O., Olsen, L. C., Glover, J. C. and Chourrout, D. (2015). Modification of the larval swimming behavior in Oikopleura dioica, a chordate with a miniaturized central nervous system by dsRNA injection into fertilized eggs. *J. Exp. Zool. Part B Mol. Dev. Evol.* **324**, 114–127.
Musso, G., Costanzo, M., Huangfu, M., Smith, A. M., Paw, J., San Luis, B.-J., Boone, C., Giaever, G., Nislow, C., Emili, A., et al. (2008). The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res.* **18**, 1092–9.
Nagatomo, K. I. and Fujiwara, S. (2003). Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated Ciona intestinalis embryos. *Gene Expr. Patterns* **3**, 273–277.
Navratilova, P., Danks, G. B., Long, A., Butcher, S., Manak, J. R. and Thompson, E. M. (2017). Sex-specific chromatin landscapes in an ultra-compact chordate genome. *Epigenetics Chromatin* **10**, 3.
Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121**, 526–41.
Nishida, H. (2008). Development of the appendicularian Oikopleura dioica: Culture, genome, and cell lineages. *Dev. Growth Differ.* **50**, 239–256.
Nishida, H. and Stach, T. (2014). Cell Lineages and Fate Maps in Tunicates: Conservation and Modification. *Zoolog. Sci.* **31**, 645–652.
Nishino, A. and Morisawa, M. (1998). Rapid Oocyte Growth and Artificial Fertilization of the Larvaceans Oikopleura dioica and Oikopleura longicauda. *Zoolog. Sci.* **15**, 723–727.
Nishino, A. and Satoh, N. (2001). The simple tail of chordates: Phylogenetic significance of Appendicularians. *Genesis* **29**, 36–45.
Nishino, A., Satou, Y., Morisawa, M. and Satoh, N. (2000). Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: Phylogenetic relationships among muscle tissues in the urochordates. *J. Exp. Zool.* **288**, 135–150.
O'Malley, M. A., Wideman, J. G. and Ruiz-Trillo, I. (2016). Losing Complexity: The Role of Simplification in Macroevolution. *Trends Ecol. Evol.* **31**, 608–621.
Ohno, S. (1970). *Evolution by gene duplication.*
Olivera-Martinez, I. and Storey, K. G. (2007). Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. *Development* **134**, 2125–35.
Olsson, R. (1963). Endostyles and endostylar secretions: A comparative histochemical study. *Acta Zool.* **44**, 299–328.
Olsson, R. (1965). Comparative morphology and physiology of the Oikopleura notochord. *Isr. J. Zool.* **14**, 213–220.
Olsson, R., Holmberg, K. and Lilliemarck, Y. (1990). Fine structure of the brain and brain nerves ofOikopleura dioica (Urochordata, Appendicularia). *Zoomorphology* **110**, 1–7.

Omotezako, T., Nishino, A., Onuma, T. A. and Nishida, H. (2013). RNA interference in the appendicularian Oikopleura dioica reveals the function of the Brachyury gene. *Dev. Genes Evol.* **223**, 261–267.

Omotezako, T., Onuma, T. A. and Nishida, H. (2015). DNA interference: DNA-induced gene silencing in the appendicularian Oikopleura dioica. *Proc. R. Soc. B* **282**, 20150435.

Omotezako, T., Matsuo, M., Onuma, T. A. and Nishida, H. (2017). DNA interference-mediated screening of maternal factors in the chordate Oikopleura dioica. *Nat. Publ. Gr.* 1–10.

Onai, T., Lin, H. C., Schubert, M., Koop, D., Osborne, P. W., Alvarez, S., Alvarez, R., Holland, N. D. and Holland, L. Z. (2009). Retinoic acid and Wnt/β-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus. *Dev. Biol.* **332**, 223–233.

Onuma, T. A., Isobe, M. and Nishida, H. (2017a). Internal and external morphology of adults of the appendicularian, Oikopleura dioica: an SEM study. *Cell Tissue Res.* **367**, 213–227.

Onuma, T. A., Matsuo, M. and Nishida, H. (2017b). Modified whole-mount in situ hybridisation and immunohistochemistry protocols without removal of the vitelline membrane in the appendicularian Oikopleura dioica. *Dev. Genes Evol.* **227**, 367–374.

Ordoñez, G. R., Hillier, L. W., Warren, W. C., Grützner, F., López-Otín, C. and Puente, X. S. (2008). Loss of genes implicated in gastric function during platypus evolution. *Genome Biol.* **9**, R81.

Paffenhöfer, G.-A. (1973). The cultivation of an appendicularian through numerous generations. *Mar. Biol.* **22**, 183–185.

Pang, K., Ryan, J. F., Comparative Sequencing Program, N., Mullikin, J. C., Baxevanis, A. D. and Martindale, M. Q. (2010). Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore Mnemiopsis leidyi. *Evodevo* **1**, 10.

Papp, B., Notebaart, R. A. and Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* **12**, 591–602.

Pasini, A., Manenti, R., Rothbächer, U. and Lemaire, P. (2012). Antagonizing Retinoic Acid and FGF/MAPK Pathways Control Posterior Body Patterning in the Invertebrate Chordate Ciona intestinalis. *PLoS One* **7**,.

Petersen, C. P. and Reddien, P. W. (2008). Smed-betacatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science* **319**, 327–30.

Petersen, C. P. and Reddien, P. W. (2009). Wnt Signaling and the Polarity of the Primary Body Axis. *Cell* **139**, 1056–1068.

Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R. and Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38**, 107–11.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., et al. (2007). Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science (80-.).* **317**, 86–94.

Raduan, M. A., Blanco, C., Soler, E., del Rio, J. G. and Raga, J. A. (1985). The seasonal distribution of the Appendicularia in the Bay of Cullera, Spain. *Rapp. Procès-Verbaux des Réunions - Comm. Int. pour l'Exploration Sci. la Mer*

Méditerranée **29**, 293–294.

Rhinn, M. and Dolle, P. (2012). Retinoic acid signalling during development. *Development* **139**, 843–858.

Riddiford, N. and Olson, P. D. (2011). Wnt gene loss in flatworms. *Dev. Genes Evol.* **221**, 187–97.

Rocheleau, C. E., Downs, W. D., Lin, R., Wittmann, C., Bei, Y., Cha, Y. H., Ali, M., Priess, J. R. and Mello, C. C. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. *Cell* **90**, 707–16.

Ronshaugen, M., McGinnis, N. and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* **415**, 914–917.

Sagane, Y., Zech, K., Bouquet, J. M., Schmid, M., Bal, U. and Thompson, E. M. (2010). Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans. *Development* **137**, 1483–1492.

Salensky, W. (1903). Etudes anatomiques sur les Appendicularies. I. Oikopleura vanhoeffeni Lohmann. *Mémoires l'Académie Impériale des Sci. St.-pétersbg.* **13**, 1–44.

Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–5.

Sasakura, Y., Ogasawara, M. and Makabe, K. W. (1998). HrWnt-5: A maternally expressed ascidian Wnt gene with posterior localization in early embryos. *Int. J. Dev. Biol.* **42**, 573–579.

Sasakura, Y., Kanda, M., Ikeda, T., Horie, T., Kawai, N., Ogura, Y., Yoshida, R., Hozumi, A., Satoh, N. and Fujiwara, S. (2012). Retinoic acid-driven Hox1 is required in the epidermis for forming the otic/atrial placodes during ascidian metamorphosis. *Development* **139**, 2156–60.

Sato, R., Yu, J., Tanaka, Y. and Ishimaru, T. (1999). New apparatuses for cultivation of appendicularians. *Plankt. Biol. Ecol.* **46**, 162–164.

Sato, R., Tanaka, Y. and Ishimaru, T. (2001). House Production by Oikopleura dioica (Tunicata, Appendicularia) Under Laboratory Conditions. *J. Plankton Res.* **23**, 415–423.

Satoh, N., Tagawa, K. and Takahashi, H. (2012). How was the notochord born? *Evol. Dev.* **14**, 56–75.

Saxena, S. K. (2009). Controversial role of smallpox on historical positive selection at the CCR5 chemokine gene (CCR5-Delta 32). *J. Infect. Dev. Ctries.* **3**, 324–6.

Seo, H.-C. (2001). Miniature Genome in the Marine Chordate Oikopleura dioica. *Science (80-.).* **294**, 2506–2506.

Seo, H.-C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., et al. (2004). Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **431**, 67–71.

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D. and Kingsley, D. M. (2006). Corrigendum: Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **439**, 1014–1014.

Simões-costa, M. S., Azambuja, A. P. and Xavier-Neto, J. (2008). The search for non-chordate retinoic acid signaling: lessons from chordates. *J. Exp. Zool. Part B Mol. Dev. Evol.* **310B**, 54–72.

Simões-Costa, M. S., Vasconcelos, M., Sampaio, A. C., Cravo, R. M., Linhares, V. L., Hochgreb, T., Yan, C. Y. I., Davidson, B. and Xavier-Neto, J. (2005). The evolutionary origin of cardiac chambers. *Dev. Biol.* **277**, 1–15.

Sire, J.-Y., Delgado, S. C. and Girondot, M. (2008). Hen's teeth with enamel cap: from dream to impossibility. *BMC Evol. Biol.* **8**, 246.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., et al. (2006). The Genome of the Sea Urchin Strongylocentrotus purpuratus. *Science (80-.).* **314**, 941–952.

Somorjai, I. M. L., Martí-Solans, J., Diaz-Gracia, M., Nishida, H., Imai, K. S., Escrivà, H., Cañestro, C. and Albalat, R. (2018). Wnt evolution and function shuffling in chordate genomes: from conservative amphioxus stasis to liberal ascidian gains and losses. *Genome Biol.*

Sönnichsen, B., Koski, L. B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.-M., Artelt, J., Bettencourt, P., Cassin, E., et al. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434**, 462–9.

Søviknes, A. M. and Glover, J. C. (2007). Spatiotemporal patterns of neurogenesis in the appendicularian Oikopleura dioica. *Dev. Biol.* **311**, 264–275.

Søviknes, A. M. and Glover, J. C. (2008). Continued growth and cell proliferation into adulthood in the notochord of the appendicularian Oikopleura dioica. *Biol. Bull.* **214**, 17–28.

Søviknes, A. M., Chourrout, D. and Glover, J. C. (2005). Development of putative GABAergic neurons in the appendicularian urochordate Oikopleura dioica. *J. Comp. Neurol.* **490**, 12–28.

Spada, F., Steen, H., Troedsson, C., Kalles??e, T., Spriet, E., Mann, M. and Thompson, E. M. (2001). Molecular Patterning of the Oikoplastic Epithelium of the Larvacean Tunicate Oikopleura dioica. *J. Biol. Chem.* **276**, 20624–20632.

Stach, T. and Anselmi, C. (2015). High-precision morphology: bifocal 4D-microscopy enables the comparison of detailed cell lineages of two chordate species separated for more than 525 million years. *BMC Biol.* **13**, 113.

Stach, T., Winter, J., Bouquet, J.-M., Chourrout, D. and Schnabel, R. (2008). Embryology of a planktonic tunicate reveals traces of sessility. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7229–7234.

Stavridis, M. P., Collins, B. J. and Storey, K. G. (2010). Retinoic acid orchestrates fibroblast growth factor signalling to drive embryonic stem cell differentiation. *Development* **137**, 881–90.

Stern, C. D., Charité, J., Deschamps, J., Duboule, D., Durston, A. J., Kmita, M., Nicolas, J.-F., Palmeirim, I., Smith, J. C. and Wolpert, L. (2006). Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? *Int. J. Dev. Biol.* **50**, 3–15.

Subramaniam, G., Campsteijn, C. and Thompson, E. M. (2014). Lifespan extension in a semelparous chordate occurs via developmental growth arrest just prior to meiotic entry. *PLoS One* **9**,.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64–119.

Szathmary, E. (2001). MOLECULAR BIOLOGY AND EVOLUTION: Can Genes Explain Biological Complexity? *Science (80-.).* **292**, 1315–1316.

Tao, Q., Yokota, C., Puck, H., Kofron, M., Birsoy, B., Yan, D., Asashima, M., Wylie, C. C., Lin, X. and Heasman, J. (2005). Maternal Wnt11 activates the canonical Wnt signaling pathway required for axis formation in Xenopus embryos. *Cell* **120**, 857–871.

Thompson, E. M., Kallesøe, T. and Spada, F. (2001). Diverse genes expressed in distinct regions of the trunk epithelium define a monolayer cellular template for construction of the oikopleurid house. *Dev. Biol.* **238**, 260–73.

Troedsson, C., Bouquet, J., Aksnes, D. L. and Thompson, E. M. (2002). Resource allocation between somatic growth and reproductive output in the pelagic chordate. *Mar. Ecol. Prog. Ser.* **243**, 83–91.

Tsagkogeorga, G., Turon, X., Hopcroft, R. R., Tilak, M.-K., Feldstein, T., Shenkar, N., Loya, Y., Huchon, D., Douzery, E. J. and Delsuc, F. (2009). An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol. Biol.* **9**, 187.

Tsagkogeorga, G., Turon, X., Galtier, N., Douzery, E. J. P. and Delsuc, F. (2010). Accelerated evolutionary rate of housekeeping genes in tunicates. *J. Mol. Evol.* **71**, 153–167.

Ukolova, S. (2012). Characterization of the Wnt signalling system in the coral Acropora millepora.

Uye, S. I. and Ichino, S. (1995). Seasonal variations in abundance, size composition, biomass and production rate of *Oikopleura dioica* (Fol) (Tunicata: Appendicularia) in a temperate eutrophic inlet. *J. Exp. Mar. Bio. Ecol.* **189**, 1–11.

Volff, J. N., Lehrach, H., Reinhardt, R. and Chourrout, D. (2004). Retroelement dynamics and a novel type of chordate retrovirus-like element in the miniature genome of the tunicate Oikopleura dioica. *Mol. Biol. Evol.* **21**, 2022–2033.

Wang, X.-D., Russell, R. M., Liu, C., Stickel, F., Smith, D. E. and Krinsky, N. I. (1996). β-Oxidation in Rabbit Liver in Vitro and in the Perfused Ferret Liver Contributes to Retinoic Acid Biosynthesis from β-Apocarotenoic Acids. *J. Biol. Chem.* **271**, 26490–26498.

Wang, K., Dantec, C., Lemaire, P., Onuma, T. A. and Nishida, H. (2017). Genome-wide survey of miRNAs and their evolutionary history in the ascidian, Halocynthia roretzi. *BMC Genomics* **18**, 314.

Weill, M., Philips, A., Chourrout, D. and Fort, P. (2005). The caspase family in urochordates: distinct evolutionary fates in ascidians and larvaceans. *Biol. cell* **97**, 857–66.

Westfall, T. A., Brimeyer, R., Twedt, J., Gladon, J., Olberding, A., Furutani-Seiki, M. and Slusarski, D. C. (2003). Wnt-5/pipetail functions in vertebrate axis formation as a negative regulator of Wnt/beta-catenin activity. *J. Cell Biol.* **162**, 889–98.

Widjaja-Adhi, M. A. K., Lobo, G. P., Golczak, M. and Von Lintig, J. (2015). A genetic dissection of intestinal fat-soluble vitamin and carotenoid absorption. *Hum. Mol. Genet.* **24**, 3206–3219.

Wikramanayake, A. H., Huang, L. and Klein, W. H. (1998). beta-Catenin is essential for patterning the maternally specified animal-vegetal axis in the sea urchin embryo. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9343–8.

Wikramanayake, A. H., Peterson, R., Chen, J., Huang, L., Bince, J. M., McClay, D. R. and Klein, W. H. (2004). Nuclear beta-catenin-dependent Wnt8 signaling in vegetal cells of the early sea urchin embryo regulates gastrulation and differentiation of endoderm and mesodermal cell lineages. *Genesis* **39**, 194–205.

Wolbach, S. B. and Howe, P. R. (1925). Tissue changes following deprivation of fat-soluble A vitamin. *J. Exp. Med.* **42**, 753–77.

Wongsiriroj, N., Piantedosi, R., Palczewski, K., Goldberg, I. J., Johnston, T. P., Li, E. and Blaner, W. S. (2008). The Molecular Basis of Retinoid Absorption. *J. Biol. Chem.* **283**, 13510–13519.

Yadetie, F., Butcher, S., Førde, H. E., Campsteijn, C., Bouquet, J.-M., Karlsen, O. A., Denoeud, F., Metpally, R., Thompson, E. M., Manak, J. R., et al. (2012). Conservation and divergence of chemical defense system in the tunicate Oikopleura dioica revealed by genome wide response to two xenobiotics. *BMC Genomics* **13**, 55.

Yagi, K. and Makabe, K. W. (2002). Retinoic acid differently affects the formation of palps and surrounding neurons in the ascidian tadpole. *Dev. Genes Evol.* **212**, 288–92.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716.

Apèndix 1

DNA methylation in amphioxus: from ancestral functions to new role in vertebrates.

Ricard Albalat, Josep Martí-Solans, and Cristian Cañestro.

Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

Resum

En els vertebrats, la metilació de l'DNA és un mecanisme epigenètic que modula la transcripció de gens i juga un paper crucial durant el desenvolupament, el manteniment de la destinació cel·lular, la pluripotència de les cèl·lules germinals i la impremta hereditària del genoma La metilació de l'DNA també pot tenir un paper com a mecanisme de defensa del genoma contra l'activitat mutacional derivada de la mobilitat dels transposons. En contrast amb els genomes fortament metilats dels vertebrats, la majoria dels genomes dels invertebrats estan escassament o parcialment metilats, i la funció de la metilació de l'DNA no està clara. Aquí, es revisa el sistema de metilació de l'DNA del cefalocordat amfiox, que pertany al grup que va divergir més aviat de la resta del nostre fílum, els cordats. En primer lloc, les cerques en la base de dades del genoma de l'amfiox revelen la presència de la maquinària de metilació d'DNA, les metiltransferases de l'DNA i les proteïnes de domini de metil-CpG. En segon lloc, la genòmica comparativa i l'anàlisi de la sintènia conservada entre amfiox i vertebrats proporcionen una evidència sòlida de que la maquinària de metilació d'DNA que presenta l'amfiox podria representa maquinària del cordat ancestral, i que la seva expansió en els vertebrats que s'observa en els vertebrats es va originar per les dues rondes de duplicació del genoma que es van produir a durant l'evolució dels vertebrats. En tercer lloc, en l'anàlisi *in silico* de les proporcions de CpGo/e al llarg del genoma de l'amfios suggereixen una distribució bimodal de la metilació de l'DNA, consistent en un patró de mosaic que comprèn dominis de DNA metilat intercalat amb dominis d'DNA no metilat, similar a la situació descrita en ascidis, però radicalment diferent als genomes globalment metilats dels vertebrats. Finalment, es discuteixen els rols potencials del sistema de metilació de l'DNA en l'amfiox en el context de l'evolució del genoma cordat i l'origen dels vertebrats.

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL II. NO 2. 142–155 doi:10.1093/bfgp/els009

DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates

Ricard Albalat, Josep Martí-Solans and Cristian Cañestro

Advance Access publication date 2 March 2012

Abstract

In vertebrates, DNA methylation is an epigenetic mechanism that modulates gene transcription, and plays crucial roles during development, cell fate maintenance, germ cell pluripotency and inheritable genome imprinting. DNA methylation might also play a role as a genome defense mechanism against the mutational activity derived from transposon mobility. In contrast to the heavily methylated genomes in vertebrates, most genomes in invertebrates are poorly or just moderately methylated, and the function of DNA methylation remains unclear. Here, we review the DNA methylation system in the cephalochordate amphioxus, which belongs to the most basally divergent group of our own phylum, the chordates. First, surveys of the amphioxus genome database reveal the presence of the DNA methylation machinery, DNA methyltransferases and methyl-CpG-binding domain proteins. Second, comparative genomics and analyses of conserved synteny between amphioxus and vertebrates provide robust evidence that the DNA methylation machinery of amphioxus represents the ancestral toolkit of chordates, and that its expansion in vertebrates was originated by the two rounds of whole-genome duplication that occurred in stem vertebrates. Third, *in silico* analysis of CpGo/e ratios throughout the amphioxus genome suggests a bimodal distribution of DNA methylation, consistent with a mosaic pattern comprising domains of methylated DNA interspersed with domains of unmethylated DNA, similar to the situation described in ascidians, but radically different to the globally methylated vertebrate genomes. Finally, we discuss potential roles of the DNA methylation system in amphioxus in the context of chordate genome evolution and the origin of vertebrates.

Keywords: *DNA methylation; amphioxus; chordate and vertebrate genome evolution; transposable elements; conserved genomic synteny; Dnmt and Mbd*

INTRODUCTION

DNA methylation is an epigenetic mechanism that has been mainly associated with regulation of gene expression. DNA methylation marks can be inherited cell to cell, and play important roles maintaining differential gene expression profiles between different cell lineages during embryo development, states of cell differentiation in adult tissues or

Corresponding author. Cristian Cañestro, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Edifici Prevosti, Av. Diagonal, 643, E-08028 Barcelona, Spain. Tel: +34 93 403 5304; Fax: +34 93 403 4420; E-mail: canestro@ub.edu

Ricard Albalat, PhD (Universitat de Barcelona), Postdoc (Skirball Institute; New York University) is Associate Professor of Genetics at the Universitat de Barcelona. His current research is focused on the evolution of the retinoid and steroid signaling pathways during chordate evolution. He is also interested in the impact of transposable elements and the changes of epigenetics mechanisms in the evolution of chordate genomes.

Josep Martí-Solans is a Master student in the Developmental Biology and Genetics program at the Department of Genetics in the University of Barcelona. His research project focuses on understanding mechanisms of gene losses and gene gains in the retinoid signaling pathway of nonvertebrate chordates (e.g. amphioxus and *Oikopleura*) that might help to understand the evolution of stem chordates and the origin of functional innovations in stem vertebrates.

Cristian Cañestro is an Assistant Professor at the Department of Genetics, in the University of Barcelona, Spain, since 2009. He made his postdoctoral research in the Institute of Neuroscience at the University of Oregon (2002–09). He received his PhD from the Department of Genetics, University of Barcelona, Spain (2001). He is interested in Genomics, Functional Evo-Devo and Epigenetics, and his research focus on the impact of gene losses on developmental programs. His projects aim to understand the evolution of chordates and the origin of vertebrate innovations. Amphioxus, ascidians, larvaceans and zebrafish are the main model organisms in which he studies his favorite signaling molecule, the retinoic acid and its epigenetic role on embryo development.

pluripotency in germ cells, as well as preserving chromosomal imprinting [1–4]. DNA methylation consists in the addition of a methyl group covalently bound to the 5-carbon of cytosine of a CpG dinucleotide. In animals, DNA methylation is mediated by DNA methyltransferases (Dnmt), which are evolutionarily conserved enzymes that have been classified into three groups, Dnmt1, Dnmt2 and Dnmt3 [5]. Dnmt3 enzymes are responsible for *de novo* methylation during development [6, 7]. Dnmt1 enzymes are important in maintaining methylation patterns after DNA replication on hemimethylated symmetric motifs [8]. Dnmt2 (also termed Trdmt1) enzymes show a robust methyltransferase activity of tRNA molecules, and only present low activity against DNA [9, 10].

The addition of methyl groups to the DNA regulates gene expression in two ways, either by direct interference with the binding of proteins that interact with DNA elements [11–13], or indirectly, through recruitment of proteins that contain a methyl-CpG-binding domain (MBD), which are called Mbd1, Mbd2, Mbd3, Mbd3L, Mbd4 and Mecp2 (methyl CpG-binding protein 2) [14, 15]. Most Mbd proteins specifically recognize and bind to methylated DNA, and it is thought that they can associate with histone deacetylases (HDACs) and other chromatin remodeling proteins, contributing to the transcriptional repression of promoters ([16, 17] reviewed in [18]). Each Mbd protein has, however, its own features. MeCP2 was the first discovered Mbd protein with methyl-CpG DNA-binding activity, and it has been related to the HDAC complex in the maintenance of transcriptional silencing [17]. Mbd1 is the only methyl-CpG-binding protein that is capable of repressing transcription from both methylated and unmethylated promoters, at least in cell culture [19–21]. The DNA binding activity of Mbd2 and Mbd3 proteins is mediated by their MBD, and although such activity has been lost in the mammalian Mbd3, it is still preserved in amphibians and fishes [22]. In mammals, Mbd2 represses transcription in a methylation density-dependent fashion [23]. Mbd3L1 is a protein with significant homology to Mbd2 and Mbd3. It lacks, however, the MBD, although it might interact with Mbd2 as a methylation-dependent transcriptional repressor, likely competing with Mbd3 [24]. Finally, Mbd4 is the only member of the MBD family that does not seem to be involved in transcriptional repression, but it plays a role in decreasing the mutability of methyl-CpG in the genome [25]. Additional Mbd proteins have also been described (e.g. Mbd5 and Mbd6), and in contrast to the rest of the MBD proteins, they do not bind methylated DNA, but might play a role in the formation of heterochromatin and epigenetic reprogramming after fertilization [26].

Despite that the DNA methylation is present throughout the phylogenetic ladder, substantial differences in the multiplicity of *Dnmt* and *Mbd* genes, and in the amount and genomic distribution of methylated cytosines are found among different taxa. Thus, vertebrates have more *Dnmt* and *Mbd* genes, and higher methylated genomes than non-vertebrates. Vertebrates have at least one *Dnmt1*, one *Dnmt2*, one to five *Dnmt3* genes and at least six *Mbd* genes in their DNA methylation toolkit, whereas invertebrates have a variable lower number of *Dnmt* and *Mbd* genes ([5, 22, 27–32] and Figure 1). *Caenorhabditis elegans*, for instance, lacks *Dnmt* genes, whereas related nematodes preserve one *Dnmt2*-related gene [33]. Among insects, fruit flies and mosquitoes only have *Dnmt2*, the silk moth has both *Dmnt1* and *Dmnt2*, and the honeybee possesses the full set of *Dmnt* genes [34, 35]. The protostome *Mbd* complement also varies depending on the species analyzed. Most lophotrochozoan and ecdysozoan species have both *Mbd1/2/3* (formerly *Mbd2/3*) and *Mbd4/MeCP2* (formerly *Mbd4*) genes, although some ecdisozoans have a truncated or even no *Mbd4/MeCP2* genes ([28, 31, 33, 36] and Figure 1). Outside bilaterians, a full complement of *Dnmt* is present in cnidarians and sponges, but so far a sole *Dnmt2* has been found in placozoans [31, 37, 38]. *Mbd1/2/3* and *Mbd4/MeCP2* have been also identified in cnidarians as well as in placozoans [31]. In addition to the diversity of *Dnmt* and *Mbd* genes, the extent of CpG methylation and its genomic distribution also differ among non-vertebrate species, from undetectable or very low levels to mosaic patterns of substantial methylation in different non-vertebrate genomes, although never reaching the high levels of global methylation of vertebrates [39–43].

Amphioxus belongs to the subphylum of the cephalochordates, the most basally divergent group in our own phylum, the chordates. The privileged position of amphioxus permits the comparison of its methylation system with that of vertebrates, thereby allowing the inference of the methylation condition of the last chordate common ancestor. Such evolutionary framework helps to understand how changes

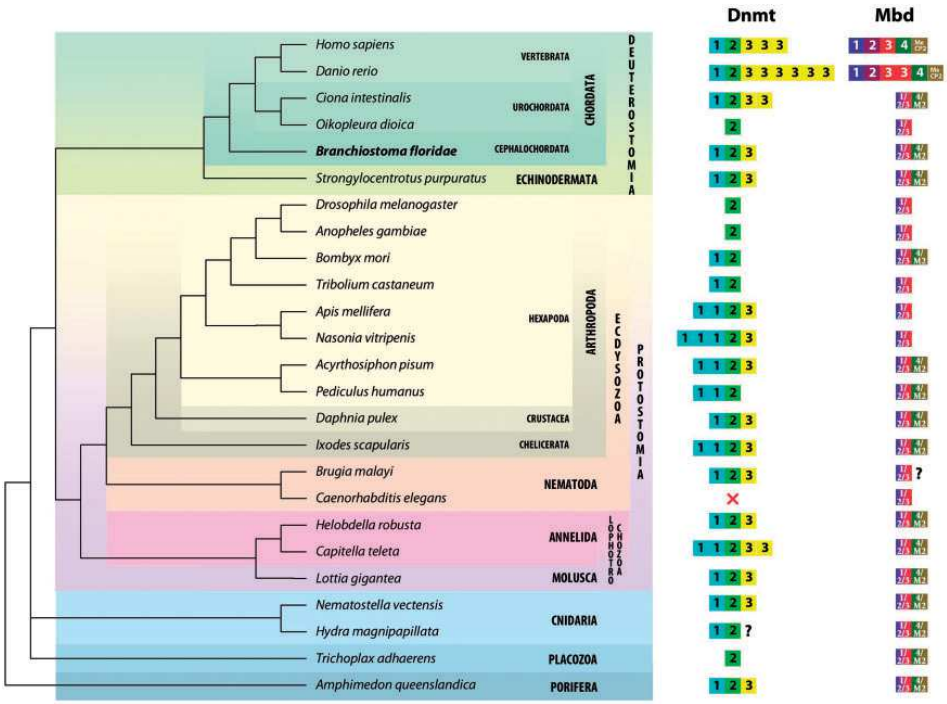


Figure 1: Phylogenetic distribution of *Dnmt* and *Mbd* toolkit. The *Dnmt* groups (*Dnmt1*, *Dnmt2* and *Dnmt3*, numbered boxes) and *Mbd* genes (*Mbd1*, *Mbd2*, *Mbd3*, *Mbd1/2/3*, *Mbd4*, *MeCP2* and *Mbd4/MeCP2*, numbered boxes) are shown in several animal species. The number of boxes represents the number of gene copies found in each species. Partial sequences similar to the glycosylase domain of Mbd4 are detected in *Bombyx mori*, *Acyrthosiphon pisum*, *Pediculus humanus* and *Amphimedon queenslandica*. The '?' denotes that the gene has not been found in incomplete genomes or EST databases. This figure summarizes the *Dnmt* and *Mbd* toolkit described in [31, 38, 44], as well as new sequences retrieved by BLAST searches from public databases: *Anopheles gambiae* Mbd1/2/3 (XP.318432), *B. mori* Mbd1/2/3 (BGIBMGA001425 in SilkDB) and Mbd4/MeCP2(BGIBMGA012841 in SilkDB), *Tribolium castaneum* Mbd1/2/3 (XP.969537), *Nasonia vitripenis* (NP.001164526), *A. pisum* Mbd1/2/3 (NP.001156167) and Mbd4/MeCP2(XP.003244914), *P. humanus* Mbd1/2/3 (XP.002428735) and Mbd4/MeCP2(XP.002425074), *Brugia malayi* Mbd1/2/3 (XP.001896671) and *A. queenslandica* Mbd1/2/3 (XP.003389039) and Mbd4/MeCP2(XP.003386267).

in the DNA methylation system might have led to functional innovations in the stem vertebrate genome. Here, we first review the main components of the DNA methylation machinery (i.e. *Dnmt* and *Mbd* genes) in amphioxus, and perform a comparative analysis with those in vertebrates. Second, we describe the pattern of distribution of the DNA methylation throughout the amphioxus genome, comparing it with that reported in other chordates. Finally, we discuss the evolution of the DNA methylation machinery in chordates, in the context of the major genomic changes related to the two rounds of whole-genome duplication (2R-WGD) that were concomitant with the origin and diversification of vertebrate innovations.

EVOLUTION OF THE CHORDATE DNA METHYLATION TOOLKIT: *DNMT* AND *MBD* GENES

The amphioxus DNA methylation toolkit

Genomic surveys revealed that amphioxus has three *Dnmt* genes, namely *Bf.Dnmt1*, *Bf.Dnmt2* and *Bf.Dnmt3*, each one ortholog to one of the three *Dnmt* groups [31, 37]. Structurally, the C-terminal

region of Dnmt enzymes contains the cytosine methyltransferase active site, organized into 10 motifs, of which 6 of them (I, IV, VI, VIII, IX and X) are evolutionary conserved [45–47]. These six motifs are recognizable in the three Bf_Dnmt sequences (Figure 2), suggesting the enzymatic conservation of the amphioxus Dnmt proteins. Each amphioxus Dnmt contains additional domains specific for each Dnmt group (Figure 2), which agrees with the tree topology obtained by phylogenetic analysis [31]. Hence, Bf_Dnmt1 contains the typical Dmap1-binding region, the nuclear localization signal (NLS), the Cys-rich zinc-binding domain and the two bromo-adjacent homology (BAH) motifs. Bf_Dnmt2 presents the characteristic Dnmt2 CFTXXYXXY (CFT) sequence between motifs VIII and IX. Bf_Dnmt3 contains the typical PWWP domain and the Cys-rich region in the N-terminal region of the protein.

Genomic surveys also revealed that amphioxus has two *Mbd* genes. Considering phylogenetic reconstructions [22, 31] and analyses of conserved synteny (discussed in the next section), here we suggest a new nomenclature that reflects their evolutionary relationships with vertebrate genes: *BfMbd1/2/3* (formerly *BfMbd2/3*) because it is the co-ortholog of vertebrate *Mbd1*, *Mbd2* and *Mbd3* paralogs, and *BfMbd4/MeCP2* (formerly *BfMbd4*) because it is the co-ortholog of vertebrate *Mbd4* and *MeCP2* paralogs. Structurally, Bf_Mbd1/2/3 contains, in addition to the MBD, the coiled-coil domain at the C-terminal region characteristic of vertebrate Mbd2 and Mbd3 [48], but lacks the CXXC motifs and the hydrophobic transcriptional repression domain (hTRD) of Mbd1 (Figure 2). Bf_Mbd4/MeCP2 concurs with the structural features of Mbd4 proteins, including the thymine DNA glycosylase domain and the Fas-associated death-domain (FADD) protein binding region in its C-terminal half [31], but lacks the highly basic transcriptional repression domain (bTRD) present in MeCP2 (Figure 2). Domain analysis reveals, therefore, that Bf_Mbd1/2/3 is more similar to vertebrate Mbd2 and 3 than to Mbd1, and Bf_Mbd4/MeCP2 is more similar to Mbd4 than to MeCP2. These similarities suggest that Bf_Mbd1/2/3, Bf_Mbd4/MeCP2 and vertebrate Mbd2, Mbd3 and Mbd4 might have maintained ancestral roles, whereas Mbd1 and MeCP2 might have acquired vertebrate-specific novel functions.

Most Dnmt and Mbd protein domains are involved in protein–protein interactions. Vertebrate Dnmt1 is able to bind different proteins, including Dnmt3 [49], histone deacetylase 2 (Hdac2), DNA methyltransferase 1 associated protein 1 (Dmap1) [50] and proliferating cell nuclear antigen (Pcna) [51]. Dnmt3 binds zinc fingers and homeoboxes 1 protein (Zhx1) [52], and Mbd2 and Mbd3 interact with many proteins in the NuRD/Mi-2 complex, including Hdac2, retinoblastoma binding protein 4 (Rbbp4), metastasis associated 1 (Mta1) and GATA zinc finger domain containing 2A (Gatad2a) [53, 54]. Homologs of these Dnmt- and Mbd partners have been identified in amphioxus (e.g. *Dmap1*, *Hdac2*, *Pcna*, *Rbbp4*, *Mta1* and *Gatad2a*) [31], and *in silico* reconstruction of a putative interaction network in amphioxus is therefore feasible (Figure 3). Although this *in silico* approximation is only an initial step that requires experimental validation, the identification in amphioxus of bona fide Dnmt and Mbd proteins and their plausible partners suggests the presence of an interaction network similar to that in vertebrates.

Other chordates such ascidians have also a full complement of the *Dnmt* and two *Mbd* genes [22, 29, 31], whereas the larvacean *Oikopleura dioica* has only *Dnmt2* and *Mbd1/2/3* (Figure 1) [37]. Outside chordates, the identification in sea urchin of *Dnmt1*, *Dnmt2*, *Dnmt3*, *Mbd1/2/3* and *Mbd4/MeCP2* genes [31, 56] indicates that the basic *Dnmt/Mbd* genetic toolkit was already present at least in the last common ancestor of deuterostomes, probably in stem bilaterians [31]. Hence, amphioxus appears to have preserved a DNA methylation machinery that represents the ancestral chordate toolkit. In contrast, within urochordates, though some species have also maintained an intact basic methylation toolkit, others such as larvaceans have lost most of their components [31, 37].

Vertebrate expansion of the DNA methylation toolkit

In contrast to non-vertebrate chordates, vertebrate *Dnmt* and *Mbd* gene families have suffered distinct degrees of expansions. *Dnmt3*, for instance, has been duplicated in vertebrates, and analysis of the human genome reveals that *DNMT3A* and *DNMT3B* paralogs are in conserved syntenic chromosomal segments in *Hsa*2 and *Hsa*20, arguing in favor of the hypothesis that both genes likely originated during the 2R-WGD, rather than via local duplications (Figure 4A). Genomic analysis of conserved synteny also reveals that *Mbd* gene family expansion was due to the 2R-WGD, and synteny

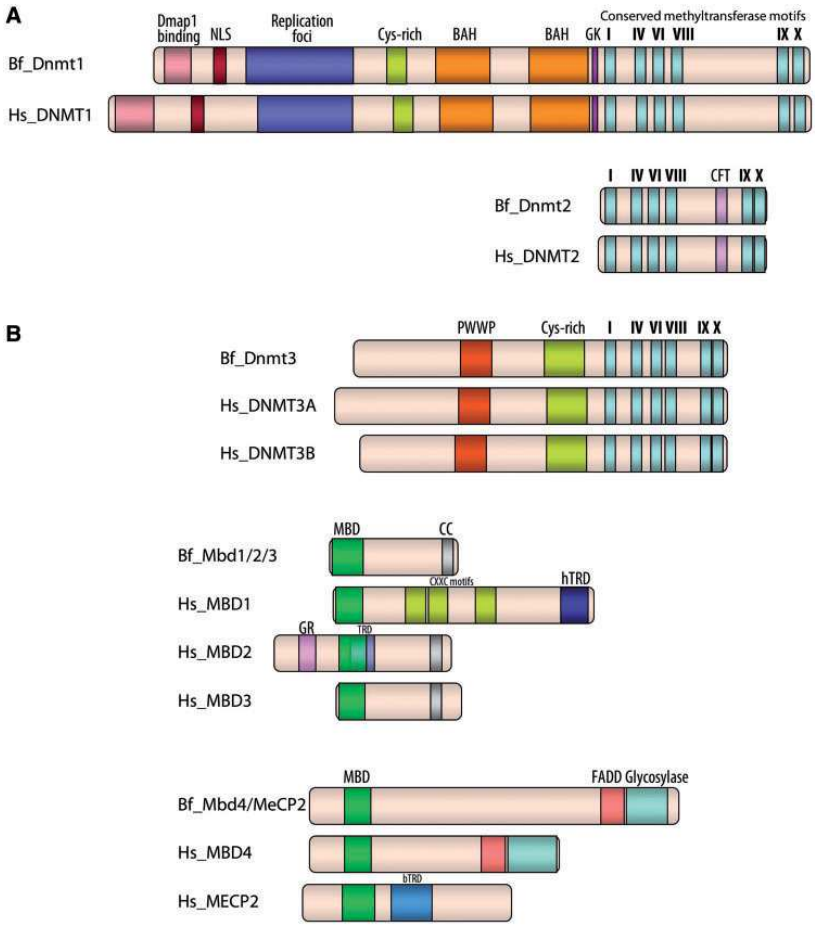


Figure 2: Structural protein domains of the DNA methylation toolkit. (**A**) Structural comparison of Dnmt enzymes in amphioxus (Bf) and human (Hs). Dnmt enzymes share six highly conserved methyltranferase motifs (I, IV, VI, VIII, IX and X) at the C-terminal region, and several group-specific domains, mostly found at the N-terminal region. A Dmap1-binding region, a nuclear localization signal (NLS), a region involved in targeting to replication foci, a Cys-rich zinc-binding domain, and two bromo-adjacent homology (BAH) motifs are distinctive of the N-terminal region of Dnmt1 enzymes, which is joined to the C-terminal catalytic region by a tract of glycine and lysine residues (GK). A CFTXXYXXY (CFT) sequence between motifs VIII and IX is characteristic of Dnmt2. A NLS, a PWWP domain and a Cys-rich region are typically found in the N-terminal region of Dnmt3. (**B**) Structural comparison of amphioxus and human Mbd proteins. The methyl-CpG binding domain (MBD), distinctively found at the N-terminal half of all these proteins, is capable to recognize and bind specifically to methylated DNA. In addition to the MBD, amphioxus Mbd1/2/3 contains a coiled-coil domain (CC) at the C-terminal region characteristic of vertebrate Mbd2 and Mbd3, but lacks the CXXC motifs and the hydrophobic transcription repression domains (hTRD) of the human MBD1. Amphioxus Mbd4/MeCP2 includes a thymine DNA glycosylase domain and a Fas-associated death-domain protein binding region (FADD) in its C-terminal half, distinctive of vertebrate Mbd4, but lacks the basic transcription repression domain (bTRD) of MECP2.

conservation is observed between the *MBD1/MBD2*-bearing paralogon in *Hsa*18 and the *MBD3/MBD3L*-bearing paralogon in *Hsa*19, as well as between the *MBD4*-bearing paralogon in *Hsa*3 and the MECP2-bearing paralogon in *Hsa*X (Figure 4B and C). Notably, genomic analysis reveals conserved synteny between the amphioxus *BfMbd1/2/3*-bearing paralogon within a 2-Mb fragment in

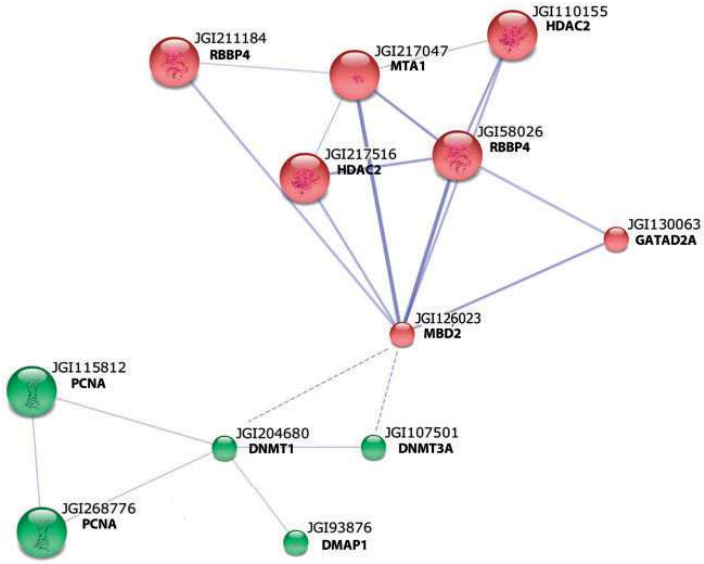


Figure 3: Amphioxus network of protein interactions linked to Dnmt and Mbd inferred by STRING v.9.0 [55]. Spheres represent amphioxus proteins connected to interacting partners by solid lines (confidence score > 0.3). Stronger interactions are represented by thicker lines. Amphioxus protein IDs from JGI and their respective human partners with the highest bit-score value in sequence comparisons are shown.

scaffold *Bf*_V2_32 and the *MBD1/MBD2* and *MBD3*/MBD3L paralogons in 10-Mb fragment in *Hsa*18 and a 20-Mb fragment in *Hsa*19, respectively (Figure 4D). It is tempting to speculate that there might be still unknown evolutionary constraints maintaining such high degree of synteny conservation during >500 Mya, the time of divergence between cephalochordates and vertebrates. Currently, different projects are being developed to uncover the causes that could explain the maintenance of macro- and micro-synteny conservation across such distantly related taxa [57].

Dnmt1 and *Dnmt2*, however, have been maintained as single copy genes despite the vertebrate 2R-WGD events. Although genomic analysis reveals conserved synteny between the *DNMT1* region on human chromosome 9 (*Hsa*9) and chromosome segments in *Hsa*1 and *Hsa*19, as well as between the *DNMT2* region on chromosome *Hsa*10 and chromosome segments in *Hsa*2 and *Hsa*17 (CC, unpublished data), none of the *DNMT1* or *DNMT2* ohnologs (i.e. paralogs originated by genome duplication) survived in their respective paralogous chromosomal segments. *Dnmt1* and *Dnmt2* have remained as single copy genes in chordates, suggesting that these genes might be refractory to copy number variation, a phenomenon observed in ohnologs that have dosage-balance evolutionary constraints [60].

DNA METHYLATION DISTRIBUTION PATTERN IN AMPHIOXUS

The extent of CpG methylation in the amphioxus genome was initially studied by comparing electrophoretic patterns of genomic DNA digested with methylation-sensitive and methylation-insensitive restriction endonucleases [39]. Southern blots of these digestions were hybridized with gene-specific probes to assess the methylation status of certain amphioxus genes, providing an approximate idea of the genomic distribution of this epigenetic mark. These early studies led to the conclusion that there are two roughly comparable fractions of methylated and unmethylated DNA in the amphioxus genome, with genes distributed in both fractions [39].

Sequencing of the whole genome of *Branchiostoma floridae* has allowed an estimate of the DNA methylation pattern throughout the amphioxus genome. The distribution of DNA methylation in

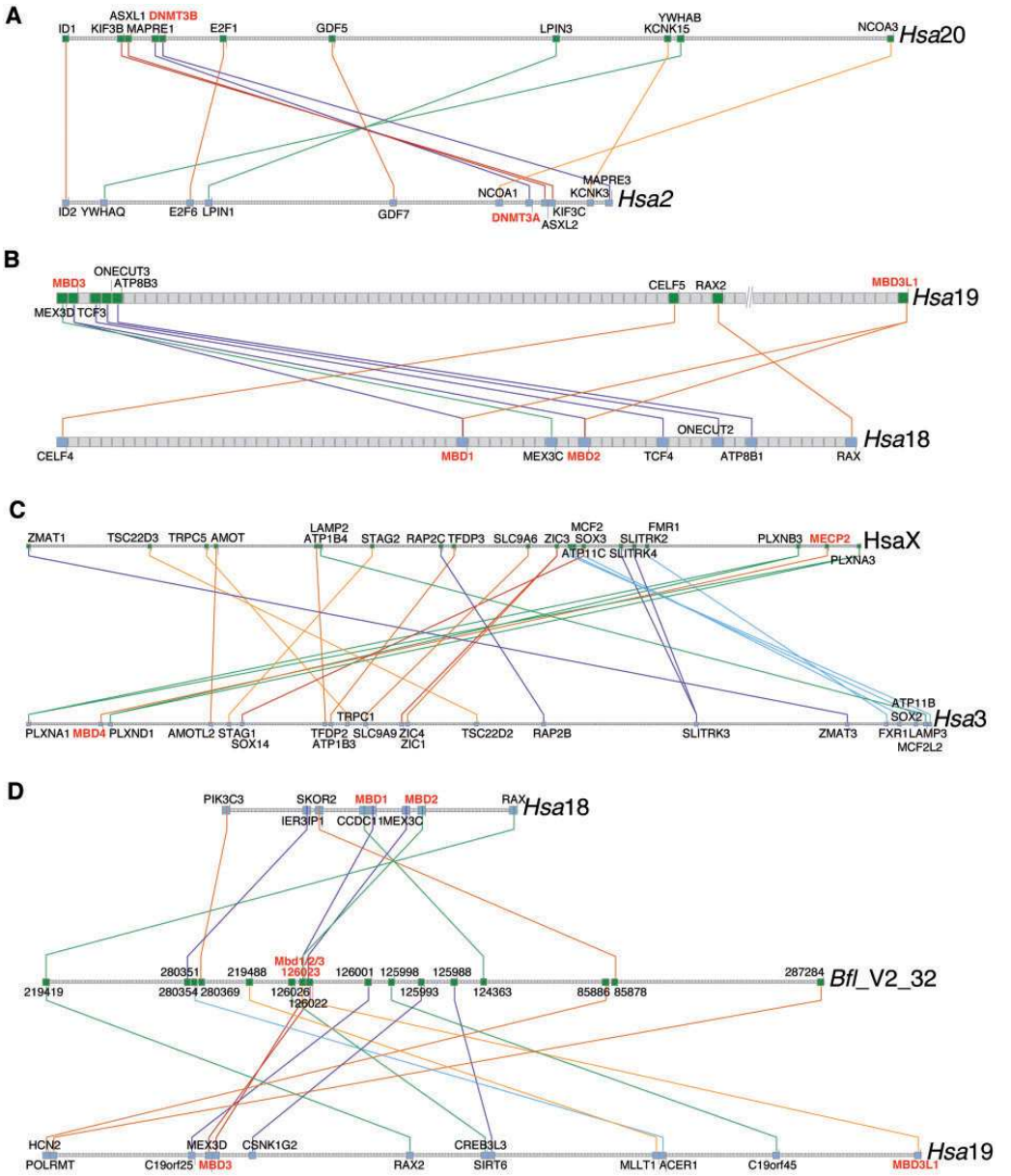


Figure 4: Conserved syntenies between *DNMT* (**A**) and *MBD* (**B** and **C**) bearing paralogons in human chromosomes provide evidence that the expansion of the DNA methylation toolkit was originated by the two rounds of whole-genome duplication (2R-WGD) that occurred in stem vertebrates. Interestingly, synteny for the MBD bearing paralogons has been preserved between human and amphioxus chromosomes for >500 Mya (**D**), which may suggest that selective evolutionary constrains might preserved the structural organization of this genomic region because of a potential functional role. *Hsa*N: human chromosome N; *Bfl*.v2.N, *B. floridae* scaffold N. The analysis of synteny conservation is inferred from the Synteny Database [58] using a 100-gene sliding window as previously described [59].

the genome can be inferred from the ratio between the observed and expected CpG content, a metric termed CpGo/e ratio or CpG score [41, 61]. In general, unmethylated regions show a CpGo/e ratios close to 1, whereas hiper-methylated regions tend to contain few CpG sites due to the mutability of 5-methylcytosine, therefore showing CpGo/e ratios far smaller than 1 (e.g. <0.5 as in humans). Using this indirect approach, Okamura *et al.* [42] analyzed one genomic region of 2 Mb, and concluded that amphioxus genome has a mosaic pattern of DNA methylation. The mean length of methylated and unmethylated regions was estimated 4.8 kb (SD = 5.0 kb) and 7.7 kb (SD = 11.1 kb), respectively, comparable to those of other non-vertebrate deuterostome genomes (e.g. 3.6 and 6.6 kb for sea urchin, and 5.3 and 6.8 kb for ascidians, respectively), but distinct from the global methylated genomes of mammals (e.g. 25.5 and 2.6 kb for human, respectively). This indirect approach has been experimentally validated in ascidian and human genomes [42], but no experimental analysis has been reported in amphioxus. To validate whether or not CpGo/e ratio is a good estimate of the DNA methylation status in amphioxus, we have examined here the CpGo/e ratios of 18 amphioxus genes that had been experimentally verified to be either methylated (8 genes) or unmethylated (10 genes). Our analysis shows that the CpGo/e ratios of 17 out of the 18 genes analyzed correlate with their methylation status (Figure 5A), supporting Okamura's analysis [42].

To test if the inferences about the methylation pattern made by Okamura and collaborators from the analysis of 2 Mb could be extended to other genomic regions, here we have analyzed the CpGo/e ratio across 17 Mb from three arbitrarily selected scaffolds of the amphioxus genome (Figure 5B–D). According to the CpGo/e values, the distribution of DNA methylation across different scaffolds shows a mosaic pattern, with methylated segments interspersed with similar segments of unmethylated DNA. Consistent with Okamura's analysis, the plot of the distribution of the frequencies of the CpGo/e ratios reveals statistically significant bimodal distributions in all regions analyzed (G^2 over 5% significance level >5.99; left and right curves in Figure 5E–G). The bimodal distributions have very similar means among different scaffolds. One of the distributions might represent the methylated fraction (average CpGo/e ratio ~0.4, left curve), whereas the other distribution may represent the hypo-methylated fraction (average CpGo/e ratio ~0.8, right curve). Interestingly, the relative weight of each fraction shows small differences among scaffolds, but globally the methylated fraction tends to be slightly smaller than the unmethylated one (Figure 5H). Future experimental analyses are needed to test these CpGo/e-based inferences that suggests that amphioxus follows an overall methylation pattern similar to that in ascidians, in which about half of the genome is heavily methylated, whereas the other half is poorly methylated [41, 42]. These non-vertebrate patterns are in sharp contrast with those of globally methylated vertebrate genomes, in which 80% of the CpGs are methylated [4].

In mammals and plants, DNA methylation has been associated not only to gene silencing, but also to transcriptional repression of transposable elements (TEs), decreasing their mobility, and thereby functioning as a mechanism of genome defense against their mutational activity [66–68]. The methylation status of the amphioxus transposons remains mostly unknown. Experimental data have been provided only for one type of TE, in which analysis of methylation-sensitive DNA digestion revealed that the non-LTR retrotransposon *Bf*-CR1 was predominantly free of methylation [69, 70]. Important caveats need to be considered if CpGo/e ratios are used to infer the methylation status of TEs due to their peculiar behavior. For instance, populations of TEs may suffer sequence bias due to bottlenecks derived from the burst of mobilization of a reduced number of TEs, and thereby the CpGo/e ratios of the new TEs could reflect the CpG ratio of the original genomic hosting region, rather than a biased ratio due to TE-specific targeted methylation. Another aspect to be considered is the possibility of horizontal transfers of TEs between different phyla, in which CpG ratios of recently acquired TEs might still reflect the methylation status of the donor genome rather than that of the host genome. However, considering that TE horizontal transfers between different phyla are rare, and that the TEs of amphioxus do not appear to have suffered significant recent expansions [71], it is plausible to assume that most of the current amphioxus TEs might be ancient and vertically inherited. We argue, therefore, that if TEs had been major targets for DNA methylation over a long evolutionary time in the amphioxus lineage, the CpGo/e ratios of most TEs should be biased toward low values.

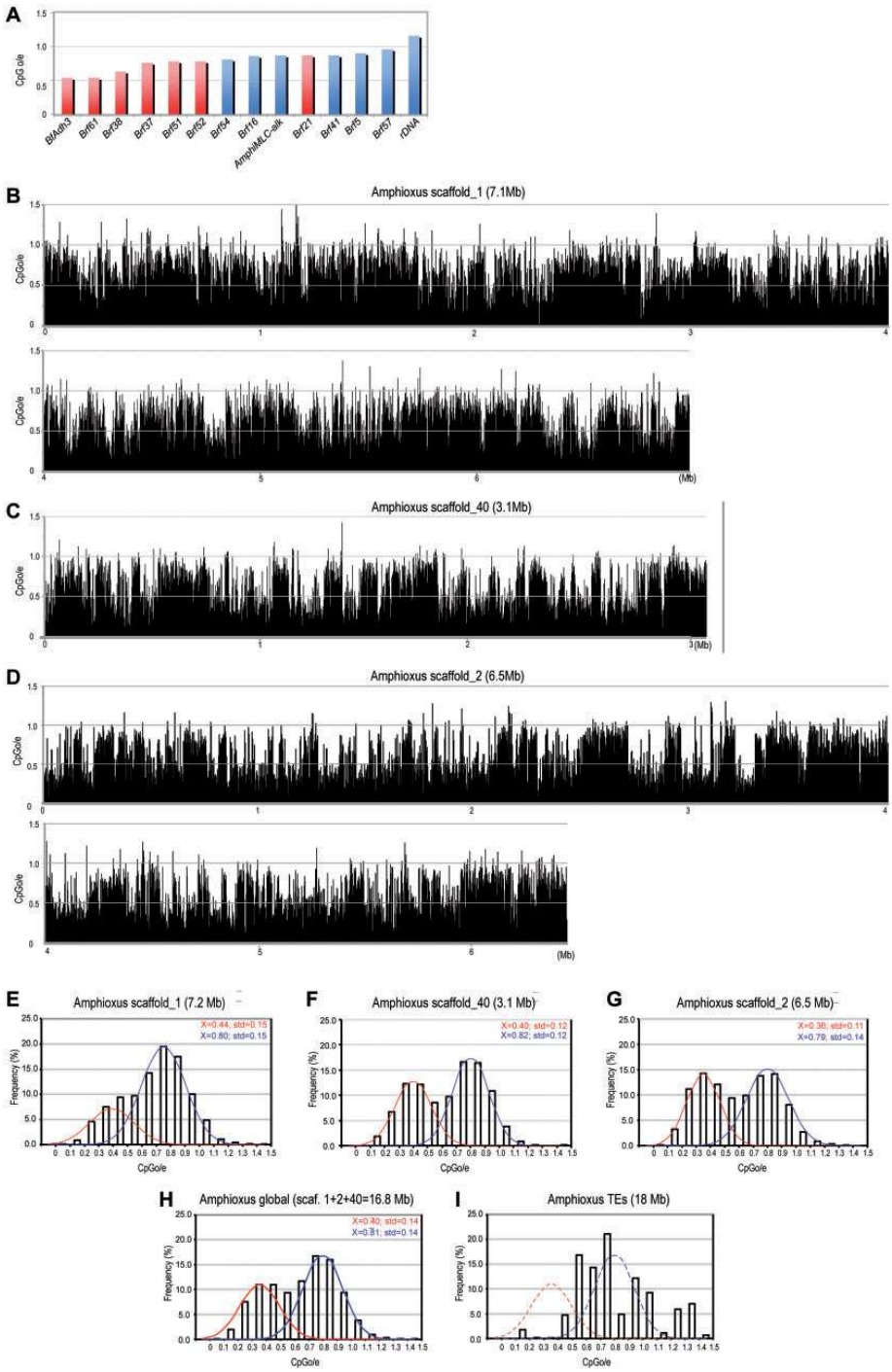


Figure 5: Mosaic pattern of DNA methylation in the amphioxus genome. (**A**) Histograms ordered by increasing values of the CpGo/e ratios of 18 amphioxus genes of known methylation status. Methylated genes include

(continued)

To test this possibility, we have analyzed the CpGo/e ratios of all the 623 TE families provided in the amphioxus genome database (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html), and the results do not show any obvious bias toward low CpGo/e ratios. In fact, only ~7% of TE sequences showed CpGo/e ratios <0.5 (0.4 is the average of the methylated genomic fraction in amphioxus), whereas >62% of TEs have CpGo/e ratios >0.7 (0.8 is the average of the hypo-methylated genomic fraction in amphioxus) (Figure 5I). These theoretical inferences suggest that, globally, TEs are not major targets of methylation, a similar condition to that described in ascidians, in which experimental evidence has proved that most TEs are poorly methylated [41, 70, 72]. If future experimental analysis (e.g. bisulfite sequencing) of amphioxus TEs corroborates this inference, we could conclude that the methylation of TEs as a defense mechanism might be an evolutionary innovation acquired in vertebrates, not present in cephalochordates and urochordates.

DNA METHYLATION, FROM ANCESTRAL FUNCTIONS TO NEW ROLES

What functions might DNA methylation be playing in amphioxus? In ascidians, which similarly to amphioxus also have a mosaic methylation pattern, DNA methylation is targeted to 'gene bodies' rather than to intergenic regions or TEs [41]. It has been proposed that DNA methylation in gene bodies might prevent aberrant transcription from cryptic initiation sites, which may interfere with the normal transcription machinery [73]. Zilberman *et al.* [73] have hypothesized that a large number of elongating RNA polymerases might prevent the use of cryptic sites in highly transcribed genes, whereas in moderately transcribed genes, methylation of gene bodies is needed to avoid aberrant transcripts. This hypothesis predicts that the methylated fraction of a given genome will be enriched in genes moderately transcribed, and that this enrichment will be perceptible in genomes in which DNA methylation has not been recruited for other functions such as silencing of TEs. The analyses of DNA methylation of ascidians and other protostome genomes agree with these predictions, and genes with weak–moderate expression (e.g. housekeeping genes) belong to the heavily methylated fraction, while genes with high tissue-specific expression belong to the poorly methylated fraction [41, 74–76]. Analysis of a small set of genes in amphioxus shows a tempting correlation between housekeeping genes in the methylated fraction [e.g. ribosomal proteins P2 (Brf6) and S6 (Brf61), mitochondrial carrier protein (Brf38)], and tissue-specific genes in the unmethylated fraction [e.g. Lipase

Figure 5: Continued
Brf6 (60S ribosomal protein P2, NW.003101422), Brf16 (NADH dehydrogenase subunit, NW.003101225), Brf32 (unidentified cDNA, NW.003101419), Brf38 (mitochondrial proton carrier-like protein, NW.003101433), Brf43 (snRNP-like protein, NW.003101422), Brf61 (40S ribosomal protein S6, NW.003101524) [39], Adh3 (alcohol dehydrogenase class III, from exons 2 to 6, NW.003101484) [62], and presinilin (AF369890) [63], and unmethylated genes are AmphiMLC-alk (alkali myosin light chain, NW.003101500), Brf5 (unidentified cDNA, NW.003101462), Brf21 (triacylglycerol lipase, NW.003101520), Brf37 (α-tubulin, NW.003101522), Brf41 (unidentified cDNA, NW.003101534), Brf51 (β-actin, NW.003101548), Brf52 (unidentified cDNA, NW.003101505), Brf54 (unidentified cDNA, NW.003101377), Brf57 (unidentified cDNA, NW.003101328) and rDNA (AF061796) [39]. (**B–D**) Pattern of DNA methylation inferred by CpGo/e ratios across 17 Mb from three arbitrarily selected genomic regions (scaffold_1: 7.2 Mb, scaffold_2: 6.5 Mb and scaffold 40: 3.1 Mb) splitted in 2 kb segments. This plot reveals a mosaic pattern of interspersed regions of heavily methylated and poorly methylated fractions across the amphioxus genome. Estimations of the CpGo/e ratio were performed with CpGproD algorithm [64]. (**E–G**) Analysis of the frequency distribution of CpGo/e ratios in 0.1 intervals revealed statistically significant (G^2 over 5% significance level >5.99) bimodal distributions of methylated (left curve) and unmethylated (right curve) fractions in the three regions analyzed. Test for uni- or bi-modal distributions was performed with the NOCOM software [65]. (**H**) Global analysis of the three previously selected scaffolds. (**I**) Analysis of the frequency distribution of CpGo/e ratios of all the 623 TE sequences provided in the amphioxus genome database (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html), representing a total of 18 Mb (Supplementary Table S1). The distribution of the frequencies of the CpGo/e ratios of all amphioxus TE families does not show a bimodal distribution as does the rest of the genome in amphioxus [dashed lines in (I)], and importantly, the distribution is not biased toward low values, suggesting that most TE might not be main targets of methylation in amphioxus.

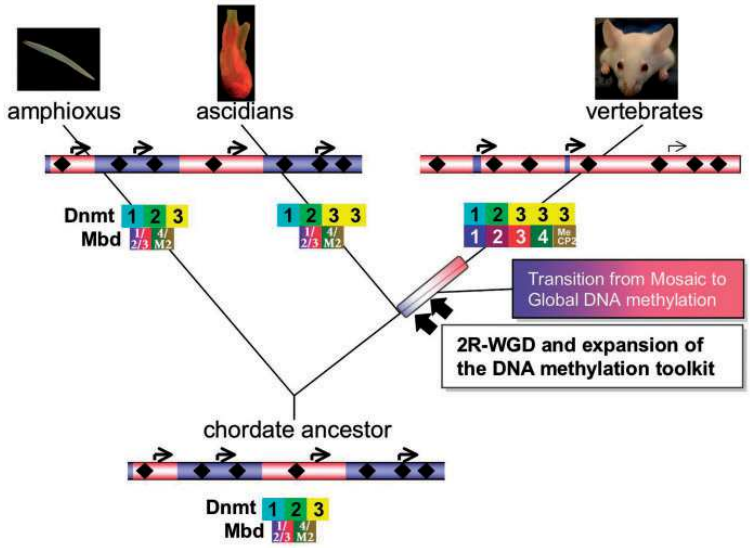


Figure 6: Evolution of DNA methylation in chordates. Mosaic patterns of DNA methylation in amphioxus and ascidian genomes include methylated segments (light shadow) interspersed with unmethylated fragments (dark shadow), in contrast with the globally methylated genomes of vertebrates (arrows represent gene promoters). TEs (black diamonds) are heavily methylated in vertebrates, whereas there is no evidence that TEs are direct targets of CpG methylation in nonvertebrate chordate asicidian and amphioxus. The transition from mosaic to global DNA methylation correlates with the 2R-WGD (solid arrows), and the expansion of the DNA methylation machinery (boxes as in Figure 1) may have facilitated the acquisition of novel or refined functions of the DNA methylation system in vertebrates. The amphioxus and ascidian toolkit and their mosaic pattern of DNA methylation likely reflects the ancestral chordate condition.

(Brf21), insulin-like protein, alkali myosin light chain] [39]. A definitive answer about the function of DNA methylation in amphioxus, however, awaits future analysis of large-scale DNA methylation footprinting in order to test if the mosaic-methylated amphioxus genome consists of poorly methylated segments enriched in highly expressed genes, whereas the significantly methylated segments mostly span throughout gene-bodies of weak–moderately expressed genes.

In summary, the mosaic pattern of DNA methylation in amphioxus and ascidians likely represents the ancestral condition of the chordate genome, while the global pattern of DNA methylation might be a vertebrate innovation (Figure 6). This radical change of DNA methylation patterns correlates with the occurrence of the 2R-WGD, and with the amplification of the *Dnmt* and *Mbd* gene families that originated during this event (Figure 6). It is plausible to postulate that after the 2R-WGD, positive selection might have facilitated the expansion of the DNA methylation toolkit, leading to the acquisition of new roles, improving, for instance, the silencing of spurious transcription of the increased number of vertebrate genes [77], counteracting deleterious gene dosage effects [78] or decreasing the mutational activity of TEs [66] in the newly polyploidized genome of stem vertebrates.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bfg.oxfordjournals.org/.

FUNDING

This work has been supported by grant 2009SGR336 from Generalitat de Catalunya and

grant BFU2010-14875 from the Ministerio de Ciencia e Innovación (Spain).

Key Points

- Phylogenetic inferences and analysis of conserved synteny suggest that amphioxus has a DNA methylation toolkit that represents the ancestral chordate condition.
- The expansion of the DNA methylation toolkit in vertebrates was originated by the 2R-WGD that occurred in stem vertebrates.
- Genomic sequence analysis suggests that amphioxus has a mosaic pattern of distribution of DNA methylation, similar to ascidians, but radically different to the globally methylated genome of vertebrates.
- The methylation status of amphioxus TEs remains unknown, but no evidence has been found that suggests that TEs are major targets of DNA methylation.
- The methylation of gene bodies, which might prevent transcription from cryptic initiation sites of moderately expressed genes, appears to be a potential role of DNA methylation in ancestral chordates.

Acknowledgements

We would like to thank to Jerzy Jurka for kindly sharing valuable TE information on the amphioxus genome. We appreciate the valuable comments of three anonymous reviewers for helping to improve the manuscript.

References

1. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. *Trends Genet* 2012;**28**:33–42.
2. Barlow DP. Genomic imprinting: a Mammalian epigenetic discovery model. *Annu Rev Genet* 2011;**45**:379–403.
3. Bird A, Macleod D. Reading the DNA methylation signal. *Cold Spring Harb Symp Quant Biol* 2004;**69**:113–118.
4. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;**16**:6–21.
5. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 2005;**74**:481–514.
6. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 1998;**19**:219–20.
7. Okano M, Bell DW, Haber DA, *et al*. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;**99**: 247–57.
8. Leonhardt H, Page AW, Weier HU, *et al*. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 1992;**71**:865–73.
9. Goll MG, Kirpekar F, Maggert KA, *et al*. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 2006;**311**:395–8.
10. Jeltsch A, Nellen W, Lyko F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 2006;**31**:306–8.
11. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev* 1989;**3**:612–9.
12. Comb M, Goodman HM. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res* 1990;**18**: 3975–82.
13. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 2000;**405**:482–5.
14. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 2003;**33**(Suppl):245–54.
15. Fatemi M, Wade PA. MBD family proteins: reading the epigenetic code. *J Cell Sci* 2006;**119**:3033–7.
16. Jones PL, Veenstra GJ, Wade PA, *et al*. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998;**19**:187–91.
17. Nan X, Ng HH, Johnson CA, *et al*. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998;**393**:386–9.
18. Wade PA. Methyl CpG-binding proteins and transcriptional repression. *Bioessays* 2001;**23**:1131–7.
19. Cross SH, Meehan RR, Nan X, *et al*. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet* 1997;**16**: 256–9.
20. Ng HH, Jeppesen P, Bird A. Active repression of methylated genes by the chromosomal protein MBD1. *Mol Cell Biol* 2000;**20**:1394–406.
21. Fujita N, Shimotake N, Ohki I, *et al*. Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. *Mol Cell Biol* 2000;**20**:5107–18.
22. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet* 2003;**19**:269–77.
23. Bird AP, Wolffe AP. Methylation-induced repression–belts, braces, and chromatin. *Cell* 1999;**99**:451–4.
24. Jiang CL, Jin SG, Pfeifer GP. MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. *J Biol Chem* 2004;**279**:52456–64.
25. Wong E, Yang K, Kuraguchi M, *et al*. Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci USA* 2002;**99**:14937–42.
26. Laget S, Joulie M, Le Masson F, *et al*. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS ONE* 2010;**5**:e11982.
27. Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 1998;**18**:6538–47.
28. Tweedie S, Ng HH, Barlow AL, *et al*. Vestiges of a DNA methylation system in Drosophila melanogaster? *Nat Genet* 1999;**23**:389–90.
29. Ponger L, Li WH. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 2005;**22**:1119–28.
30. Wang X, Grus WE, Zhang J. Gene losses during human origins. *PLoS Biol* 2006;**4**:e52.

31. Albalat R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus Branchiostoma floridae. *Dev Genes Evol* 2008;**218**:691–701.
32. Mandrioli M. A new synthesis in epigenetics: towards a unified function of DNA methylation from invertebrates to vertebrates. *Cell Mol Life Sci* 2007;**64**:2522–4.
33. Gutierrez A, Sommer RJ. Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis briggsae. *Nucleic Acids Res* 2004;**32**:6388–96.
34. Mita K, Kasahara M, Sasaki S, *et al*. The genome sequence of silkworm, Bombyx mori. *DNA Res* 2004;**11**:27–35.
35. Marhold J, Rothe N, Pauli A, *et al*. Conservation of DNA methylation in dipteran insects. *Insect Mol Biol* 2004;**13**: 117–23.
36. Wang Y, Jorda M, Jones PL, *et al*. Functional CpG methylation system in a social insect. *Science* 2006;**314**:645–7.
37. Cañestro C, Yokoi H, Postlethwait JH. Evolutionary developmental biology and genomics. *Nat Rev Genet* 2007;**8**: 932–942.
38. Lyko F, Maleszka R. Insects as innovative models for functional studies of DNA methylation. *Trends Genet* 2011;**27**: 127–31.
39. Tweedie S, Charlton J, Clark V, *et al*. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 1997;**17**:1469–75.
40. Rollins RA, Haghighi F, Edwards JR, *et al*. Large-scale structure of genomic methylation patterns. *Genome Res* 2006;**16**:157–63.
41. Suzuki MM, Kerr AR, De Sousa D, *et al*. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 2007;**17**:625–31.
42. Okamura K, Matsumoto KA, Nakai K. Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution. *BMC Bioinformatics* 2010; **11**(Suppl. 7):S2.
43. Feng S, Cokus SJ, Zhang X, *et al*. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* 2010;**107**:8689–94.
44. Werren JH, Richards S, Desjardins CA, *et al*. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* 2010;**327**:343–8.
45. Lauster R, Trautner TA, Noyer-Weidner M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J Mol Biol* 1989;**206**:305–12.
46. Posfai J, Bhagwat AS, Posfai G, *et al*. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res* 1989;**17**:2421–35.
47. Kumar S, Cheng X, Klimasauskas S, *et al*. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* 1994;**22**: 1–10.
48. Matsumoto M, Toraya T. cDNA cloning, expression, and characterization of methyl-CpG-binding domain type 2/3 proteins from starfish and sea urchin. *Gene* 2008;**420**: 125–34.
49. Kim GD, Ni J, Kelesoglu N, *et al*. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J* 2002;**21**: 4183–95.
50. Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet* 2000;**25**:269–77.
51. Chuang LS, Ian HI, Koh TW, *et al*. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 1997;**277**:1996–2000.
52. Kim SH, Park J, Choi MC, *et al*. Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. *Biochem Biophys Res Commun* 2007;**355**: 318–23.
53. Humphrey GW, Wang Y, Russanova VR, *et al*. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem* 2001;**276**:6817–24.
54. Brackertz M, Boeke J, Zhang R, *et al*. Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. *J Biol Chem* 2002;**277**:40958–66.
55. Szklarczyk D, Franceschini A, Kuhn M, *et al*. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2010; **39**:D561–8.
56. Aniello F, Villano G, Corrado M, *et al*. Structural organization of the sea urchin DNA (cytosine-5)-methyltransferase gene and characterization of five alternative spliced transcripts. *Gene* 2003;**302**:1–9.
57. Lv J, Havlak P, Putnam N. Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinformatics* 2011;**12**:S11.
58. Catchen JM, Conery JS, Postlethwait JH. Automated identification of conserved synteny after whole-genome duplication. *Genome Res* 2009;**19**:1497–505.
59. Cañestro C, Catchen JM, Rodriguez-Mari A, *et al*. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet* 2009;**5**: e1000496.
60. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci USA* 2010;**107**:9270–4.
61. Zeng J, Yi SV. DNA methylation and genome evolution in honeybee: gene length, expression, functional enrichment covary with the evolutionary signature of DNA methylation. *Genome Biol Evol* 2010;**2**:770–80.
62. Cañestro C, Hjelmqvist L, Albalat R, *et al*. Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. *Eur J Biochem* 2000;**267**:6511–8.
63. Martinez-Mir A, Cañestro C, Gonzalez-Duarte R, *et al*. Characterization of the amphioxus presenilin gene in a high gene-density genomic region illustrates duplication during the vertebrate lineage. *Gene* 2001;**279**:157–64.
64. Ponger L, Mouchiroud D. CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics* 2002;**18**: 631–3.
65. Ott J. Detection of rare major genes in lipid levels. *Hum Genet* 1979;**51**:79–91.
66. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997;**13**: 335–40.

67. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 2007;**8**: 272–85.

68. Rigal M, Mathieu O. A ''mille-feuille'' of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta* 2011;**1809**:452–8.

69. Albalat R, Permanyer J, Cañestro C, *et al*. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. *Cell Mol Life Sci* 2003;**60**:803–9.

70. Permanyer J, Gonzalez-Duarte R, Albalat R. The non-LTR retrotransposons in Ciona intestinalis: new insights into the evolution of chordate genomes. *Genome Biol* 2003;**4**:R73.

71. Cañestro C, Albalat R. Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. *Brief Funct Genomics* (in press).

72. Simmen MW, Leitgeb S, Charlton J, *et al*. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 1999;**283**:1164–7.

73. Zilberman D, Gehring M, Tran RK, *et al*. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 2007;**39**:61–9.

74. Lyko F, Foret S, Kucharski R, *et al*. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* 2010;**8**:e1000506.

75. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008;**9**: 465–76.

76. Foret S, Kucharski R, Pittelkow Y, *et al*. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics* 2009;**10**:472.

77. Bird AP. Gene number, noise reduction and biological complexity. *Trends Genet* 1995;**11**:94–100.

78. Chang AY, Liao BY. DNA methylation rebalances gene dosage after mammalian gene duplications. *Mol Biol Evol* 2012;**29**:133–44.

Apèndix 2

Ecological impact and developmental genetic responses of appendicularians to diatom-bloom derived biotoxins.

Nuria Paz Torres-Águila,[1] Josep Martí-Solans,[1] Alfonso Ferrández-Roldán,[1] Alba Almazán-Almazán,[1] Vittoria Roncalli,[1] Salvatore D'Aniello,[2] Giovanna Romano,[3] Anna Palumbo A,[2] Ricard Albalat[1] and Cristian Cañestro[1]

[1]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona. Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale 80121, Napoli, Italy

[3]Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale 80121, Napoli, Italy

Resum

La comprensió de l'impacte que les biotoxines derivades dels afloraments de microalgues podria tenir en el desenvolupament de les apendicularies, un dels components més abundants del mesozooplàncton, és d'interès ecològic primari, sobretot en la futura intensificació dels afloraments associats a l'escalfament i l'acidificació dels oceans. Aquest treball, després de realitzar tractaments amb aldehids poliinsaturats i extractes de diatomees productores d'oxilipines, revela que el desenvolupament embrionari de l'apendiculària *Oikopleura dioica* es veu compromès per aquestes biotoxines, fins i tot a concentracions del mateix rang que les mesurades després dels aflorament de mcroalgues. El nostre estudi exhaustiu dels marcadors genètics del desenvolupament revela un nou fenotip de "pilota de golf" causat per les biotoxines produïdes per les diatomees, que inclou el bloqueig de la morfogènesi, l'allargament de la cua i la convergència de la línia mitja, revelant-ne algunes de les seves bases moleculars. A més, la resposta genètica dels embrions exposats a les biotoxines inclou un silenciament global de la transcripció zigòtica dels gens del desenvolupament i una ràpida activació del sistema de defensa. El nostre treball, per tant, ens alerta sobre el possible impacte que pot tenir un augment de les biotoxines produïdes durant els aflorament de microalgues en cadenes tròfiques marines, i assenyala l'ús de gens de defensoma com a biosensors moleculars que els ecòlegs marins podrien utilitzar per controlar l'estrès genètic de les poblacions naturals d'apendiculàries exposades a afloraments d'algues en el context del canvi climàtic.

Ecological impact and developmental genetic responses of appendicularians to diatom-bloom derived biotoxins

Nuria P.Torres-Águila[1], Josep Martí-Solans[1], Alfonso Ferrández-Roldán[1], Alba Almazán-Almazán[1], Vittoria Roncalli[1], Salvatore D'Aniello[2], Giovanna Romano[3], Anna Palumbo[2], Ricard Albalat[1]* and Cristian Cañestro[1]*

[1] Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona. Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale 80121, Napoli, Italy

[3] Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale 80121, Napoli, Italy

*Corresponding authors

E-mail: canestro@ub.edu (CC)

E-mail: ralbalat@ub.edu (RA)

Keywords: Oxylipin and polyunsaturated aldehyde diatom biotoxins; Marine Ecotoxicity; Urochordate appendicularian *Oikopleura dioica*; Chordate EvoDevo; Defensome and Developmental mechanisms of embryonic stress response

Abstract

Understanding the impact that biotoxins derived from microalgal blooms might have on the development of appendicularians, one of the most abundant components of the mesozooplankton, is of prime ecological interest, especially upon the conceivable future intensification of blooms associated to ocean warming and acidification. This work, after performing treatments with polyunsaturated-aldehydes and extracts from oxylipin-producing diatoms, reveals that embryo development of the appendicularian *Oikopleura dioica* is compromised by these biotoxins, even at concentrations in the same range than those measured after blooms. Our exhaustive study of developmental gene markers reveals a novel “golf ball” phenotype caused by diatom biotoxins, which includes blockage of morphogenesis, midline convergence and tail elongation, and uncovers some of its molecular bases. Moreover, the genetic response of embryos exposed to biotoxins includes global silencing of zygotic transcription of developmental genes, and a rapid upregulation of the defensome system. Our work, therefore, alerts us on the potential impact that an increase of bloom biotoxins may have on marine food webs, and points to the use of defensome genes as molecular biosensors that marine ecologists could use to monitor the genetic stress of natural populations of appendicularians exposed to algal blooms in the context of climate change.

1. Introduction

Climate change pressures such as ocean warming and acidification may conceivably increase the frequency and severity of harmful algal blooms (HABs) in the near future, globally influencing marine planktonic systems [1]. Phytoplankton species causing HABs include diatoms, which are among the most important photoautotrophic organisms that drive food web dynamics in marine ecosystems [2]. Diatoms can produce different biotoxins, some of which may act as a defense mechanism against grazers due to its negative consequences on reproductive and developmental processes [3-6]. Among diatom-derived biotoxins, oxylipins are of prime interest because of their negative effects on the reproduction of copepods [4], the main grazers of these algae and one of the most abundant components of the mesozooplankton around the globe [7]. Oxylipins –including polyunsaturated aldehydes (PUAs), which are among the most investigated of the diatom-oxylipin family– are secondary metabolic end-products of a lipoxygenase/hydroperoxide lyase pathway that are toxic when released to the environment [8]. During algal blooms, cell membranes are easily broken by cellular friction and massive grazing or by diatom senescence at the end of the bloom, generating oxylipin-rich microzones that can alter the biology of neighboring organisms [9].

Oxylipin toxicity does not only affect copepods, but recent studies have shown that several marine species are compromised, including sea urchins [9-13], ascidian urochordates [14-16], as well as mollusks and annelids [10, 17, 18]. Despite analyses of different organisms have shown that PUA's toxicity can affect a wide range of physiological processes, including oocyte maturation, sperm mobility, fertilization, cell proliferation, embryogenesis, hatching, metamorphosis and apoptosis (reviewed in [19]), the molecular bases of PUA's toxicity remain unclear. Studies in copepods and ascidians have revealed, for instance, that PUAs affect the expression of stress response genes (i.e. defensome [20]) associated with the metabolism of the glutathione

system (e.g. *Gclm*, *Ggt* and *Gst*) [15, 16, 21, 22]. The expression of defensome genes related to aldehyde detoxification derived from lipid peroxidation (e.g. *Aldh2* and *Aldh8*) are also altered by PUAs, suggesting that these compounds may also induce oxidative stress [21, 23, 24].

While the physiological and defensive effects of PUA's toxicity have been investigated, the developmental genetic mechanisms affected by PUAs causing such functional alterations and embryo malformations remain largely unknown. In sea urchins and ascidians, expression analyses by qRT-PCR of developmental genes related to embryo patterning (e.g. *Hox*, *Parahox*) or signaling pathways (e.g. *Wnt and nitric oxide/Erk*) have revealed differences in the expression levels of some of them [15, 16, 24]. The mechanism by which PUAs induced those quantitative changes remains uncertain. Moreover, to our knowledge, there is no data describing what developmental processes and germ layers are altered by PUAs, nor what developmental gene pathways are deregulated by these toxins.

In this work, we study if PUAs do affect the embryonic development of the appendicularian urochordate specie *Oikopleura dioica* for two main reasons: first, from an ecological perspective, appendicularians (a.k.a. larvaceans) are relevant for marine food webs and carbon cycle; and second, from a evo-devo perspective, *O. dioica* is an amenable model animal for studying the evolution of developmental genetic mechanisms in response to environmental stressors, and investigating potential adaptions by gene loss [25].

From an ecological perspective, appendicularians are relevant because they are cosmopolitan pelagic filter-feeding organisms distributed in oceans all around the globe, considered the second most abundant after copepods in marine mesozooplankton [26], although in some offshore waters they are even more abundant than copepods [27]. Appendicularians graze about 10% of the ocean's primary productivity [28]. The appendicularian capability of trapping a wide range of particle sizes – from pico to nanoplankton (0.2–20 μm), thanks to their unique filter-feeding

apparatus that involves recurrent secretion of elaborated mucous houses– makes them to occupy an important trophic position in food webs. Appendicularians act as an important short-circuit that allows a rapid energy transfer from colloidal carbon and phytoplankton primary producers to zooplanktivorous predators such as fish larvae [29, 30]. Moreover, they contribute to at least 8% of the vertical transport of carbon to the deep ocean thanks to the rapid sinking of their discarded houses that can concentrate bacteria 1000 fold [31-33]. Therefore, the study of PUAs toxicity on the biology of appendicularians is fundamental from an ecological perspective to better understand the potential effect of increased oxylipins on marine food webs and carbon cycle, upon intensification of harmful diatom blooms due to ocean warming and acidification.

From a evodevo perspective, the chordate *O. dioica* is an attractive animal model because it has undergone a process of genetic and morphological simplification during the evolution of urochordates [25, 34-36]. The low genetic redundancy of *O. dioica* genome, –in comparison with the twice-duplicated vertebrate genome (reviewed in [37])–, together with the extraordinary amount of gene losses [25], make that its functional redundancy, as well as its mutational robustness, are lower than those in other animal species (e.g. vertebrates). In some cases, gene loss events have led this organism to be considered as an 'evolutionary knockout' that can facilitate the dissection of complex genetic networks and the understanding of sophisticated biological processes [25, 38-41]. *O. dioica* has lost, for instance, the retinaldehyde dehydrogenase 1a gene (*Aldh1a*) and most of the components of the metabolic pathway responsible for the synthesis of retinoic acid (RA) [42], a signaling molecule that in all other chordates plays a fundamental role in the regulation of embryo development, adult organ homeostasis and gametogenesis [43-45]. The loss of the *Aldh1a* in *O. dioica* is particularly relevant for the present work because it has been proposed that PUAs such as *trans,trans*-2,4-decadienal (DD) can compete with retinaldehyde for the substrate binding site of the Aldh1a enzymes [46]. This finding suggested the hypothesis that the teratogenicity of DD could be due to its interference with RA-synthesis, impairing thereby normal RA signaling during embryo development.

129 Thus, our work here aims to study *O. dioica* as an 'evolutionary knockout' for *Aldh1a*
130 for testing this hypothesis, because DD should produce minor or no alterations on
131 embryo development in *O. dioica* if RA signaling is the main developmental pathway
132 targeted by DD. On the contrary, if DD severely affects the development of *O. dioica*, it
133 would imply that other developmental mechanisms, distinct from RA signaling, are
134 affected by DD exposure. Altogether, our work demonstrates that *O. dioica* can be
135 used as an attractive and amenable model for marine embryo toxicology, as it has
136 been recently shown in other urochordate species [47, 48].
137

138 # 2. Results

139

140 ## 2.1. Doses-dependent effects of DD on embryo
141 ## development of *Oikopleura dioica*

142 To investigate the possible teratogenic effects and dose-response of DD on *O.*
143 *dioica* embryo development, we analyzed the embryos that successfully completed the
144 hatch after treatments with DD at different concentrations (from 0.05 to 2.5 µg/mL). In
145 contrast to DMSO-control conditions, DD-treatments altered embryo development in a
146 dose-dependent manner, in which the severity of the phenotype and the proportion of
147 affected embryos depended on the concentration of the DD treatment (**Fig 1**). We
148 classified the aberrant phenotypes in three categories: *i*) abnormal hatchlings, *ii*) pre-
149 tailbud arrested embryos with "golf-ball" morphology, *iii*) and 1-cell arrested zygotes
150 (**Fig1A-D**). According to the differences in the proportions of aberrant phenotypes
151 (statistical analyses are provided in **S1 Fig**), we classified DD concentrations in four
152 categories (**Fig 1E**): *i*) 'innocuous' (≤0.05 µg/mL), in which no significant differences
153 were observed in comparison with DMSO-controls; *ii*) 'mild' (from 0.075 to 0.125
154 µg/mL), in which most embryos did hatch, but the presence of aberrant morphologies

with shorten or kinked tails was higher than in the DMSO-control condition; *iii*) 'moderate' (from 0.25 to 2.0 µg/mL), in which most animals did not hatch, remaining arrested in a pre-tailbud stage with the appearance of a "golf ball" with no evidence of morphogenesis or organogenesis; *iv*) 'severe' (≥2.5 µg/mL), in which most zygotes remained at 1-cell stage and did not proceed with embryo cleaving.

2.2. "1-cell arrest" embryos by severe DD concentrations

To test if the absence of cell divisions at severe concentrations could be due to impairment of the fertilization, we scored for the formation of polar bodies after fertilization in the presence or absence of DD in 760 and 610 embryos, respectively (**Fig 2A-C**). Results showed no significant difference in the number of oocytes with polar bodies between DMSO-controls and DD-treatments (ANOVA test, P-value= 0.432), suggesting therefore that DD was not impairing oocyte fertilization. Careful inspection of oocytes by DIC microscopy did not reveal any evident morphological difference, neither in the shape and size of the internal granules of the oocytes, nor in the changes of membrane surface rugosity that normally occurs during fertilization (**Fig 2B-C**). To investigate if the internal organization of the cytoplasm of oocytes could be affected by DD, we performed whole-mount *in situ* hybridizations to detect maternal *Wnt* transcripts (**Fig 2D-E**). In DMSO-control 2-cell stage embryos at 30-minute post-fertilization (mpf), maternal *Wnt* signal appeared asymmetrically distributed, mostly accumulated near the cell membrane that will form the prospective posterior vegetal pole of the embryo (yellow arrowheads in **Fig 2D**, and unpublished data). In DD treated embryos at 30-mpf, maternal *Wnt* signal also appeared asymmetrically distributed near the membranes of the putative posterior vegetal pole, despite the absence of embryo cleavage (**Fig 2E**). These results, therefore, altogether demonstrated that DD did not impair fertilization in *O. dioica*, and the blocking of the constriction of the membrane to

form the cleavage furrow in DD-treated embryos seemed not to be due to changes of cellular architecture that drives the primary axial polarity of the zygote.

2.3. “Golf ball” pre-tailbud arrested embryos induced by moderate DD concentrations

At moderate DD concentrations (e.g. 1 µg/mL), the majority of embryos proceeded normally with embryo cleavage, but most of them never developed a proper tailbud morphology (**Figs 1C** and **3A**). Step-by-step comparison between the development of DMSO control and DD-treated embryos revealed no obvious differences up to the incipient-tailbud stage about 3 hpf. At that stage, however, while control embryos showed a clear indentation at the initial demarcation of the trunk and tail, DD-treated embryos did not show that indentation, and their morphology looked like a ‘golf ball’ (**Fig 3A**). Then, during tailbud stages, while control embryos continued with the progressive differentiation of their trunk and tail until the hatch, DD-treated embryos maintained the ‘golf ball’ morphology in an apparently arrested “pre-tailbud” stage, in which no trunk nor tail could be distinguished (**Figs 1C** and **3A**).

To test if the failure of the embryos to proceed to tailbud stages was due to the arrest of cell divisions, we counted the number of nuclei labeled with Hoechst in control and DD-treated embryos fixed at 1 h 10 min (64-cell stage), 1 h 30 min (≈100-cell stage), 1 h 50 min (≈200-cell stage) and 3 h post-fertilization (mid-tailbud stage) (**Fig 3B**). Results showed no significant differences in the number of cells between control and DD-treated embryos (ANOVA test, P-values>0.05), **Fig 3C**), suggesting that the absence of the typical differentiation of the trunk and tail, and the lack of obvious internal structures were not due to a slow-down or blockage of development caused by a deceleration or arrest of nuclear divisions, but to the failure of the process of morphogenesis that normally start at incipient-tailbud stage.

2.4 Genetic response of the molecular mechanisms of embryo development upon DD exposure.

2.4.1. DD causes a global delay in the zygotic expression of developmental genes

The fact that *O. dioica* has lost RA signaling [42] allowed us to discard that these abnormalities were due to alterations of this signaling pathway, contrary to what has been reported in other organisms [46]. To uncover what genetic mechanisms could be affected by DD resulting in a failure of morphogenesis and organogenesis, we performed expression analyses by whole-mount *in situ* hybridization with eight specific gene markers across different embryonic germ layers comparing six developmental stages between DMSO-control (**Fig 4A-AV**) and DD-treated embryos (**Fig 4A'-AV'**). With the exception of *Brachyury* (*Bra*) [49, 50], *Actin* (*Act*) [51], and *Nkx2.3/5/6* (comp 20628 in [52]), all other *O. dioica* genes have not been described yet, and due to their complex developmental expression patterns their exhaustive description will be published in future works. Here, they have been exclusively used as makers of derivatives of the three germ cell layers: mesoderm (notochord labeled by *Bra, Wnt* and *SoxBb*, and muscle labeled by *Act*), endoderm (entire endoderm labeled by *Tis11b*, only endoderm of the trunk labeled by *SoxBa* and *SoxBb*), and ectoderm (tail epidermis labeled by *SoxBb*, epidermis of the trunk labeled by *Tis11a,* and ventral epidermis of the trunk by *Nkx2.3/5/6*; neural cells by *Tis11a* and *SoxBa*, and presumptive placodal regions by *Wnt11* and *SoxBa*).

Expression analyses by whole-mount *in situ* hybridization revealed that the main alteration produced by DD exposure was that most embryos showed a systematic delay in the expression onset of all developmental genes analyzed. For instance, the first signal corresponding to the expression onset of *Bra*, *Act, Tis11a* and *SoxBa* was delayed from 32- to at least 64-cell stage **(Fig 4A, B'; G, I'; M, N'; S, U'**), the zygotic

expression onset of *Wnt11* was delayed from 64- to at least to 200-cell stage (**Fig 4AF, AH'**), and the expression onset of *SoxBb* was delayed from 200-cell stage to an embryo with a golf ball appearance equivalent to a mid-tailbud stage in the control condition (**Fig 4AB, AC'**). This delay was maintained during development since expression domains at later developmental stages in DD-treated embryos reminded those at earlier stages of control embryos (e.g. compare in **Fig 4** *Bra*: **A/B'-C'**, **C/E'** and **D/F';** *Act***: H/J'**, **I/K'**; *Tis11a* **O/P'**; *SoxBa* **T/U'**; *SoxBb* **AB/AC'**; *Wnt11* **AF/AH'**; *Nkx2.3/5/6* **AN/AP'**; *Tis11b* **AT/AV'**). The fact that we had found out that the total number of cells at each stage did not significantly differ between control and DD-treated embryos (**Fig 3B,C**) suggested that the delay of expression could not be explained by an overall developmental slowdown due to a decrease of the rate of cell divisions, but rather to alterations in the regulatory mechanisms of the transcription of these genes.

When we scored the number of DD-treated embryos that showed or not expression signal in each stage (insets in **Fig 4**), we found that the delay of the expression onset could reach extreme situations of detecting no signal at all in a majority of DD-treated embryos at stages in which gene expression was already detected in 100% of their control counterparts (**Fig 4:** *Bra* **A';** *Act* **G';** *Tis11a* **M'***SoxBa* **T';** *SoxBb* **AB';** *Wnt11* **AF';** *Nkx2.3/5/6* **AN';** *Tis11b* **AR'**). Remarkably, while the proportion of DD-treated embryos with no expression between 32- and 200-cell stages was high (73%, 210 out of 288), this proportion decreased in "golf ball" embryos equivalent to tailbud stages (31%, 58 out of 190). To determine whether the absence of signal was due to RNA degradation or repression of zygotic transcription, we focused on the expression signal of *Wnt11* gene, which has both maternal and zygotic components. Results revealed that while signal from the maternal component was observed up to the 64-cell stage in control embryos, time at which zygotic transcription started (**Fig 4 AE, AF**), in DD-treated embryos, however, only signal from the maternal component was observed and no signal from zygotic expression of *Wnt11* was detected until the 200-cell stage (**Fig 4**

AE'-AG´). This finding suggested that the absence of expression signal in DD-treated embryos was not due to transcript degradation, but it was consistent with a global silencing of zygotic transcription of developmental genes, which was more conspicuous at earlier stages up to 200-cell stage than at later stages equivalent to tailbud in control embryos.

2.4.2. DD does not delay gastrulation, but alters midline convergence and posterior elongation at later stages

To test if the global delay of the expression of developmental genes caused by DD was also accompanied by a delay of the onset of cell invagination during gastrulation, we analyzed *Bra* expression, which in agreement with previous reports [49] labeled one pair of blastomeres (A21) at the surface in embryos at 32-cell stage, and two pair of cells that already occupied an internal position (A211 and A212) after invagination as gastrulation proceeded at 64-cell stage in control embryos (**Fig 4A, B**). In DD-treated embryos, however, no signal was detected at 32-cells stage, and the first *Bra* signal was observed in only one pair of internal cells rather than two pairs as expected at the 64-cell stage (**Fig 4A', B'**). The fact the first two cells expressing *Bra* were already located in an internal position of the embryo (rather than the surface) suggested that DD did not delay the process of gastrulation, but simply altered the time at which *Bra* started to be expressed.

Analyses of the temporal dynamics of markers such *Bra* and *Act* revealed that the process of midline convergence seemed to be clearly delayed in DD-treated embryos. For instance, while the processes of notochord intercalation and muscle midline bilateral-convergence were completed by 2 hpf at 200-cell stage in control embryos (**Fig 4D**, **J**), in DD-treated embryos these processes were not completed up to 2 hour later by the time at which normal animals were already hatching (**Fig 4F'** and **L'**). Moreover, the length of the *Bra* and *Act* expression domains appeared to be much

shorter in DD-treated embryos than control ones, suggesting that not only the midline convergence was affected, but also posterior elongation was impaired.

2.4.3. DD blocks morphogenesis, but does not alter axial patterning or cell fate determination

Analysis of expression of eight developmental markers in DD-treated embryos with "golf ball" phenotypes revealed that the setup of the embryonic axes had been already established, and the fate determination of tissues derived from the three germ layers did not appear to be affected. For instance, the setup of the anterior-posterior (AP) axis in embryos with "golf ball" morphology was revealed by the presence of *Bra, Act* and *SoxBb* expression domains restricted to the presumptive posterior half of the embryo – (labeling the mesodermal precursor of the notochord, muscle cells and the ectoderm of the tail, respectively, **Fig 4 E'-F', J'-L´** and **AC'-AD'**), as well as *SoxBa*, *SoxBb*, *Tis11a*, and *Nkx2.3/5/6* expression domains restricted to the presumptive anterior half of the embryo (labeling the endoderm, and the ectoderm of the trunk **Fig 4 W'-X', AD', Q'-R´** and **AO'-AP'**). The setup of the dorso-ventral (DV) axis was revealed by the restricted expression of *Nkx2.3/5/6* to the presumptive ventral epidermal domain of the trunk (**Fig 4 AO'-AP'**). In addition to the endodermal, mesodermal and ectodermal precursors labeled with the aforementioned markers in DD-treated embryos, the presence of ectodermal *SoxBb* and *Wnt11* expression domains in precursor cells of the Langerhans receptors in the tail-trunk transition area and in the most anterior part of the embryo (**Fig 4 AD', AI'** and **AJ'**) suggested that the fate cell commitment of placode and neural cell precursors was not affected by DD. Thus, despite that DD seems to block morphogenesis and the formation of the trunk and tail in embryos with a "golf ball" phenotype, the setup of the embryonic axes, as well as the fate determination of tissues derived from the three germ layers, seemed not to be altered by DD.

2.5. Genetic response of 'defensome' during embryo development upon DD exposure

Previous works on sea urchins had revealed the existence of a "defensome", a set of stress response genes that activates their expression during development to protect embryos upon environmental stress [20]. To test if DD upregulates defensome genes during *O. dioica* development, we have analyzed the expression pattern of four defensome genes –three previously described dehydrogenase genes *Adh3* [53], *Aldh2* [54] and *Aldh8* [42], and a here newly described glutathione-related gene (Glutamate-cysteine ligase modifier subunit; *Gclm*)– in six developmental stages from 8-cell stage to late hatchling of DMSO-control (**Fig 5A-X**) and DD-treated embryos at mild concentrations (e.g. 0.1 µg/mL) (**Fig 5A'-X'**). In control embryos, the expression signal of all four genes was almost negligible at early developmental stages, and it was only detected in mid/late hatchling stages, in which *Gclm* and *Adh3* showed obvious expression domains restricted to digestive compartments (e.g. mouth, pharynx, esophagus and stomach lobes) (**Fig 5A-F** and **G-L,** respectively), and *Aldh2* and *Aldh8* showed either no signal or a weak and scattered signal (**Fig 5M-R** and **S-X,** respectively). In DD-treated embryos, however, the expression of all four genes was upregulated since very early developmental stages, showing broader and stronger staining than control embryos (**Fig 5A'-X'**). The obvious up-regulation of these four genes in response to DD, therefore, suggested that DD was able to trigger a genetic response of members of the 'defensome' involved in aldehyde detoxification, and especially members that respond to alterations of the metabolism of the glutathione system such as *Gclm*, which was the gene that showed the strongest up-regulation in response to DD (**Fig 5A'-F'**). Moreover, the fact that the expression of defensome genes was already noticeably upregulated by 8-cell stage, a time at which normal zygotic expression of most genes had not started yet, suggested that these genes had an immediate response capability upon DD challenge (**Fig 5A', G', M'** and **S'**).

2.6. Effects of diatom extracts on *Oikopleura dioica* embryogenesis

To test if similar developmental abnormalities induced by DD on *O. dioica* could be also caused by natural PUAs or other oxylipins produced by diatoms frequently involved in algal blooms, we tested the effect of the extracts from different microalga species on *O. dioica* embryogenesis. These species included *Skeletonema marinoi (Sm)*, which has been characterized by its high content of the PUAs heptadienal and octadienal and the lack of decadienal [5, 55]; *Chaetoceros affinis* (*Ca*), which produces other oxylipins different from PUAs [8]; *Prorocentrum minimum* (*Pm*), a dinoflagellate species used as a negative control that does not produce oxylipins [8]; and *Chaetoceros calcitrans* (*Cc*), a diatom included in the diet used to feed *O. dioica* in our animal facility [41], but for which the capability to produce PUAs was unknown. First, we studied the dose-response of *O. dioica* development at different concentrations of *S. marinoi* extracts (from 5 to 100 µg/mL **S2A** and **B Fig**). Results revealed that the most concentrated extracts reproduced the same phenotypes observed during DD treatments (**S2C** and **D Fig**). Our results from treatments with *S. marinoi* extracts at 100 µg/mL – which according to calculations reported in previous studies [55] contained PUAs at a concentration of approximately 1.0 µg/mL– revealed that the "golf ball" morphology of pre-tailbud arrested embryos was the most abundant phenotype, suggesting that the effects of the *S. marinoi* PUAs were comparable with the moderate, but not severe, condition observed in DD-treatments (**Figs 1** and **S2**). These results were consistent with previous observations showing that higher concentrations of heptadienal and octadienal, the main PUAs of *S. marinoi*, were required to cause similar teratogenic effects to those of DD during sea urchin embryonic development [12].

To test whether treatments with *S. marinoi* extracts were also capable to trigger a defensome response as observed with DD, we performed *in situ* hybridization experiments with *Gclm*. Results revealed that *Gclm* expression was upregulated even

after treatments with extracts at concentrations of only 10 µg/mL, in which no significant difference was observed in the proportion of aberrant phenotypes between Sm-extract treated embryos and the DMSO-control condition (P-value= 0.2611; **Figs 6A** and **S2**). We concluded therefore that potential secondary metabolites produced by *S. marinoi* were also capable to activate the expression of the defensome.

Finally, we tested whether *O. dioica* development was also affected by extracts of other microalgae that produce other secondary metabolites different from those of *S. marinoi*, using diatom extracts of *C. affinis* and *C. calcitrans*, as well as the dinoflagellate *P. minimum* as negative control. In all cases, with the exception of the negative control *P. minimum*, extracts affected *O. dioica* development giving rise to the same phenotypes as observed with DD and *S. marinoi* extracts (**Figs 6B** and **S2D**). Results revealed that the proportion of normal embryos in *C. affinis* extract treatments seemed even lower than those observed with *S. marinoi* extracts, suggesting that *O. dioica* development could be even more sensitive to the toxicity of other oxylipins different than PUAs (**Fig 6B**). In the case of *C. calcitrans*, a diatom whose production of PUAs or other oxylipins had not been investigated to our knowledge, we observed that embryo development was similarly affected by treatments with *S. marinoi* extracts (**Fig 6B**). To our knowledge this is the first finding suggesting that *C. calcitrans* is able to produce toxic secondary metabolites, which indeed is consistent with personal observations during the feeding strategy in our *O. dioica* facility in which excess of this alga in the diet made *O. dioica* cultures to decline [41].

3. Discussion

3.1. Ecological impact of biotoxins derived from diatoms on appendicularians

Understanding the potential impact that global climate change may have on oceans' health, and particularly how the potential increase of HABs may impact marine trophic webs is one of the main challenges that our Society is currently facing [1] (see focus areas and agenda of UNESCO-IOC). In this context, results from this work reveal that biotoxins from the oxylipin family that are abundantly released at the end of diatom blooms can compromise embryonic development of appendicularian species, and therefore their reproduction. This finding is ecologically relevant because appendicularians together with copepods, which are also affected by these biotoxins [4], represent the two most abundant components of the mesozooplankton around the globe, playing a crucial role in the global dynamics of marine food webs [26]. Our analysis of the development of *O. dioica* embryos exposed to DD, which is the most investigated PUA among those synthesized by many bloom-forming diatoms species [3, 56], reveals a characteristic set of embryonic aberrant morphologies with frequencies that follow a defined dose-response pattern (**Fig 1**). Our work, furthermore, reveals that the same set of embryonic aberrant morphologies is also induced by treatments with extracts from diatom species that produce high levels of heptadienal and octadienal (*S. marinoi* [5, 55]), or other oxylipins different from PUAs (*C. affinis* [8]). Also extracts from species such as *C. calcitrans*, whose oxylipin content is still unknown, affect *O. dioica* embryogenesis.

At present, our understanding of how PUAs toxicity released after blooms affect marine organisms in their natural habitats remains uncertain. Previous studies have quantified dissolved PUAs after blooms events in a range from 7.61×10^{-6} to 0.02 µg/mL

[57], questioning laboratory experiments in which the concentrations of PUAs are of an order of magnitude higher than those described in the field [5, 56]. Although in our study the most dramatic embryonic malformations mostly appear at concentrations of orders of magnitude higher than in the field (i.e. one-cell arrest at >2.5 µg/mL or "golf ball" morphology at >0.25 µg/mL), we want to remark that we observe lethal embryonic abnormalities at concentrations as low as 0.075 µg/mL, which is in the same order of the maximum value measured in the field after diatom blooms (e.g. 0.02 µg/mL after a bloom of *S. marinoi* in the northwestern Adriatic Sea [57]). Thus, despite we do not find significant differences between the development of control embryos and those that have been exposed during 4hpf to DD at concentrations lower than 0.05 µg/mL, we cannot exclude that long or repeated exposure of appendicularians to peaks of PUAs released after blooms (see [56, 58]), or even synergistic effect of simultaneous exposure to multiple PUAs may have detrimental effects on appendicularians' biology in natural habitats. Moreover, our finding that longer exposures before fertilization, that facilitate eggs to accumulate DD through the chorion, increase the sensitivity of appendicularian development to DD toxicity makes plausible to suggest that adult animals exposed to DD throughout generations may accumulate this biotoxin in their gonads and become, therefore, more vulnerable to its detrimental effects. This possibility might explain the observation that overfeeding *O. dioica* cultures with diatom *C. calcitrans* appears to lead culture declining [41]. Future experiments of longer exposure of *O. dioica* adults at low DD doses or overfeeding with excess of oxylipin-producing diatoms will be necessary to further investigate the vulnerability of appendicularians to the accumulation of these biotoxins throughout multiple generations.

Finally, our finding of the highly sensitive expression response of defensome genes (e.g. *Gclm*, **Fig 5**) to low concentrations of PUAs and other oxylipins opens the possibility to use this marker as a biosensor to test for genetic stress of natural appendicularian populations that may have been exposed to algal blooms in their habitats. Interestingly, increased expression of *Gclm* has been also reported in *Ciona*

intestinalis larvae treated with DD [16], suggesting that the use of *Gclm* expression as a biosensor can be used in other urochordates, and possibly other organisms, to monitor ecosystems with different exposition to blooms, thus providing a better understanding and a prediction of the potential impact of increasing HABs associated for instance to ocean warming and acidification.

3.2 Developmental genetic responses and alterations upon exposure to biotoxins derived from diatoms

Despite PUAs and other oxylipins produced by diatoms can interfere with a wide spectrum of physiological processes in many marine animals [9-18], the genetic pathways altered by these biotoxins remain unclear, especially during embryo development. The fact that PUAs such as DD had been described to be able to compete with retinaldehyde for the substrate binding site of the Aldh1a enzymes had led to suggest that genetic alterations due to a disruption of RA signaling could be a potential explanation for the physiological and developmental alterations caused by DD [46]. Contrary to this suggestion, our study takes advantage that *O. dioica* is an evolutionary knockout of RA signaling [42] and demonstrates that DD can cause dramatic developmental alterations independently of RA signaling.

To our knowledge, this work is the first general attempt to fill the gap between the teratogenic phenotypes and the developmental genetic responses induced by oxylipins by expression analyses of whole-mount in situ hybridization. Our results show for the first time which developmental processes are altered and which ones are not affected by DD. Indeed, early developmental processes, such as cleaving pattern, gastrulation, and cell division rate do not seem to be affected by this biotoxin. DD-treated embryos, however, fail to begin the process of morphogenesis, and develop a “golf ball” morphology in which the indentation that normally separates the trunk from the tail at incipient-tailbud stage never appears. Molecular characterization of DD-treated

embryos reveals a systematic delay of the onset of the expression of developmental genes, completely abolishing in many cases zygotic expression until late developmental stages. The fact that defensome genes highly upregulate their expression as a quick response to the presence of DD, even as early as at the 8-cell stage, when most genes have not yet started normal zygotic transcription, indicates that downregulation of developmental genes does not result from a general transcriptional repression. On the contrary, it points to the activation of a silencing mechanism for developmental genes, especially during early stages, that could be part of the embryonic response to environmental challenges.

Despite the absence of morphogenesis, the expression of specific cell markers in "golf ball" reveals a correct cell fate differentiation of the main derivatives of the ectodermal, mesodermal and endodermal germ lines, as well a correct establishment of AP and DV axes that confer a correct spatial regionalization in DD-treated embryos (**Fig 7**). Analyses of the position and size of expression domains of developmental markers show that midline convergence and tail elongation are notably affected in DD-treated embryos, suggesting that genes regulating these two developmental processes may be linked to the formation of the "golf ball" morphology. Future transcriptomic profiling by RNAseq and functional approaches will be necessary to uncover the entire set of defensome genes that respond to DD exposure, as well as the developmental gene pathways altered by DD which are responsible of the failure of morphogenesis in *O. dioica*.

4. Material and Methods

4.1 Animal culture and embryo collection

Oikopleura dioica embryos were obtained by *in vitro* fertilization from adults cultured in our animal facility at the University of Barcelona as described in [41].

4.2 Decadienal treatments

Previous studies about the effect of *trans,trans*-2,4-decadienal (DD, Sigma-Aldrich®, W313505) on ascidians embryos had shown that the chorion could act as a barrier that slow down the access of DD to embryos [14, 15]. Since our experience had shown that dechorionation is not possible in *O. dioica* embryos because it lethally affects development, to test if the chorion of *O. dioica* could also be a barrier against DD, we scored the hatching success of embryos that had been pre-incubated in DD previous to fertilization during different times (**S3 Fig**). Thus, we empirically determined that despite the chorion of *O. dioica* also act as a barrier for exogenous DD, 10 minutes (min) of pre-incubation with DD before fertilization was enough to overcome the barrier. We therefore implemented a 10-min egg pre-incubation as the standard procedure in all our DD treatment experiments.

For DD treatments, pools of *O. dioica* eggs were incubated before fertilization and during development with nine increasing concentrations of *DD* 0.05, 0.075, 0.1, 0.125, 0.25, 0.5, 1.0, 2.0 and 2.5 µg/mL. Stock solution of 15 mg/mL of DD was prepared in dimethyl sulfoxide (DMSO), avoiding light exposure. To minimize the toxic effect of DMSO in the development of *O. dioica,* the stock solution was diluted 1:500 (30 µg/mL) in sterilized seawater (sSW). This diluted stock was used to create each incubation solution containing the corresponding concentration of DD.

Eggs were pre-incubated for 10 min before fertilization into 900 µL of DD incubation solution. Then eggs were fertilized by adding 100 µL of sperm dilution (the sperm of 2 males diluted into 1 mL of sSW). 5 min after the *in vitro* fertilization, fertilized eggs were transferred into a 50 mm-diameter-glass dish with 3 mL of the corresponding incubation solution. Control embryos were incubated in DMSO at 0.02%(v/v) without DD. Embryos were cultured at 19°C until hatching (≈4 hours post fertilization, hpf), time

at which the effects of DD were scored. At least three replicas (r) containing hundreds of embryos (n) were analyzed for each condition.

4.3 Expression analyses and nuclear staining

Developmental gene markers in *O. dioica* were identified by *in silico* screening of the Oikobase [59], using *Ciona intestinalis* proteins as starting queries for tBLASTn searches against EST and genomic databases. The orthology of proteins was initially deduced by reciprocal best BLAST search against the *Ciona* genome (KH from Aniseed database), and corroborated by phylogenetic analysis (data not shown). Whole-mount *in situ* hybridizations using probes of the genes of interest, which were generated by PCR amplification of cDNA with gene-specific primers (**Sup. table 1**), were performed as described in [42]. Embryos were fixed in 4% paraformaldehyde in fix buffer (0.1M MOPS, 0.5M NaCl, 2mM MgSO4, 1mM EGTA) for 1 h at room temperature, dehydrated in 70% ethanol and stored at -20°C. Fixed embryos were dechorionated manually in petri dishes with poly-lysine as described in [60].

For nuclear staining, fixed whole embryos were rehydrated by washing in PBST during 10 min and were incubated in staining solution (1 µM Hoechst-33342 Invitrogen-62249 in PBST) for 5 to 10 min. ProLong Live Antifade Reagent (Invitrogen-P36975, dilution 1:100) was added to prevent photobleaching during fluorescence image capture. Nuclei were manually counted by analyzing the images of 10 to 15 optical sections for each embryo.

4.5 Diatom extracts and treatments

Extracts of the diatoms *Skeletonema marinoi*, *Chaetoceros affinis*, *Chaetoceros calcitrans* and the dinoflagellate *Prorocentrum minimum* were performed as described in [55] with minor modifications: microalgal pellets were frozen in liquid nitrogen and stored at -80ºC until cell lysis was performed by sonication. Stock solutions of 500

mg/mL of each extract were prepared in DMSO, avoiding light exposure. To minimize the toxic effect of DMSO in the embryo development of *O. dioica*, the stock solution was diluted 1:100 in sSW, which was used to create three incubation solutions at 5, 10 and 100 µg/mL. Treatments were performed as described for DD.

4.6 Statistical analysis

Statistical analysis was performed using RStudio Version 0.98.507 for Mac OS X. The results were reported as means ± SD (standard deviation) and analyzed by Tukey HSD (honest significant difference) test for multiple comparisons between control and treatments and between treatments, and by ANOVA the pairwise comparisons.

Acknowledgements

The authors thank all team members of the C.C. and R.A. laboratories for assistance with animal facility and fruitful discussions. We thank to Miriam Diaz-Gracia for her technical support on Wnt analyses, and to Valeria Mazziotti for her technical support with algal cultivations and extractions. C.C. was supported by BFU2016-80601-P. R.A. was supported by BIO2015-67358-C2-1-P grant from Ministerio de Economía y Competitividad (Spain). C.C. and R.A. were also supported by grant 2017-SGR-1665 from Generalitat de Catalunya. C.C and A.P. were also supported by EU-FP7 Assemble 1553.

Authors' contributions

N.T-A. carried out treatments, whole-mount *in situ* hybridizations, phenotype characterizations. N.T-A. and J.M-S. made microalgal extractions under the supervision of A.P., G.R. and S.D.. N.T-A, J.M-S., A.F-R. and A.A-A. contributed with cloning and expression analyses of developmental and defensome markers. N.T-A.

and C.C. interpret the data and made the figures. N.T-A., R.A. and C.C. wrote the MS with input from A.P., G.R. and S.D.. N.T-A. and V.R. performed statistical analyses. C.C. conceptualized the project. R.A. and C.C. supervised the project. All authors commented on the manuscript and agreed to its final version.

Figure legends

Figure 1. Effects and dose-response of DD on *Oikopleura dioica* embryo development. After 4.5 hours of DD treatment, embryo phenotypes could be classified in four categories: **(A)** normal hatchling (blue), **(B)** aberrant hatchling (yellow), **(C)** Pre-tailbud arrest with a "golf-ball" phenotype (salmon), and **(D)** 1-cell arrest (cherry). **(E)** The proportion of each aberrant phenotype at different DD concentrations allowed us to characterize the dose response and to define four ranges of concentrations depending on their severity: innocuous, mild, moderate and extreme. Colors correspond to phenotypes in (A). Number of analyzed embryos (n) and number replicas (r) are indicated on top of each treatment.

Figure 2. Characterization of 1-cell arrest phenotype caused by severe DD concentrations (2.5 µg/mL). (A) Percentage of fertilized embryos assessed by the formation of polar bodies do not show significant differences (P-value= 0.432) between control **(B)** and DD-treated embryos **(C)**, in which moreover no morphological differences in the presence of internal granules nor surface rugosity are observed. Maternal *Wnt* signal appeared asymmetrically distributed, mostly accumulated near the cell membrane that will form the prospective posterior vegetal pole both in control **(D)** and DD-treated embryos despite the absence of the first cleavage **(E)**, suggesting that DD does not alter the cellular architecture of the oocyte that drives primary axial polarity**.** Number of analyzed embryos (n) and number replicas (r) are indicated on top of each treatment.

Figure 3. Characterization of the "golf-ball" morphology at the pre-tailbud arrest stage caused by moderate DD concentrations (1 µg/mL). (A) Developmental progression of DD-treated embryos does not show any obvious abnormality until 3 hpf, when morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a 'golf-ball' morphology appears in an arrested "pre-tailbud" stage of DD-treated embryos, in which no trunk nor tail can be distinguished. **(B)** Nuclear staining with HOECHST reveals no significant differences in the number of cells between control and DD-treated embryos at 64-, 100-, 200-cell and mid-tailbud stage (ns, P-

values 0.725, 0.262, 0.206, respectively) **(C)**, suggesting that DD does not induce a general slow-down or blockage of development by the pre-tailbud arrest that produces the "golf-ball" morphology.

Figure 4. Alteration of developmental gene expression patterns by DD causing the "golf ball" morphology. Whole mount *in situ* hybridizations on DMSO control embryos **(A-AV)** and DD-treated embryos at 1 µg/mL **(A'-AV')** was performed throughout different developmental stages until hatch (i.e. 32-cell, 64-cell, 100-cell, 200-cell, mid-tailbud stage and just hatched) using markers for derivatives of all three germ cell layers: mesoderm (notochord labeled by *Bra, Wnt11* and *SoxBb*, and muscle labeled by *Act*), endoderm (entire endoderm labeled by *Tis11b*, only endoderm of the trunk labeled by *SoxBa* and *SoxBb*), and ectoderm (epidermis of the tail labeled by *SoxBb*, epidermis of the trunk labeled by *Tis11a,* and ventral epidermis of the trunk by *Nkx2.3/5/6*; neural cells by *Tis11a* and *SoxBa*, and presumptive placodal regions by *Wnt11* and *SoxBb*). Numbers of scored embryos with signal or without signal (insets) are shown in the top-right and bottom-right corners of each panel. All DMSO-control embryos presented signal as shown.

Figure 5. Defensome gene response to DD during embryo development. Whole mount *in situ* hybridizations on DMSO control embryos **(A-X)** and DD-treated embryos at 1 µg/mL **(A'-X')** was performed throughout different developmental stages until hatch (i.e. 8-cell, 64-cell and mid-tailbud stages and early-, mid- and late-hatchlings) using genes *Gclm, Adh3*, *Aldh2* and *Aldh8* reveals that DD activates a rapid response of the defensome during *O. dioica* development. Numbers of scored DD-treated embryos with signal are shown in the top-right corner.

Figure 6. Effects and defensome response of *Oikopleura dioica* embryos treated with extracts of different diatom species. (A) Whole mount *in situ* hybridizations with *Gclm* on DMSO control embryos and embryos treated with extracts of *S. marinoi* at different concentrations suggest that other PUAs than DD (e.g. heptadienal and octadienal) also activates a quick response of the defensome system. **(B)** Embryo phenotypes observed after treatments with extracts of *S. marinoi* (Sm), *C. affinis* (Ca), *C. calcitrans* (Cc) corresponds to the same observed after DD-treatment (**Fig. 1**). Extract of *P. minimum* (Pm), a non-producing aldehyde dinoflagellate, was used as additional control.

Figure 7. Schematic representation of the genetic patterning of *Oikopleura dioica* embryos with "golf ball" phenotype induced by DD exposure. Despite the failure to begin the process of morphogenesis of DD-treated embryos with "golf-ball" morphology, at the time in which control embryos has hatched, developmental gene markers show that the initial cell fate determination of derivatives of all three germ cell layers is not affected, and embryos show a correct AP and DV axial patterning. The main developmental processes that seem to be mostly affected are the failure of the formation of the indentation that separates trunk and tail, a delayed midline convergence, and absence of tail elongation.

Figure S1. Boxplot statistical representation of phenotype frequency distribution throughout treatments at different DD concentrations. **(A)** normal hatchling, **(B)** aberrant hatchling, **(C)** Pre-tailbud arrest with a "golf-ball" phenotype, and **(D)** 1-cell arrest. DD concentration at 0.05 µg/mL was considered innocuous, since no significant difference was observed in the amount of normal hatchlings in comparison with the control condition (ns, P-value 0.99843). At higher concentrations than 0.05 µg/mL we started to observe severe abnormalities than normally we do not observe in DMSO-control conditions (see **Fig. 1B**; P-values 0.1522 and <0.001, for 0.075 and 0.125 µg/mL, respectively). Lower and upper limits of the box represent the first and third quartile, respectively; the midline in the box represents the median; the bottom and top whiskers represent the minimum and maximum values, respectively; and the circles represent outliers.

Figure S2. Effects and dose-response of *Skeletonema marinoi* extracts on *Oikopleura dioica* embryo development. (A) After 4.5 hours of treatment with extracts of *S. marinoi* (*Sm*) at different concentrations (5, 10 and 100 µg/mL) the same embryo phenotypes were observed after DD-treatments (see **Fig. 1**: normal hatchling (blue), aberrant hatchling (yellow), Pre-tailbud arrest with a "golf-ball" phenotype (salmon), and 1-cell arrest (cherry) could be observed at different proportions depending on the concentration. Number of analyzed embryos (n) and number replicas (r) are indicated on top of each treatment. **(B)** Boxplot statistical representation of phenotype frequency distribution throughout treatments at different concentrations of *S. marinoi*. No significant differences were observed between embryos developed in sSW or DMSO (P-value 0.9996). Extracts at 5 and 10 µg/mL did not show significant differences with DMSO control (ns, P-value 0.0573 and 0.2301 for normal hatchling phenotype, respectively). **(C)** The developmental progression of embryos treated with *S. marinoi* extract at 100 µg/mL reproduced the same alterations observed with DD at

moderate concentrations (see **Fig. 2**). Thus, no obvious abnormalities are observed until 3 hpf, time at which morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a 'golf-ball' phenotype appears in arrested "pre-tailbud" stage of treated embryo. **(D)** Boxplot statistical representation of phenotype frequency distribution among treatments with different microalgal extracts at 100 µg/mL – *S. marinoi* (*Sm*), *C. affinis* (*Ca*), *C. calcitrans* (*Cc*) and *P. minimum* (*Pm*). Lower and upper limits of the boxes represent the first and third quartile, respectively; the midline in the box represents the median; the bottom and top whiskers represent the minimum and maximum values, respectively; and the circles represent outliers.

Figure S3. Determination of the requirement of pre-incubation time of oocytes during DD treatments. Using a DD concentration of 0.28 µg/mL –as previously used in ascidian experiments [13]–, oocytes were pre-incubated with DD during 0, 5, 10 and 20 minutes before fertilization, in at least 2 replicas, and at least 100 embryos analyzed (n) in each condition. After fertilization and continuous DD-treatment, we scored the percentage of animals that hatched (red), and from those what proportion had a normal morphology (blue). Results showed that the relative abundance of animals that hatched and from those that had normal morphology in relation to non-treated embryos (negative control) was significantly lower when pre-incubations were shorter than 10 minutes, and no significant (n.s.) decrease of hatch or normal hatched was observed with longer pre-incubation times (P-values 1.0 and 0.99992, respectively). We therefore always used pre-incubation times of 10 minutes as the standard procedure in our treatment experiments.

References

1. Wells, M.L. et al. Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae* **49**, 68-93 (2015).
2. Falkowski, P.G. The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth Res* **39**, 235-58 (1994).
3. Miralto, A. et al. Embryonic Development in Invertebrates Is Arrested by Inhibitory Compounds in Diatoms. *Mar Biotechnol (NY)* **1**, 401-402 (1999).
4. Ianora, A. et al. Aldehyde suppression of copepod recruitment in blooms of a ubiquitous planktonic diatom. *Nature* **429**, 403-7 (2004).
5. Wichard, T. et al. Survey of the chemical defence potential of diatoms: screening of fifty one species for alpha,beta,gamma,delta-unsaturated aldehydes. *J Chem Ecol* **31**, 949-58 (2005).
6. Ianora, A. & Miralto, A. Toxigenic effects of diatoms on grazers, phytoplankton and other microbes: a review. *Ecotoxicology* **19**, 493-511 (2010).
7. Longhurst, A.R. The structure and evolution of plankton communities. *Progress in Oceanography* **15**, 1-35 (1985).
8. Fontana, A. et al. Chemistry of oxylipin pathways in marine diatoms. *Pure Appl Chem* **79**, 481-90 (2007).

9. Romano, G., Russo, G.L., Buttino, I., Ianora, A. & Miralto, A. A marine diatom-derived aldehyde induces apoptosis in copepod and sea urchin embryos. *J Exp Biol* **206**, 3487-94 (2003).
10. Caldwell, G.S., Olive, P.J. & Bentley, M.G. Inhibition of embryonic development and fertilization in broadcast spawning marine invertebrates by water soluble diatom extracts and the diatom toxin 2-trans,4-trans decadienal. *Aquat Toxicol* **60**, 123-37 (2002).
11. Romano, G., Miralto, A. & Ianora, A. Teratogenic effects of diatom metabolites on sea urchin Paracentrotus lividus embryos. *Mar Drugs* **8**, 950-67 (2010).
12. Varrella, S. et al. Molecular response to toxic diatom-derived aldehydes in the sea urchin Paracentrotus lividus. *Mar Drugs* **12**, 2089-113 (2014).
13. Hansen, E., Even, Y. & Geneviere, A.M. The alpha, beta, gamma, delta-unsaturated aldehyde 2-trans-4-trans-decadienal disturbs DNA replication and mitotic events in early sea urchin embryos. *Toxicol Sci* **81**, 190-7 (2004).
14. Tosti, E. et al. Bioactive aldehydes from diatoms block the fertilization current in ascidian oocytes. *Mol Reprod Dev* **66**, 72-80 (2003).
15. Lettieri, A., Esposito, R., Ianora, A. & Spagnuolo, A. Ciona intestinalis as a marine model system to study some key developmental genes targeted by the diatom-derived aldehyde decadienal. *Mar Drugs* **13**, 1451-65 (2015).
16. Castellano, I., Ercolesi, E., Romano, G., Ianora, A. & Palumbo, A. The diatom-derived aldehyde decadienal affects life cycle transition in the ascidian Ciona intestinalis through nitric oxide/ERK signalling. *Open Biol* **5**, 140182 (2015).
17. Adolph, S. et al. Cytotoxicity of diatom-derived oxylipins in organisms belonging to different phyla. *J Exp Biol* **207**, 2935-46 (2004).
18. Caldwell, G.S., Lewis, C., Pickavance, G., Taylor, R.L. & Bentley, M.G. Exposure to copper and a cytotoxic polyunsaturated aldehyde induces reproductive failure in the marine polychaete Nereis virens (Sars). *Aquat Toxicol* **104**, 126-34 (2011).
19. Caldwell, G.S. The influence of bioactive oxylipins from marine diatoms on invertebrate reproduction and development. *Mar Drugs* **7**, 367-400 (2009).
20. Goldstone, J.V. et al. The chemical defensome: environmental sensing and response genes in the Strongylocentrotus purpuratus genome. *Dev Biol* **300**, 366-84 (2006).
21. Lauritano, C. et al. Molecular evidence of the toxic effects of diatom diets on gene expression patterns in copepods. *PLoS One* **6**, e26850 (2011).
22. Marrone, V. et al. Defensome against toxic diatom aldehydes in the sea urchin Paracentrotus lividus. *PLoS One* **7**, e31750 (2012).
23. Lauritano, C., Procaccini, G. & Ianora, A. Gene expression patterns and stress response in marine copepods. *Mar Environ Res* **76**, 22-31 (2012).
24. Carotenuto, Y. et al. Insights into the transcriptome of the marine copepod Calanus helgolandicus feeding on the oxylipin-producing diatom Skeletonema marinoi. *Harmful Algae* **31**, 153–162 (2014).
25. Albalat, R. & Canestro, C. Evolution by gene loss. *Nat Rev Genet* **17**, 379-391 (2016).
26. Gorsky, G. & Fenaux, R. in The Biology of Pelagic Tunicates (ed. Bone, Q.) (Oxford University Press, Oxford, 1998).
27. Capitanio, F.L. et al. Seasonal cycle of Appendicularians at a coastal station (38º28´S, 57º41´W) of the SW Atlantic Ocean. *Bull. Mar. Sci.* **82**, 171-184 (2008).
28. Acuña, J.L. et al. Phytoplankton ingestion by appendicularians in the North Water. *DEEP-SEA RESEARCH PART II-TOPICAL STUDIES IN OCEANOGRAPHY* **49**, 5101-5115 (2002).
29. Flood, P.R. & Deibel, D. in The Biology of Pelagic Tunicates (ed. Bone, Q.) 105-137 (Oxford University Press, Oxford, 1998).
30. Diego, F.Â.n., ÃƒÂ ngel, L.Â.p.-U., Antonio, F.Â.n., JosÃƒÂ© Luis, A.Â.a. & Roger, H. Retention efficiency of 0.2 to 6 µm particles by the appendicularians Oikopleura dioica and Fritillaria borealis. *Marine Ecology Progress Series* **266**, 89-101 (2004).
31. Davoll, P. & Youngbluth, M. Heterotrophic activity on appendicularian (Tunicata: Appendicularia) houses in mesopelagic regions and their potential contribution to particle flux. *Deep-Sea Research* **37**, 285-294. (1990).
32. Robison, B.H., Reisenbichler, K.R. & Sherlock, R.E. Giant larvacean houses: rapid carbon transport to the deep sea floor. *Science* **308**, 1609-11 (2005).

33. Troedsson, C. et al. Effects of ocean acidification, temperature and nutrient regimes on the appendicularian Oikopleura dioica: a mesocosm study. *Marine Biology* **160**, 2175-2187 (2013).
34. Seo, H.C. et al. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **431**, 67-71 (2004).
35. Cañestro, C., Yokoi, H. & Postlethwait, J.H. Evolutionary developmental biology and genomics. *Nat Rev Genet* **8**, 932-42 (2007).
36. Nishida, H. Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages. *Dev Growth Differ* **50 Suppl 1**, S239-56 (2008).
37. Cañestro, C., Albalat, R., Irimia, M. & Garcia-Fernandez, J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin Cell Dev Biol* **24**, 83-94 (2013).
38. Edvardsen, R.B. et al. Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr Biol* **15**, R12-3 (2005).
39. Cañestro, C., Postlethwait, J.H., Gonzàlez-Duarte, R. & Albalat, R. Is retinoic acid genetic machinery a chordate innovation? *Evol Dev.* **8**, 394-406. (2006).
40. Denoeud, F. et al. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* **330**, 1381-5 (2010).
41. Marti-Solans, J. et al. Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* **53**, 183-93 (2015).
42. Marti-Solans, J. et al. Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. *Mol Biol Evol* **33**, 2401-16 (2016).
43. Cunningham, T.J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol* **16**, 110-23 (2015).
44. Rhinn, M. & Dolle, P. Retinoic acid signalling during development. *Development* **139**, 843-58 (2012).
45. Duester, G. Retinoid signaling in control of progenitor cell differentiation during mouse development. *Semin Cell Dev Biol* **24**, 694-700 (2013).
46. Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. & Rahuel-Clermont, S. Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. *Chem Biol Interact* **202**, 78-84 (2013).
47. Battistoni, M. et al. The Ascidian Embryo Teratogenicity assay in Ciona intestinalis as a new teratological screening to test the mixture effect of the co-exposure to ethanol and fluconazole. *Environ Toxicol Pharmacol* **57**, 76-85 (2018).
48. Dumollard, R., Gazo, I., Gomes, I.D.L., Besnardeau, L. & McDougall, A. Ascidians: An Emerging Marine Model for Drug Discovery and Screening. *Curr Top Med Chem* **17**, 2056-2066 (2017).
49. Bassham, S. & Postlethwait, J. Brachyury (T) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of T. *Dev Biol* **220**, 322-32. (2000).
50. Nishino, A., Satou, Y., Morisawa, M. & Satoh, N. Brachyury (T) gene expression and notochord development in *Oikopleura longicauda* (Appendicularia, Urochordata). *Dev. Genes Evol.* **211**, 219-231 (2001).
51. Nishino, A., Satou, Y., Morisawa, M. & Satoh, N. Muscle actin genes and muscle cells in the appendicularian, Oikopleura longicauda: phylogenetic relationships among muscle tissues in the urochordates. *J Exp Zool* **288**, 135-50 (2000).
52. Onuma, T.A., Matsuo, M. & Nishida, H. Modified whole-mount in situ hybridisation and immunohistochemistry protocols without removal of the vitelline membrane in the appendicularian Oikopleura dioica. *Dev Genes Evol* **227**, 367-374 (2017).
53. Cañestro, C., Albalat, R. & Postlethwait, J.H. Oikopleura dioica alcohol dehydrogenase class 3 provides new insights into the evolution of retinoic acid synthesis in chordates. *Zoolog Sci* **27**, 128-33 (2010).
54. Cañestro, C. & Postlethwait, J.H. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* **305**, 522-38 (2007).
55. Barreiro, A. et al. Diatom induction of reproductive failure in copepods: The effect of PUAs versus non volatile oxylipins. *Journal of Experimental Marine Biology and Ecology* **401**, 13-19 (2011).
56. Bartual, A. et al. Polyunsaturated aldehydes from large phytoplankton of the Atlantic Ocean surface (42 degrees n to 33 degrees s). *Mar Drugs* **12**, 682-99 (2014).

57. Vidoudez, C., Casotti, R., Bastianini, M. & Pohnert, G. Quantification of dissolved and particulate polyunsaturated aldehydes in the Adriatic sea. *Mar Drugs* **9**, 500-13 (2011).
58. Ribalet, F. et al. Phytoplankton cell lysis associated with polyunsaturated aldehyde release in the Northern Adriatic Sea. *PLoS One* **9**, e85947 (2014).
59. Danks, G. et al. OikoBase: a genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. *Nucleic Acids Res* **41**, D845-53 (2013).
60. Cañestro, C., Bassham, S. & Postlethwait, J.H. Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. *Dev Biol* **285**, 298-315 (2005).

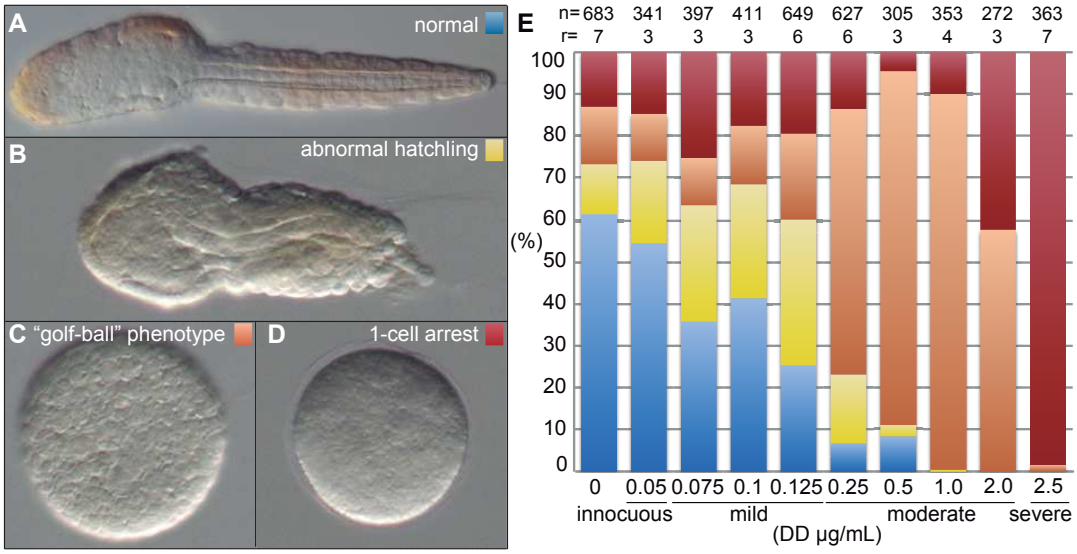


Figure 1. Effects and dose-response of DD on *Oikopleura dioica* embryo development. After 4.5 hours of DD treatment, embryo phenotypes could be classified in four categories: **(A)** normal hatchling (blue), **(B)** aberrant hatchling (yellow), **(C)** Pre-tailbud arrest with a "golf-ball" phenotype (salmon), and **(D)** 1-cell arrest (cherry). **(E)** The proportion of each aberrant phenotype at different DD concentrations allowed us to characterize the dose response and to define four ranges of concentrations depending on their severity: innocuous, mild, moderate and extreme. Colors correspond to phenotypes in (A). Number of analyzed embryos (n) and number replicas (r) are indicated on top of each treatment.

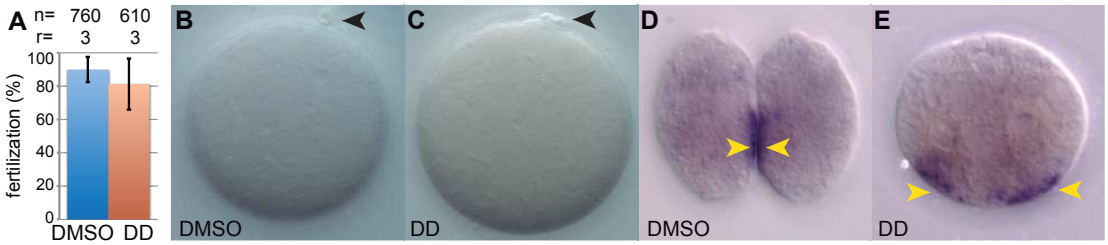


Figure 2. Characterization of 1-cell arrest phenotype caused by severe DD concentrations (2.5 µg/mL). **(A)** Percentage of fertilized embryos assessed by the formation of polar bodies do not show significant differences (P-value= 0.432) between control **(B)** and DD-treated embryos **(C)**, in which moreover no morphological differences in the presence of internal granules nor surface rugosity are observed. Maternal *Wnt* signal appeared asymmetrically distributed, mostly accumulated near the cell membrane that will form the prospective posterior vegetal pole both in control **(D)** and DD-treated embryos despite the absence of the first cleavage **(E)**, suggesting that DD does not alter the cellular architecture of the oocyte that drives primary axial polarity. Number of analyzed embryos (n) and number replicas (r) are indicated on top of each treatment.

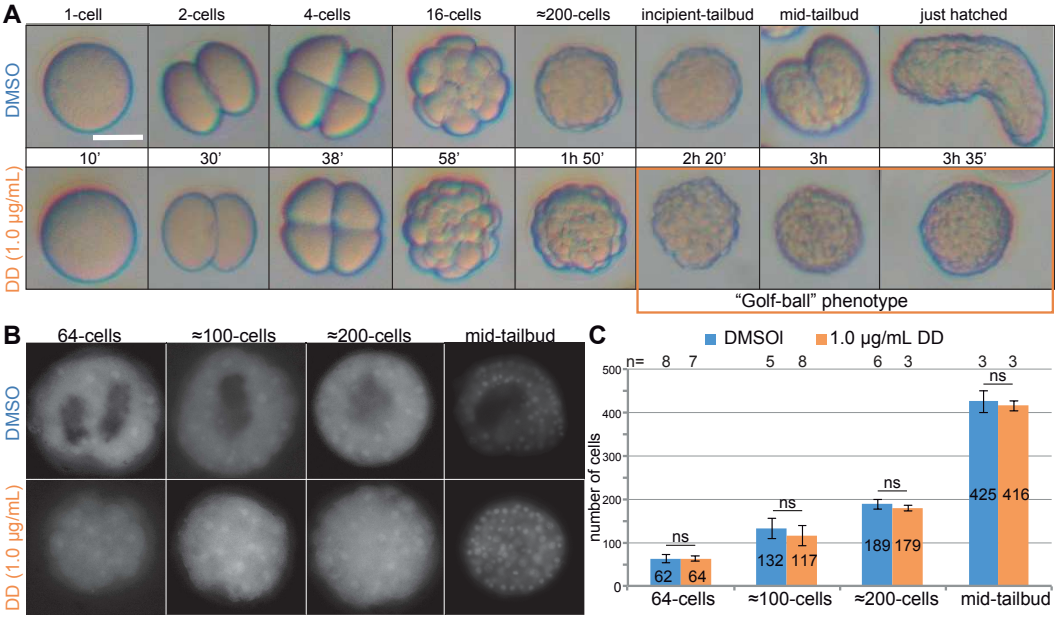


Figure 3. Characterization of the "golf-ball" morphology at the pre-tailbud arrest stage caused by moderate DD concentrations (1 µg/mL). (A) Developmental progression of DD-treated embryos does not show any obvious abnormality until 3 hpf, when morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a 'golf-ball' morphology appears in an arrested "pre-tailbud" stage of DD-treated embryos, in which no trunk nor tail can be distinguished. **(B)** Nuclear staining with HOECHST reveals no significant differences in the number of cells between control and DD-treated embryos at 64-, 100-, 200-cell and mid-tailbud stage (ns, P-values 0.725, 0.262, 0.206, respectively) **(C)**, suggesting that DD does not induce a general slow-down or blockage of development by the pre-tailbud arrest that produces the "golf-ball" morphology.

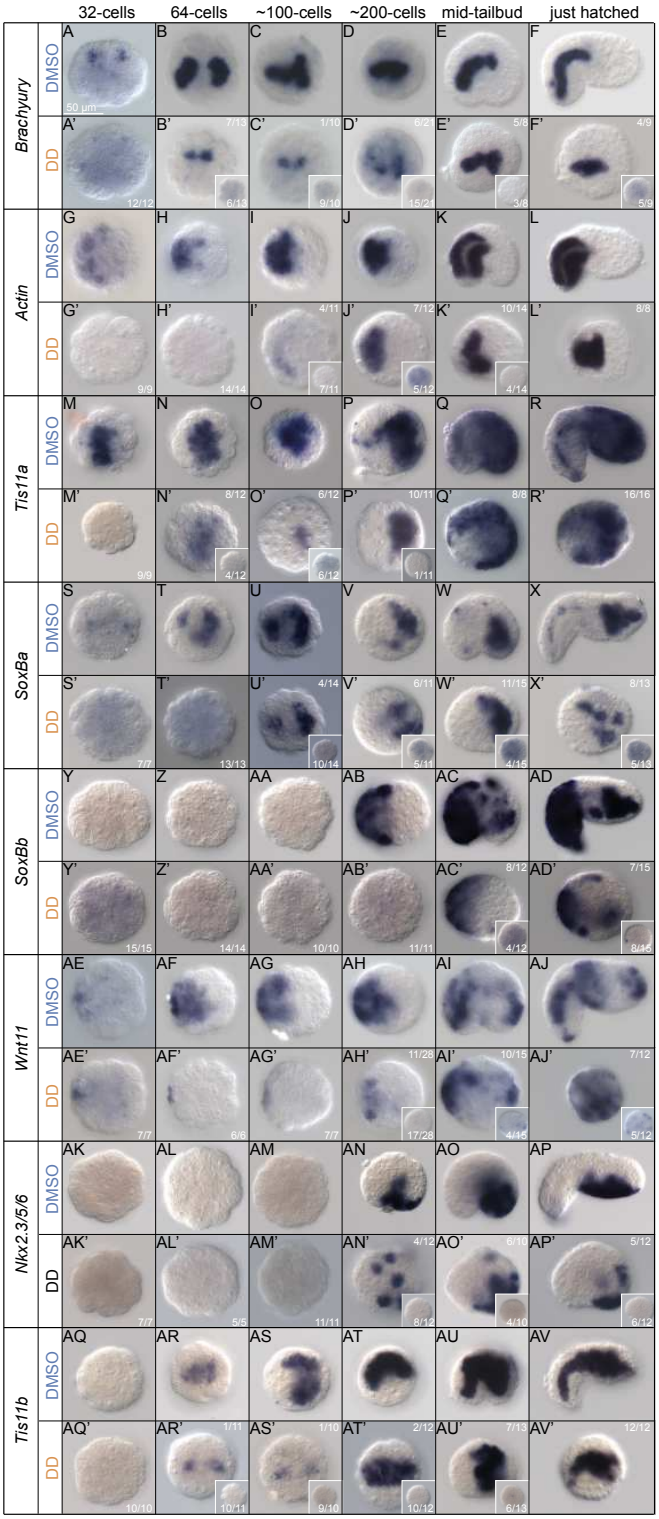


Figure 4. Alteration of developmental gene expression patterns by DD causing the "golf ball" morphology. Whole mount in situ hybridizations on DMSO control embryos **(A-AV)** and DD-treated embryos at 1 µg/mL **(A'-AV')** was performed throughout different developmental stages until hatch (i.e. 32-cell, 64-cell, 100-cell, 200-cell, mid-tailbud stage and just hatched) using markers for derivatives of all three germ cell layers: mesoderm (notochord labeled by *Bra*, *Wnt11* and *SoxBb*, and muscle labeled by *Act*), endoderm (entire endoderm labeled by *Tis11b*, only endoderm of the trunk labeled by *SoxBa* and *SoxBb*), and ectoderm (epidermis of the tail labeled by *SoxBb*, epidermis of the trunk labeled by *Tis11a*, and ventral epidermis of the trunk by *Nkx2.3/5/6*; neural cells by *Tis11a* and *SoxBa*, and presumptive placodal regions by *Wnt11* and *SoxBb*). Numbers of scored embryos with signal or without signal (insets) are shown in the top-right and bottom-right corners of each panel. All DMSO-control embryos presented signal as shown.

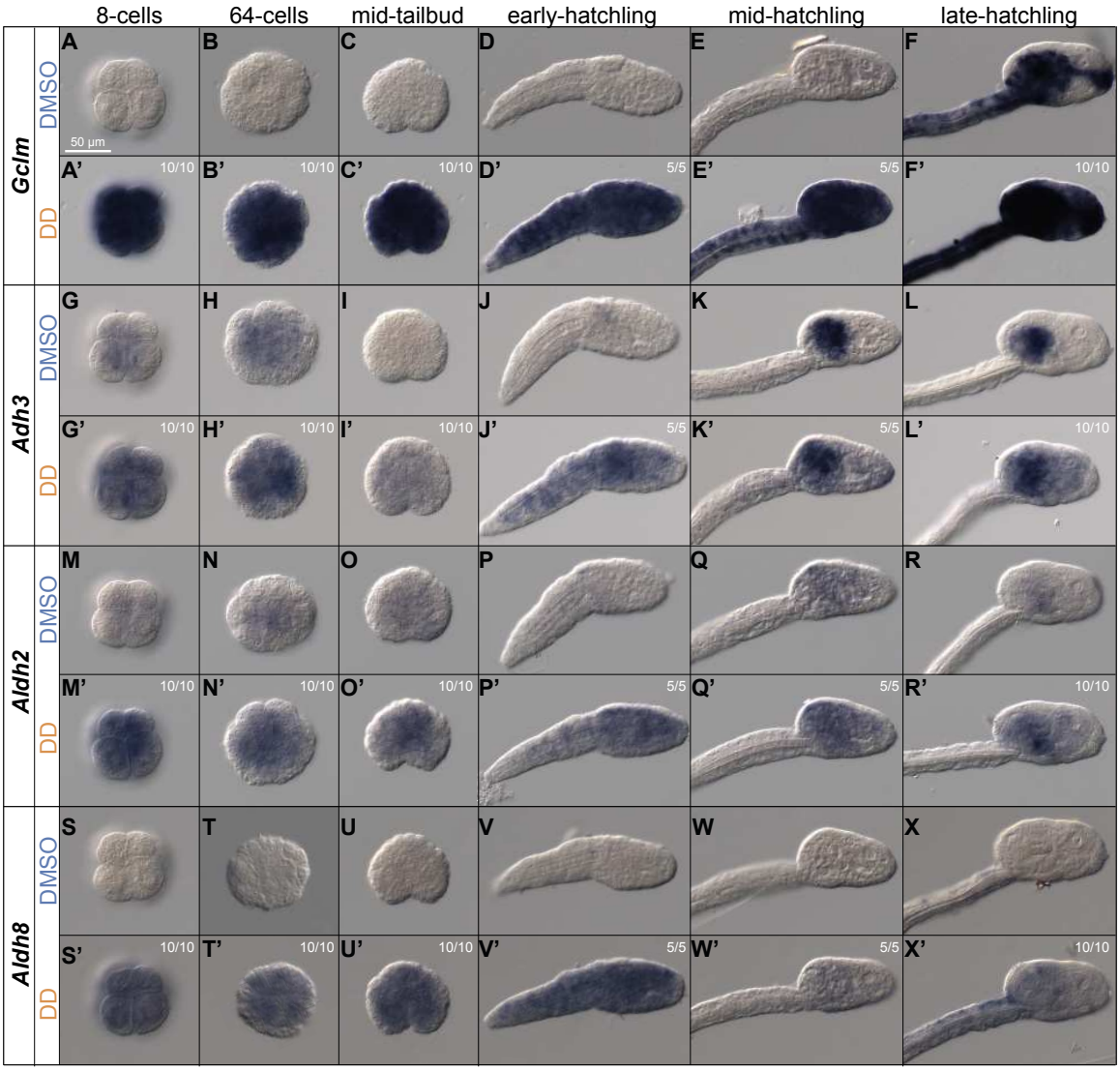


Figure 5. Defensome gene response to DD during embryo development. Whole mount in situ hybridizations on DMSO control embryos **(A-X)** and DD-treated embryos at 1 µg/mL **(A'-X')** was performed throughout different developmental stages until hatch (i.e. 8-cell, 64-cell and mid-tailbud stages and early-, mid- and late-hatchlings) using genes *Gclm*, *Adh3*, *Aldh2* and *Aldh8* reveals that DD activates a rapid response of the defensome during *O. dioica* development. Numbers of scored DD-treated embryos with signal are shown in the top-right corner.

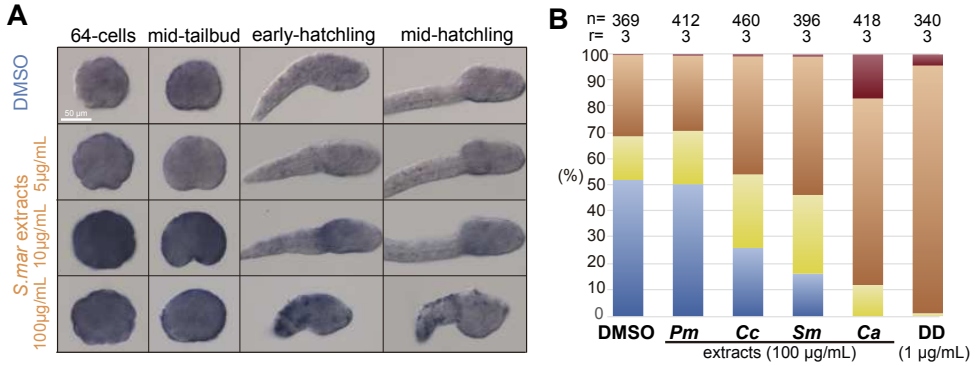


Figure 6. Effects and defensome response of *Oikopleura dioica* **embryos treated with extracts of different diatom species. (A)** Whole mount in situ hybridizations with *Gclm* on DMSO control embryos and embryos treated with extracts of *S. marinoi* at different concentrations suggest that other PUAs than DD (e.g. heptadienal and octadienal) also activates a quick response of the defensome system. **(B)** Embryo phenotypes observed after treatments with extracts of *S. marinoi* (Sm), *C. affinis* (Ca), *C. calcitrans* (Cc) corresponds to the same observed after DD-treatment (**Fig. 1**). Extract of *P. minimum* (Pm), a non-producing aldehyde dinoflagellate, was used as additional control.

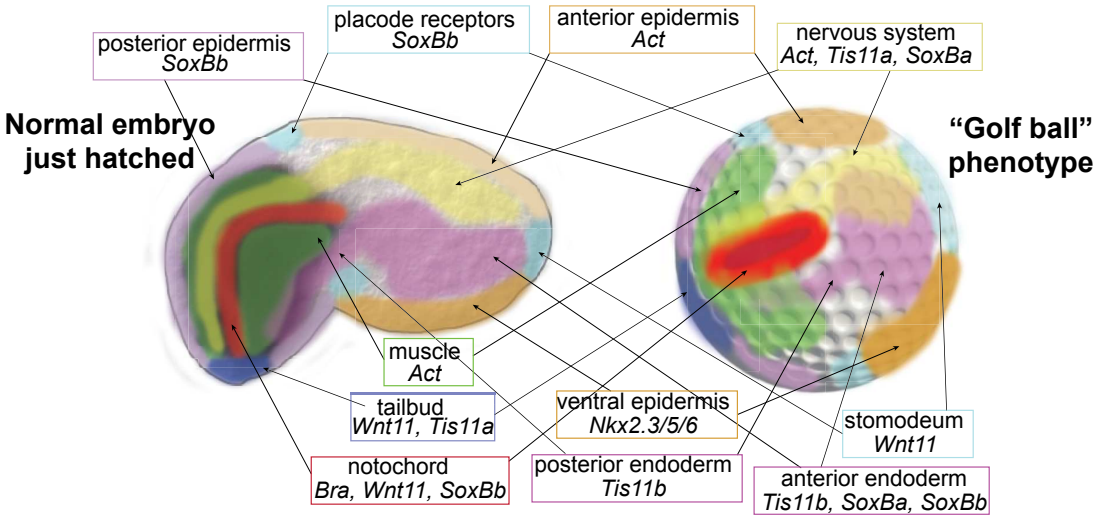
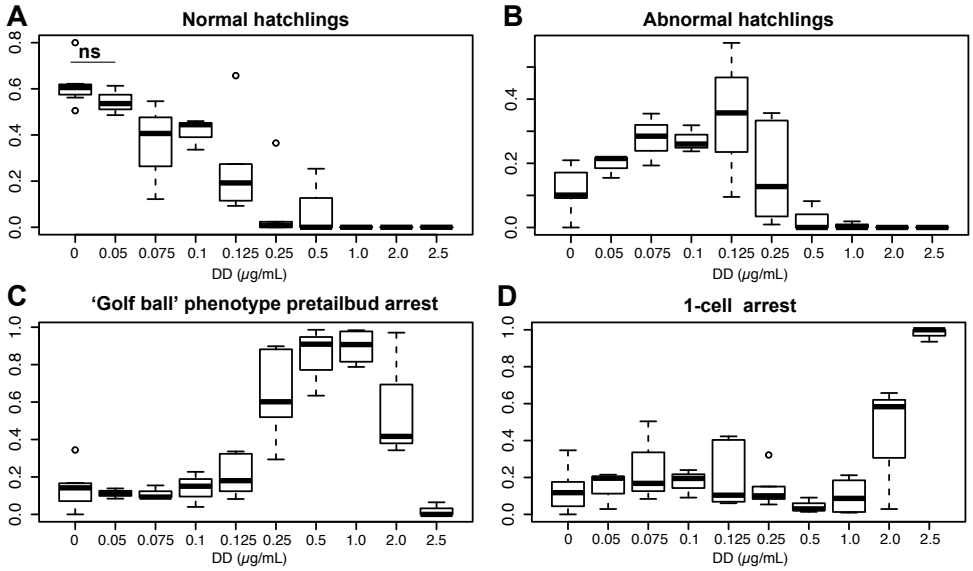
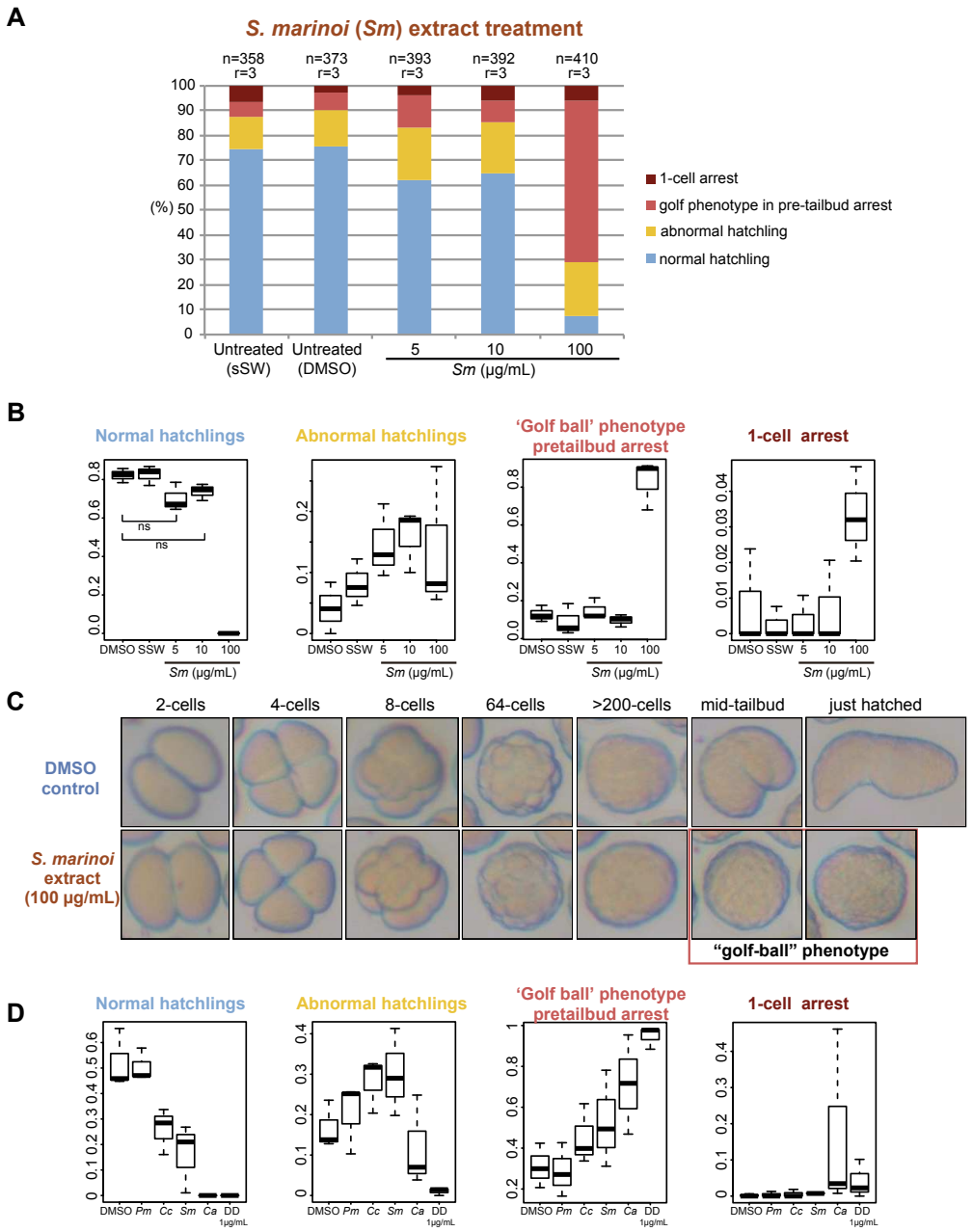


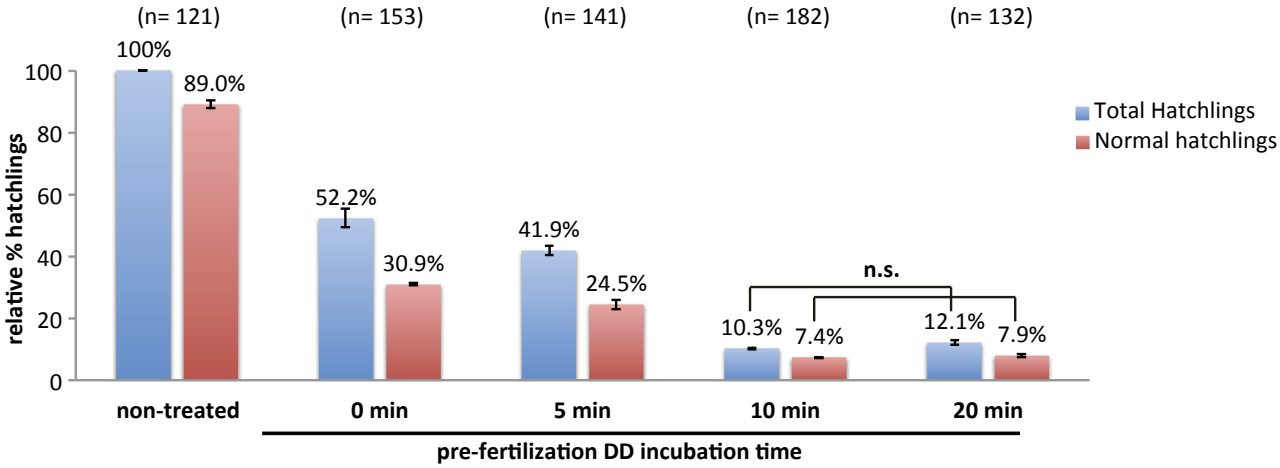
Figure 7. Schematic representation of the genetic patterning of *Oikopleura dioica* embryos with “golf ball” phenotype induced by DD exposure. Despite the failure to begin the process of morphogenesis of DD-treated embryos with “golf-ball” morphology, at the time in which control embryos has hatched, developmental gene markers show that the initial cell fate determination of derivatives of all three germ cell layers is not affected, and embryos show a correct AP and DV axial patterning. The main developmental processes that seem to be mostly affected are the failure of the formation of the indentation that separates trunk and tail, a delayed midline convergence, and absence of tail elongation.



S1 Fig. Boxplot statistical representation of phenotype frequency distribution throughout treatments at different DD concentrations. (A) normal hatchling, **(B)** aberrant hatchling, **(C)** Pre-tailbud arrest with a "golf-ball" phenotype, and **(D)** 1-cell arrest. DD concentration at 0.05 µg/mL was considered innocuous, since no significant difference was observed in the amount of normal hatchlings in comparison with the control condition (ns, P-value 0.99843). At higher concentrations than 0.05 µg/mL we started to observe severe abnormalities than normally we do not observe in DMSO-control conditions (see **Fig. 1B**; P-values 0.1522 and <0.001, for 0.075 and 0.125 µg/mL, respectively). Lower and upper limits of the box represent the first and third quartile, respectively; the midline in the box represents the median; the bottom and top whiskers represent the minimum and maximum values, respectively; and the circles represent outliers.



S2 Fig. Effects and dose-response of Skeletonema marinoi extracts on *Oikopleura dioica* embryo development. (A) After 4.5 hours of treatment with extracts of *S. marinoi* (*Sm*) at different concentrations (5, 10 and 100 µg/mL) the same e mbryo phenotypes were observed after DD-treatments (see Fig. 1: normal hatchling (blue), aberrant hatchling (yellow), Pre-tailbud arrest with a "golf-ball" phenotype (salmon), and 1-cell arrest (cherry) could be observed at different proportions depending on the concentration. Number of analyzed embryos (n) and number replicas (r) are indicated on top o f each treatment. **(B)** Boxplot statistical representation of phenotype frequency distribution throughout treatments at diffe rent concentrations of *S. marinoi*. No significant differences were observed between embryos developed in sSW or DMSO (P-value 0.9996). Extracts at 5 and 10 µg/mL did not show significant differences with DMSO control (ns, P-value 0.0573 an d 0.2301 for normal hatchling phenotype, respectively). **(C)** The developmental progression of embryos treated with S. marinoi extract at 100 µg/mL reproduced the same alterations observed with DD at moderate concentrations (see Fig. 2). Thus, no obvious abnormalities are observed until 3 hpf, time at which morphogenesis starts during the formation of tailbud stages in DMSO-control embryos, but a 'golf-ball' phenotype appears in arrested "pre-tailbud" stage of treated embryo. **(D)** Boxplot statistical representation of phenotype frequency distribution among treatments with different microalgal extracts at 100 µg/mL – *S. marinoi* (*Sm*), *C. affinis* (*Ca*), *C. calcitrans* (*Cc*) and *P. minimum* (*Pm*). Lower and upper limits of the boxes represent the first and third quartile, respectively; the midline in the box represents the median; the bottom and top whiskers represent the minimum and maximum values, respectively; and the circles represent outliers.



S3 Fig. Determination of the requirement of pre-incubation time of oocytes during DD treatments. Using a DD concentration of 0.28 µg/mL –as previously used in ascidian experiments [14]–, oocytes were pre-incubated with DD during 0, 5, 10 and 20 minutes before fertilization, in at least 2 replicas, and at least 100 embryos analyzed (n) in each condition. After fertilization and continuous DD-treatment, we scored the percentage of animals that hatched (red), and from those what proportion had a normal morphology (blue). Results showed that the relative abundance of animals that hatched and from those that had normal morphology in relation to non-treated embryos (negative control) was significantly lower when pre-incubations were shorter than 10 minutes, and no significant (n.s.) decrease of hatch or normal hatched was observed with longer pre-incubation times (P-values 1.0 and 0.99992, respectively). We therefore always used pre-incubation times of 10 minutes as the standard procedure in our treatment experiments.

S1 Table. Developmental and defensome gene markers. Oikobase IDs, names and sequences.

Gene	Gene ID	Primer name	Sequence (5'- 3')
Brachyury	GSOIDG00000279001	OdiT_F3	GGTTCGCACTGGATGAAACAGCC
		OdiT_R3	TATCCGTTCTGACACCAGTCGTTC
Actin	GSOIDG00000756001	Actin_5'_F	GTCCCCGCCATGTACGTCTG
		Actin_3'_R	GCATCGGAATCGCTCGTTACCA
Tis11a	GSOIDG00015222001	ZFC3Ha_e5_F	GGGTACTGCCCATATGGCG
		ZFC3Ha_e7_R	GCTCGAAGTTGGGCAGCTG
SoxBa	GSOIDG00010386001	SoxBa_e6_F	GCAGAAGTACCCAGCAAGGA
		SoxBa_e6_R	GTGACCACTTTCGGGCTTGT
SoxBb	GSOIDG00013526001	SoxBb_e4_F	GTTGTCGCTGGAAATGGCGA
		SoxBb_e8_R	CTCGACACGGACGCTCTGAT
Wnt11	GSOIDG00011688001	Wnt_s4_atgF	ATGAAGATTTCAGTCACCCTTTTCTCTG
		Wnt_s4_stopR	GTTATTTGCATATATGAGTGACAGTCG
Nkx2.3/5/6	GSOIDG00003812001	Nkx_5'start_F	GACCGAAAAATTACAACTATGAGC
		Nkx_ex_3_R	GCTGTAGCGCCGAGCTCAC
Tis11b	GSOIDG00017080001	ZFC3Hb_e2_F	GGCCAAATGAACGACGAAATCG
		ZFC3Hb_e4_R	GCACTCGGAGAGCGAGAG
Glcm	GSOIDG00006303001	Glcm_e2_F	GCAATAATTATCCAAGATGCCATG
		Glcm_e5_R	GTTCAGTCCTGCAAAGTATCC
Adh3	GSOIDG00000110001	Adh3_F1	CGTCGGTAAAGTGATCACGTGCA
		Adh3_R1	GCGCCCTGTGTAGCTCGGAC
Aldh2	GSOIDG00002220001	Aldh2_F	TGGAACTTCCCTCTCCTCATGCA
		Aldh2_R	TTATTTGGCGTATTGAGGAAGTTTCAT
Aldh8a1	GSOIDG00021101001	Aldh8_F	GATTTAAACAAAAAATGGAGCCGATTG
		Aldh8_R	TTTATTGCTTGCTAATTGTATTAGCTTGAAG